Problème d'extremum

Attention ! Dans ce chapitre, toutes les fonctions seront à valeurs réelles. Cela n'a pas de sens de chercher un maximum (ou un minimum) pour une fonction à valeurs vectorielles. On ne supposera pas toujours que le domaine de définition est ouvert (en particulier pour les définitions générales).

1 Définitions : local ou global

On donne les définitions dans un cas très général même si, dans la pratique X sera toujours une partie d'un espace vectoriel (voire de $\mathbb{R}^p$).

Soit f une fonction définie d'un ensemble X à valeurs dans $\mathbb{R}$ et a un point de X.

— a est un maximum (global) de f sur X si

$$\forall x \in X,\ f(a) \geqslant f(x).$$

Ce maximum est strict si l'inégalité est stricte pour $x \neq a$.

— a est un minimum (global) de f sur X si

$$\forall x \in X,\ f(a) \leqslant f(x).$$

Ce minimum est strict si l'inégalité est stricte pour $x \neq a$.

— a est un extremum si a est un maximum ou un minimum.

Pour définir la notion d'extremum local, il faut pouvoir définir des voisinages. On supposera donc que X est une partie d'un espace métrique (E, d) (dans la pratique, E sera presque toujours un espace vectoriel normé).

— a est un maximum local de f sur X si il existe $r > 0$ tel que

$$\forall x \in B(a, r) \cap X,\ f(a) \geqslant f(x).$$

Ce maximum local est strict si l'inégalité est stricte pour $x \neq a$.

— a est un minimum local de f sur X si il existe $r > 0$ tel que

$$\forall x \in B(a, r) \cap X,\ f(a) \leqslant f(x).$$

Ce minimum local est strict si l'inégalité est stricte pour $x \neq a$.

— a est un extremum local si a est un maximum local ou un minimum local.

Remarque : si X est ouvert dans E, quitte à prendre un plus petit r on peut supposer que $B(a, r) \subset X$.

Rappel : Lorsque (X, d) est un espace métrique compact (par exemple une partie compacte d'un espace vectoriel normé) et f une application **continue** de X dans $\mathbb{R}$, on sait déjà qu'il existe (au moins) un maximum global de f sur X et (au moins) un minimum global de f sur X. On fera attention qu'une partie ouverte d'un espace vectoriel normé n'est jamais compacte.

2 Une condition nécessaire d'ordre 1

2.1 Rappel en dimension 1.

Soit f une fonction définie sur un intervalle ouvert I de $\mathbb{R}$, à valeurs réelles et soit a un extremum local de f. Si f est dérivable en a alors $f'(a) = 0$.

Démonstration. On fait la preuve dans le cas d'un maximum. Comme I est ouvert et $a \in I$, il existe $r > 0$ tel que, $]a - r, a + r[\subset I$ et

$$\forall x \in]a - r, a + r[, \;\; f(x) \leqslant f(a).$$

On écrit la formule du taux d'accroissement pour $x > a$ et on fait tendre x vers a. Puisque f est dérivable en a on obtient en passant à la limite

$$f'(a) \leqslant 0.$$

Le même raisonnement pour $x < a$ donne

$$f'(a) \geqslant 0,$$

d'où la conclusion. □

Lorsque f est définie et continue sur un intervalle compact $[x_-, x_+]$, et dérivable sur l'ouvert $]x_-, x-+[$ on en déduit que ou bien le maximum est atteint en x_- ou en x_+ ou bien il est atteint en un point $a \in]x_-, x_+[$ et alors $f'(a) = 0$. Ce raisonnement est à la base de la preuve du théorème de Rolle.

2.2 Généralisation

Soit f une fonction définie sur un ouvert $\Omega \subset \mathbb{R}^p$ à valeurs dans $\mathbb{R}$. Supposons que $a \in \Omega$ est un minimum local. Pour tout vecteur $v \in \mathbb{R}^n$, on définit la fonction F_v sur un petit intervalle qui contient 0 par $F_v(t) = f(a + tv)$. Il est évident que si a est un minimum local pour f alors, pour tout v, $t = 0$ doit être un minimum local de F_v.

