Analyse dans $\mathbf{R}^n$

Chapitre no. 3, Dérivées supérieures, points critiques

Un peu d'algèbre multilinéaire. Soit E et F deux espaces vectoriels de dimensions finies et soit $L(E,F)$ l'espace des applications linéaires de E dans F. Alors $L(E,F)$ est lui-même un espace vectoriel : soit $\varphi, \psi \in L(E,F)$; on définit $\varphi + \psi$ par $(\varphi+\psi)(v) := \varphi(v)+\psi(v)$ et $\lambda\varphi$ par $(\lambda\varphi)(v) := \lambda\varphi(v)$ ($v \in E$, $\lambda \in \mathbf{R}$).

Exemple : $L(\mathbf{R}^n, \mathbf{R}^p)$ s'identifie avec l'espace des $p \times n$ - matrices.

Si E et F sont munis des normes (qu'on note par $||\cdot||$ pour les deux espaces), on peut définir une norme sur $L(E,F)$ par

$$||\varphi|| := \sup_{v\neq 0} \frac{||\varphi(v)||}{||v||}$$

(puisque $L(E,F)$ est de dimension finie, toutes les normes sont équivalentes).

Une application $\varphi : \underbrace{E \times E \times \cdots \times E}_{r} \to F$ est dite *multilinéaire* si elle est linéaire en chaque coefficient : $\varphi(v_1,\ldots,\alpha v_j+\beta w_j,\ldots,v_r) = \alpha\varphi(v_1,\ldots,v_j,\ldots,v_r)+\beta\varphi(v_1,\ldots,w_j,\ldots,v_r)$.

Si $r = 2$, on parle d'une application *bilinéaire.*

φ est *symétrique* si $\varphi(\ldots,v_i,\ldots v_j,\ldots) = \varphi(\ldots,v_j,\ldots,v_i,\ldots)$ pour tout i et j ; elle est *antisymétrique* si $\varphi(\ldots,v_i,\ldots v_j,\ldots) = -\varphi(\ldots,v_j,\ldots,v_i,\ldots)$ pour tout i et j.

Exemple : 1. Le produit scalaire $(v,w) \to v \cdot w$ définit une application bilinéaire symmetrique $\mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$.

2. Le produit vectoriel $(v,w) \to v \wedge w$ definit une application blinéaire antisymétrique $\mathbf{R}^3 \times \mathbf{R}^3 \to \mathbf{R}^3$.

L'espace $L(E, L(E,F))$ s'identifie avec $L^2(E,F)$. En effet, soit $\Phi \in L(E, L(E,F))$. Alors Φ correspond à $\varphi \in L^2(E,F)$ déterminée par la formule : $\varphi(v,w) := \Phi(v)(w)$. Récursivement, on peut identifier $L(E, L(E,\cdots, L(E,F))\cdots)$ (r copies de E) avec $L^r(E,F)$. On remarque que si E et F sont des espace normés, on peut munir $L^r(E,F)$ d'une norme.

Forme quadratique. Un polynôme est *homogène* de degré d si tous ses monômes sont de même degré d. Par exemple $x^3+2xyz+y^2z$ est homogène de degré 3 ; $xy+x^2+2y$ n'est pas homogène. Une forme quadratique définie sur $\mathbf{R}^n$ est un polynôme $Q(x_1,x_2,\ldots,x_n)$ homogène de degré 2. Une tel polynôme s'écrit

$$(1) \quad Q(x) = \sum_{i,j=1}^{n} a_{ij}x_ix_j = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

où les coefficients sont symétriques en i et j : $a_{ij} = a_{ji}\ \forall i, j = 1, \ldots, n$.

Exemple : 1. $n = 2$

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & c \\ c & b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = ax^2 + 2cxy + by^2$$

- c'est la forme générale d'un polynôme homogène de degré deux en deux variables.

2. $n = 3$

$$\begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = ax^2 + dy^2 + fz^2 + 2bxy + 2cxz + 2eyz$$

- c'est la forme générale d'un polynôme homogène de degré deux en trois variables.