De plus, si la fonction f est différentiable en a alors la fonction F_v est dérivable en $t = 0$ et

$$\begin{aligned} F_v'(0) &= \sum_{i=1}^{n} v_i \frac{\partial f}{\partial x_i}(a) \\ &= df_{|a}(v). \end{aligned}$$

D'après le critère en dimension 1, si a est un minimum local alors, nécessairement

$$\forall v \in \mathbb{R}^p, \; \sum_{i=1}^{n} v_i \frac{\partial f}{\partial x_i}(a) = 0.$$

Cela n'est possible que si

$$\forall 1 \leqslant i \leqslant n, \;\; \frac{\partial f}{\partial x_i}(a) = 0,$$

ou encore

$$df_{|a} = 0.$$

Le même raisonnement s'applique pour un maximum local, on en déduit le critère suivant.

Proposition 1. *Soit E un espace vectoriel normé et f une application définie sur un ouvert $\Omega \subset E$ à valeurs réelles. Soit a un extremum local tel que f est différentiable en a alors*

$$df_{|a} = 0.$$

Dans le cas où $E = \mathbb{R}^p$, dire que $df_{|a} = 0$ équivaut à

$$\forall 1 \leqslant j \leqslant p,\ \frac{\partial f}{\partial x_j}(a) = 0.$$

Définition 1. *Soit f une application définie d'un ouvert $\Omega \subset E$ à valeurs réelles. On dit que a est un point critique de f si f est différentiable en a et si $df_{|a} = 0$.*

D'après ce qui précède, on a donc la **condition nécessaire d'ordre** 1 : si f est différentiable sur l'ouvert $\Omega \subset E$ et si a est un extremum local de f alors a est un point critique de f.

3 Etude des points critiques

3.1 En dimension 1

Soit f une fonction définie sur un voisinage de 0 qui admet un point critique en 0. On suppose que f est au moins C^2. D'après la formule de Taylor-Young en 0 on a

$$f(x) - f(0) = \frac{1}{2}f''(0)x^2 + o(x^2) = \frac{x^2}{2}(f''(0) + r(x)),$$

où r est une fonction qui tend vers 0 en 0.

Si 0 est un minimum local, on a donc

$$\frac{x^2}{2}(f''(0) + r(x)) \geqslant 0.$$

On divise par x^2 puis on fait tendre x vers 0 pour trouver

$$f''(0) \geqslant 0.$$

Inversement, si 0 est un maximum local, on trouve que $f''(0) \leqslant 0$. Remarquons que même si l'extremum est strict, l'inégalité ci-dessus reste large puisqu'elle est obtenue en passant à la limite.

Réciproquement, si $f''(0) \neq 0$, pour x assez petit, $(f''(0) + r(x))$ est du signe de $f''(0)$. On obtient donc les conditions suffisantes suivantes :

— Si $f''(0) > 0$ alors 0 est un minimum local.

— Si $f''(0) < 0$ alors 0 est un maximum local.

Si $f''(0) = 0$, les exemples suivants montrent qu'on ne peut pas conclure (sans hypothèses supplémentaires) quant à savoir si 0 est un extremum ou pas (ni, si c'en est un, sur la nature de cet extremum)

— Pour $x \mapsto x^3$, 0 est un point critique qui n'est pas un extremum local.

— Pour $x \mapsto x^4$, 0 est un point critique qui est un minimum local.

— Pour $x \mapsto -x^4$, 0 est un point critique qui est un maximum local.

3.2 En dimension plus grande

En dimension plus grande, soit f définie de l'ouvert $\Omega \subset \mathbb{R}^p$ au moins C^2 et soit a est un point critique de f. On définit, comme précédemment, la fonction $t \mapsto f(a + tv)$ et on essaie d'utiliser la caractérisation de la dimension 1.