Une forme quadratique est équivalente à la donnée d'une application bilinéaire symétrique à valeurs réelles $B : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$: donnée B, soit $Q(x) = B(x, x)$; reciproquement donnée $Q(x)$, soit

$$B(x, y) = \frac{1}{2}\left(Q(x + y) - Q(x) - Q(y)\right)$$

(l'identité de polarisation). Si Q est donnée par (1), alors B est donnée par

$$B(x, y) == \begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Soit

$$A := \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \tag{2}$$

Une forme quadratique Q est *positive (négative) définie* si $x \neq 0 \Rightarrow Q(x) > 0$ (< 0), et *positive (négative) semi-définie* si $x \neq 0 \Rightarrow Q(x) \geq 0$ (≤ 0). Si on écrit $Q(x)$ sous la forme $Q(x) = x^t Ax$ avec matrice A donnée par (2) ; puisque A est symétrique ses valeurs propres sont réelles et il existe une transformation orthogonale des coordonnées par rapport à laquelle A devient diagonale avec ses valeurs propres le long de la diagonale : $\exists M$ ($M^t = M^{-1}$) telle que $M^t AM = D$ avec $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ (λ_i les valeurs propres de A). Il s'ensuit que Q est positive définie si $\lambda_i > 0$ pour tout i, et que Q est positive semi-définie si $\lambda_i \geq 0$ pour tout i.

Application de classe C^r. Soit U un ouvert de $\mathbf{R}^n$ et soit $f : U \to \mathbf{R}^p$ une application differentiable en tout point $x \in U$. Alors, en un point $x \in U$, la dérivée $Df(x)$ est un élement de $L(\mathbf{R}^n, \mathbf{R}^p)$ (un espace normé), c'est à dire $Df : U \to L(\mathbf{R}^n, \mathbf{R}^p)$. Si Df est continue (en tant qu'application entre espaces normés), on dit que f est de classe C^1, ou, qu'elle est simpement C^1.

Puisque $Df(x)(h) = \partial_1 f(x)h_1 + \cdots + \partial_n f(x)h_n$ ($\partial_i f = \partial f/\partial x_i$), il s'ensuit que f est C^1 si et seulement si ses dérivées partielles $\partial f/\partial x_i : U \to \mathbf{R}^p$ sont continues.

On peut maintenant poser la question de si l'application $Df : U \to L(\mathbf{R}^n, \mathbf{R}^p)$ est dérivable. Si c'est le cas en un point $x \in U$, on définit la 2ème dérivée $D^2 f(x) := D(Df)(x) \in L(\mathbf{R}^n, L(\mathbf{R}^n, \mathbf{R}^p) \cong L^2(\mathbf{R}^n, \mathbf{R}^p)$. Explicitement:

$$D^2 f(x)(h,k) = \sum_{i,j=1}^{n} h_i k_j \frac{\partial^2 f}{\partial x_j \partial x_i} \quad (h = (h_1, \ldots, h_n),\, k = (k_1, \ldots, k_n))$$

où

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)$$

Si $D^2 f$ est définie en tout point de U et $D^2 f : U \to L^2(E,F)$ est continue, on dit que f est (de classe) C^2, ce qui revient au fait que les 2èmes dérivées $\dfrac{\partial^2 f}{\partial x_i \partial x_j}$ sont continues.

Lemme de Schwarz (ou le théorème de Clairaut). Soit $f : U \to \mathbf{R}$ (U ouvert dans $\mathbf{R}^n$) de classe C^2, alors

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

Preuve : Il suffit de montrer le théorème pour $n = 2$. Pour $a = (a_1, a_2) \in U$, on pose

$$\begin{aligned}
u(h_1,h_2) &= f(a_1+h_1, a_2+h_2) - f(a_1+h_1, a_2)\\
v(h_1,h_2) &= f(a_1+h_1, a_2+h_2) - f(a_1, a_2+h_2)\\
w(h_1,h_2) &= f(a_1+h_1, a_2+h_2) - f(a_1+h_1, a_2) - f(a_1, a_2+h_2) + f(a_1,a_2)
\end{aligned}$$

définie pour $h = (h_1, h_2)$ assez petit. Par le théorème des acroissements finis

$$\begin{aligned}
w(h_1,h_2) &= u(h_1,h_2) - u(0,h_2) = h_1 \frac{\partial u}{\partial x}(\theta_1(h_1,h_2)h_1, h_2)\\
&= h_1\left(\frac{\partial f}{\partial x}(a_1 + \theta_1(h_1,h_2)h_1, a_2+h_2) - \frac{\partial f}{\partial x}(a_1+\theta_1(h_1,h_2)h_1, a_2)\right)\\
&= h_1 h_2 \frac{\partial^2 f}{\partial x \partial y}(a_1 + \theta_1(h_1,h_2)h_1, a_2 + \theta_2(h_1,h_2)h_2)
\end{aligned}$$