On doit donc chercher un développement limité pour cette fonction. Comme f est C^2 sur Ω, pour tout v, F_v sera C^2 sur un petit intervalle qui contient 0. De plus, on peut calculer la dérivée :

$$F_v'(t) = \sum_{i=1}^{p} v_i \frac{\partial f}{\partial x_i}(a + tv),$$

et la dérivée seconde :

$$F_v''(t) \;=\; \sum_{i,j=1}^{p} v_i v_j \frac{\partial^2 f}{\partial x_i \partial x_j}(a+tv).$$

La formule de Taylor en 0 pour F_v donne donc :

$$f(a+tv) \;-\; f(a) \;=\; \frac{t^2}{2}\left(Q(v)+r(t)\right)$$

où r tend vers 0 quand t tend vers 0 et Q est la forme quadratique définie par

$$Q(v) \;=\; \sum_{i,j=1}^{p} v_i v_j \frac{\partial^2 f}{\partial x_i \partial x_j}(a)$$

D'après le théorème de Schwarz sur les dérivées croisées, cette forme quadratique est associée à la forme bilinéaire symétrique représentée par la matrice hessienne

$$H(a) \;=\; \Big[\frac{\partial^2 f}{\partial x_i \partial x_j}(a)\Big]_{1\leqslant, i,j \leqslant p}.$$

C'est à dire que $Q(v) \;=\; {}^tVH(a)V$ où V est le vecteur colonne représentant v dans la base canonique.

On déduit de l'analyse précédente la **condition nécessaire d'ordre** 2 : si f est au moins C^2 sur l'ouvert Ω, et si le point critique a est un maximum (resp. minimum) local alors, pour tout v, $Q(v) \leqslant 0$. (resp. pour tout v, $Q(v) \geqslant 0$).

Par contraposée, on peut formuler différemment ce critère : si on trouve deux vecteurs v et w tels que $Q(v)Q(w) < 0$ alors le point critique a n'est pas un extremum local.

Ainsi, l'étude des points critiques est reliée à l'étude des formes quadratiques dont on rappelle dans la partie suivante les résultats principaux vus en $L2$.

3.2.1 Rappels d'algèbre linéaire

On travaille dans $\mathbb{R}^p$ muni du produit scalaire canonique. (ce qui revient à travailler dans un espace de dimension finie p muni d'un produit scalaire de référence, pour lequel on a choisi une base orthonormée).

La forme quadratique Q est donc représentée par une matrice symétrique A de sorte que

$$Q(X) \;=\; {}^tXAX.$$

La forme quadratique Q est dite **positive** si $Q(X) \geqslant 0$ pour tout X, **définie positive** si $Q(X) > 0$ pour tout X non-nul. On définit symétriquement les notions de forme quadratique **négative** et **définie négative**.

La forme quadratique Q est dite **non-dégénérée** lorsque

$$\left[{}^tYAX = 0, \forall Y \in \mathbb{R}^p\right] \implies X = 0.$$

De façon équivalent, Q est non-dégénérée si $\det A \neq 0$ (c'est à dire lorsque la matrice A est inversible).

Théorème 1. *Soit Q une forme quadratique sur $\mathbb{R}^p$ alors il existe p nombres $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_p$ et une base orthonormée $V_1 \cdots V_p$ telle que*

$$Q(\sum_{i=1}^{p} y_i V_i) \;=\; \sum_{i=1}^{p} \lambda_i y_i^2.$$

De plus, les λ_i sont les valeurs propres de la matrice qui représente Q.

Si on note P la matrice de passage de la base canonique à la base orthonormée $V_1, \cdots V_p$ (c'est à dire que $X = PY$) alors on a

$$^tPAP = \mathrm{diag}(\lambda_1, \cdots \lambda_p).$$

Puisque P est orthogonale on a $^tP = P^{-1}$ et donc on en déduit que la matrice symétrique A est diagonalisable (en base orthonormée) et les λ_i sont les valeurs propres de A.

Corollaire 1. *La forme quadratique Q est non-dégénérée si et seulement si toutes les valeurs propres sont non-nulles. Elle est positive (resp. négative) si toutes les valeurs propres sont positives (resp. négatives). Elle est définie positive (resp. définie négative) si toutes les valeurs propres sont strictement positives (resp. strictement négatives).*

Lorsque Q est définie positive, pour tout X, on a

$$Q(X) \geqslant \lambda_1 \sum_{i=1}^{p} y_i^2 \geqslant \lambda_1 \|X\|^2$$

(où, pour la dernière égalité, on a utilisé le fait que P est orthogonale).

De même, lorsque Q est définie négative, pour tout X on a

$$Q(X) \leqslant -|\lambda_p| \|X\|^2.$$

3.3 Une condition d'ordre 2

On reprend le calcul donnant

$$f(a+tv) - f(a) = \frac{t^2}{2}\left(Q(v) + r(t)\right)$$

où r tend vers 0 quand t tend vers 0 et Q est la forme quadratique représentée par la Hessienne en a dans la base canonique. Dans ce calcul le r dépend en fait de v, lorsque f est au moins C^2, on peut en fait montrer, en un point critique a, la formule de Taylor suivante (obtenue en posant $v = th$ et en montrant que le reste r tend vers 0 uniformément par rapport à la direction) :

$$f(a+h) = f(a) + \frac{1}{2}Q(h) + \|h\|^2 r(h),$$

où r tend vers 0 si h tend vers 0 (on écrit parfois $o(\|h\|^2)$ au lieu de $\|h\|^2 r(h)$).

Remarque : en un point quelconque, pas forcément critique, la formule de Taylor pour une fonction de $\mathbb{R}^p$ dans $\mathbb{R}$ de classe au moins C^2 s'écrit :

$$f(a+h) = f(a) + \sum_{i=1}^{p} h_i \frac{\partial f}{\partial x_i}(a) + \frac{1}{2}\sum_{i,j=1}^{p} h_i h_j \frac{\partial^2 f}{\partial x_i \partial x_j}(a) + o(\|h\|^2).$$

Dans le cas où Q est non-dégénérée, on a le théorème suivant.

Théorème 2. *Soit f définie d'un ouvert $\Omega \subset \mathbb{R}^p$ dans $\mathbb{R}$ une application de classe au moins C^2. Soit a un point critique tel que la hessienne en a est non dégénérée. On est alors dans un des trois cas suivants.*

- *Q est définie positive et a est un minimum local.*
- *Q est définie négative et a est un maximum local.*
- *Q est non-dégénérée mais n'est ni définie positive, ni définie négative. Alors a n'est pas un extremum local. On dit que a est un point-col (ou un point-selle).*

Preuve : On repart de la formule de Taylor-Young,

$$f(a+h)-f(a) \,=\, \frac{1}{2}Q(h) \,+\, o(\|h\|^2).$$

Si Q est définie positive, il existe $\lambda_1 > 0$ tel que, pour tout h on a

$$Q(h) \geqslant \lambda_1 \cdot \|h\|^2$$

Pour h assez petit, on a donc

$$f(a+h)-f(a) \geqslant \frac{\lambda_1}{4}\|h\|^2,$$

ce qui entraîne bien que a est un minimum local. On écrit une preuve similaire dans le cas où Q est définie négative.

Lorsque Q est non-dégénérée mais ni définie positive, ni définie négative, alors Q admet des valeurs propres de signes différents. En prenant v et w des vecteurs propres associés à des valeurs propres de signe différent, on a $Q(v)Q(w) < 0$ ce qui entraîne que a n'est pas un extremum local. □

Remarque : Si Q est dégénérée, on peut trouver v tel que $Q(v) = 0$ et donc, dans cette direction, on ne peut rien dire.

3.4 En pratique

Dans un problème d'optimisation, on peut souvent distinguer plusieurs types d'arguments :

— des arguments utilisant plutôt la continuité et la compacité.

— des arguments de type différentiel (recherche de points critiques, étude de la hessienne),

Très schématiquement, les premiers types d'arguments vont servir à assurer **l'existence** d'un certain type d'extrémum, et les deuxièmes vont permettre de le **localiser**. La méthode est compliquée par le fait que les premiers arguments demandent en général de la compacité, alors que les deuxièmes s'utilisent dans des ouverts. Ainsi, lorsqu'on est sur un compact K, on sait que les extremums existent mais, pour pouvoir utiliser les méthodes différentielles il faut montrer que tel extremum est atteint dans l'intérieur de K. Ce qui se fait typiquement en étudiant ce qui se passe sur le bord de K. Inversement, si la fonction est définie sur un ouvert, on peut mettre des conditions sur le comportement au bord pour *gagner de la compacité*.

Le prototype d'un tel résultat est la proposition suivante.