De même,

$$\begin{aligned}
w(h_1,h_2) &= v(h_1,h_2) - v(h_1,0) = h_2\frac{\partial v}{\partial y}(h_1,\theta_3(h_1,h_2)h_2) \\
&= h_2\left(\frac{\partial f}{\partial y}(a_1+h_1,a_2+\theta_3(h_1,h_2)h_2) - \frac{\partial f}{\partial x}(a_1,a_2+\theta_3(h_1,h_2)h_2)\right) \\
&= h_1h_2\frac{\partial^2 f}{\partial y\partial x}(a_1+\theta_4(h_1,h_2)h_1,a_2+\theta_3(h_1,h_2)h_2)
\end{aligned}$$

où les θ_i sont des fonctions qui tendent vers 0 avec h. On laisse $h \to 0$ $(h_1, h_2 \neq 0)$ et la formule s'ensuit. $\square$

<u>Exercice</u> : Etudier la continuité des 2èmes dérivées et la conclusion du lemme de Schwarz pour la fonction

$$f(x,y) = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} & \text{pour } (x,y) \neq (0,0) \\ 0 & \text{pour } (x,y) = (0,0) \end{cases}$$

Par récurrence, pour $x \in U$, on définit $D^r f(x) \in L^r(\mathbf{R}^n, \mathbf{R}^p)$, par $D^r f(x) := D(D^{r-1}f)(x)$. On dit que $f : U \to \mathbf{R}^p$ est (de classe) C^r si toutes les dérivées jusqu'à l'ordre r existent et sont continues, ce qui revient au fait que toutes les dérivées partielles jusquà l'ordre r sont continues. Si f est C^r pour tout $r = 1, 2, \ldots$ on dit que f est C^∞.

<u>Exemple</u>. Soit $f : \mathbf{R} \to \mathbf{R}$ définie par $f(x) = x^{7/3}$. Alors

$$f'(x) = \tfrac{7}{3}x^{4/3}, \qquad f''(x) = \tfrac{28}{9}x^{1/3}, \qquad f'''(x) = \tfrac{28}{27}x^{-2/3}$$

On voit alors que f est une fonction C^2 mais pas C^3.

<u>Formule de Taylor-Young)</u>. Soit $U \subset \mathbf{R}^n$ ouvert et soit $f : U \to \mathbf{R}^p$ de classe C^{r+1}. On suppose que le segment droit $[x, x+h] \subset U$. Alors

$$f(x+h) = f(x) + Df(x)(h) + \frac{1}{2!}D^2f(x)(h,h) + \cdots + \frac{1}{r!}D^rf(x)(\underbrace{h,h,\ldots,h}_{r}) + R(x,h)$$

où le reste $R(x,h)$ est $o(||h||^r)$. Explicitement

$$R(x,h) = \left(\int_0^1 \frac{1-\zeta)^r}{r!}D^{r+1}f(x+\zeta h)\mathrm{d}\zeta\right)(\underbrace{h,h,\ldots,h}_{r+1})$$

Si on n'est par concerné par l'expression explicite du reste, on peut affirmé que si f est r fois dérivable en x (toutes les dérivées jusqu'à l'ordre r existent), alors on a la formule de Taylor-Lagrange:

$$f(x+h) = f(x) + Df(x)(h) + \frac{1}{2!}D^2f(x)(h,h) + \cdots + \frac{1}{r!}D^rf(x)(\underbrace{h,h,\ldots,h}_{r}) + o(||h||^r)$$

Les premiers termes s'exprîment :

$$f(x+h) = f(x) + h_1\frac{\partial f}{\partial x_1}(x) + \cdots + h_n\frac{\partial f}{\partial x_n}(x) + \frac{1}{2!}\sum_{i,j=1}^{n} h_i h_j \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \cdots$$

(A noter que $f = \begin{pmatrix} f_1 \\ \vdots \\ f_p \end{pmatrix}$ et que cette expression est valable pour chaque composante f_k de f). Bien evidemment, si $f :]a,b[\to \mathbf{R}$, on obtient la formule de Taylor-Lagrange pour un fonction d'une variable réelle.

Points critiques. Soit $f : U \to \mathbf{R}^p$ ($U \subset \mathbf{R}^n$ ouvert) une application dérivable en chaque point $x \in U$. On dit que $a \in U$ est un *point critique* (ou singulier) si $Df(a) : \mathbf{R}^n \to \mathbf{R}^p$ n'est pas de rang maximal ($= \min\{n,p\}$). Dans la suite, on va supposer que $p = 1$ et que f est une fonction. Dans ce cas, a est un point critique si $Df(a) = 0$, ce qui revient au fait que $\frac{\partial f}{\partial x_i}(a) = 0$ pour tout $i = 1, \ldots, n$.