Proposition 2. *Soit f une fonction définie et continue sur un ensemble ouvert $X \subset \mathbb{R}^p$. On suppose que f est* ***propre*** *c'est à dire qu'elle vérifie la propriété suivante. Pour tout $M > 0$, il existe un compact $K \subset X$ tel que, pour tout $x \in X \setminus K$, $f(x) \geqslant M$. Alors f admet un minimum global sur X.*

Preuve : On commence par fixer un a dans X et on pose $M = f(a) + 1$. Soit K le compact associé à M par l'hypothèse, K n'est pas vide puisqu'il contient le point a. La fonction f est continue sur le compact K et donc admet un minimum x_*. Puisque $a \in K$ on a

$$f(x_*) \leqslant f(a) \,\leqslant\, M.$$

Montrons que x_* est un minimum global, pour tout $x \in X$, ou bien $x \in K$ et alors $f(x) \geqslant f(x_*)$ puisque x_* est un minimum sur K, ou bien $x \notin K$ mais alors $f(x) \geqslant M + 1 \geqslant M \geqslant f(x_*)$. □

On en déduit que lorsque f est propre, on peut trouver le minimum en cherchant les points critiques.

4 Annexe

4.1 Formule de Taylor-Young

Soit $a \in I$, on suppose que f vérifie les deux hypothèses suivantes
- f est $\mathcal{C}^{n-1}$ sur I,
- $f^{(n-1)}$ est dérivable en a.

alors, au voisinage de a on a

$$\boxed{f(x) = f(a) + \sum_{k=1}^{n} \frac{(x-a)^k}{k!} f^{(k)}(a) \,+\, o\left((x-a)^n\right)}$$

Démonstration. On écrit la dérivabilité de $f^{(n-1)}$ en a. Pour $t \in I$, on a

$$f^{(n-1)}(t) \;=\; f^{(n-1)}(a) \,+\, (t-a)f^{(n)}(a) \,+\, o(t-a).$$

On va primitiver cette égalité entre a et x. Remarquons que par définition du o, pour tout ε, il existe η tel que si $|t-a| \leqslant \eta$ alors $|o(t-a)| \leqslant \varepsilon |t-a|$. Par intégration, on a alors, pour $|x-a| \leqslant \eta$,

$$\left| \int_a^x o(t-a)\,dt \right| \leqslant \frac{\varepsilon}{2}|x-a|^2.$$

Ainsi la primitive de $o(t-a)$ est un $o\left((x-a)^2\right)$. On trouve ainsi

$$f^{(n-2)}(x) \;=\; f^{(n-2)}(a) \,+\, (x-a)f^{(n-1)}(a) \,+\, \frac{(x-a)^2}{2!} f^{(n)}(a) \,+\, o\left((x-a)^2\right).$$

Il reste alors à intégrer encore $n-2$ fois pour obtenir le résultat. □

Remarque : Cette formule est en particulier vraie en tout point de I dès que f est $\mathcal{C}^n$ sur I.

4.2 Formule de Taylor-Lagrange

Soit $a \in I$, on suppose que f vérifie les deux hypothèses suivantes
- f est $\mathcal{C}^n$ sur I,
- $f^{(n)}$ est dérivable sur I.

alors, pour tout $a < x \in I$, il existe $c \in]a, x[$ tel que

$$\boxed{f(x) = f(a) + \sum_{k=1}^{n} \frac{(x-a)^k}{k!} f^{(k)}(a) \,+\, \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(c).}$$

Démonstration. On introduit une fonction auxiliaire g, définie sur $[a,x]$ par la formule

$$g(t) := f(x) - f(t) - \sum_{k=1}^{n} \frac{(x-t)^k}{k!} f^{(k)}(t) \; - A(x-t)^{n+1},$$

où la constante A est choisie de telle sorte que $g(a) = 0$. Par hypothèse, la fonction g est continue sur $[a,x]$, dérivable sur $]a,x[$ et vérifie $g(a) = 0 = g(x)$. Le théorème de Rolle s'applique donc et il existe c tel que $g'(c) = 0$. Quand on dérive g, une simplification télescopique intervient de sorte que

$$g'(c) = \frac{(x-c)^n}{n!} f^{(n+1)}(c) - (n+1)A(x-c)^n.$$

On peut alors simplifier par $(x-c)^n$; on obtient $A = \frac{1}{(n+1)!} f^{(n+1)}(c)$. Le résultat s'en déduit en réécrivant $g(a) = 0$. $\square$

Remarque : La même formule est vraie pour $x < a$ (avec la même preuve).