Un point $a \in U$ est un *maximum local* pour f s'il existe un voisinage $V \subset U$ de a tel que $f(x) \leq f(a)$ pour tout $x \in V$; on dit que le maximum local est *strict* si $f(x) < f(a)$ pour tout $x \in V,\ x \neq a$. On a des définitions analogues pour *minimum local* et *minimum local strict*. Le point $a \in U$ s'appelle un *maximum global* pour f si $f(x) \leq f(a)$ pour tout $x \in U$, et un *maximum global strict* si $f(x) < f(a)$ pour tout $x \in U,\ x \neq a$. Un point maximal ou minimal s'appelle un *point extremal*, ou un *extremum*.

On voit que si f est dérivable en un maximum local $a \in U$, alors $Df(a) = 0$, c'est à dire a est un point critique (on raisonne par l'absurde : si $Df(a) \neq 0$, on a $f(a+h) = f(a) + Df(a)h + o(||h||)$...).

Un point critique qui est ni un maximum local ni un minimum local s'appelle un *point de selle*.

Pour trouver les extrema d'une fonction $f : U \to \mathbf{R}$ dérivable, on calcule d'abord les points $a \in U$ où $Df(a) = 0$ (les dérivées partielles s'annullent), puis on étudie le prochaine terme non-nul dans le développement de Taylor-Lagrange :

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

Il s'ensuit que si au moins une dérivée d'ordre 2 est non-nulle en a, la nature du point critique dépend de la forme quadratique: $Q(h) = h^t H_f(a) h$, où $H_h(a)$ est la *matrice* *hessienne* :

$$H_f(a) := \begin{pmatrix} \frac{\partial^2 f}{\partial {x_1}^2}(a) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(a) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(a) & \frac{\partial^2 f}{\partial {x_2}^2}(a) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(a) & \frac{\partial^2 f}{\partial x_n \partial x_2}(a) & \cdots & \frac{\partial^2 f}{\partial {x_n}^2}(a) \end{pmatrix}$$

Si f est de classe C^2, cette matrice est symétrique et on peut appliquer la théorie des formes quadratiques :

• si les valeurs propres de $H_f(a)$ sont toutes strictement positives, alors a est un minimum local strict ;

• si les valeurs propres de $H_f(a)$ sont toutes ≥ 0, alors a est un minimum local ;

• si les valeurs propres de $H_f(a)$ sont toutes strictement negatives, alors a est un maximum local strict ;

• si les valeurs propres de $H_f(a)$ sont toutes ≤ 0, alors a est un maximum local ;

• si les valeurs propres de $H_f(a)$ ont des valeurs negatives et positives, alors a est un point de selle.

Remarque : Si $\det H_f(a) \neq 0$ alors toutes les valeurs propres sont non-nulles.

Lorsque $n = 2$, l'étude devient assez simple. En écrivant $f_{ij} = \partial^2 f/\partial x_i \partial x_j$, on a (pour $h_2 \neq 0$)

$$\begin{aligned} h^t H_f(a) h &= {h_1}^2 f_{11}(a) + 2h_1 h_2 f_{12}(a) + {h_2}^2 f_{22}(a) \\ &= {h_2}^2 \left(t^2 f_{11}(a) + 2t f_{12}(a) + f_{22}(a)\right) \end{aligned}$$

où $t = h_1/h_2$. Alors la fonction $t \to t^2 f_{11}(a) + 2t f_{12}(a) + f_{22}(a)$ est une quadratique en t ; elle n'a aucune racine réelle ssi le discriminant $4(f_{12}(a)^2 - f_{11}(a) f_{22}(a)) < 0$, c'est à dire ssi $\det H_f(a) > 0$. Dans ce cas, la fonction est toujours positive si $f_{11}(a) > 0$ (ou $f_{22}(a) > 0$) (a est un minimum strict), ou négative si $f_{11}(a) < 0$ (ou $f_{22}(a) < 0$) (a est un maximum strict). Si $\det H_f(a) = 0$ et au moins une des 2èmes dérivées est non-nulles, alors il y a une racine réelle unique, et on a soit un maximum ou minumim non-strict. Si $\det H_f(a) < 0$, le point critique a est un point de selle. Pour resumer :

• $\det H_f(a) > 0 \Rightarrow$ point extremal ; soit un minimum ($f_{11}(a) > 0$), soit un maximum ($f_{11} < 0$) ;

• $\det H_f(a) < 0 \Rightarrow$ point de selle.

Exemples : 1. Soit $f(a_1, \ldots, x_n) = {x_1}^2 + \cdots + {x_n}^2$. Alors l'origine $x = 0$ est un point critique. Puisque $H_f(0) = \mathrm{diag}(2, 2, \ldots, 2)$ est positive définie, on voit qu'il s'agit un minimum local.

2. Soit $f(x,y) = x^2y - 2x - 2y$. Alors $f_x = 2xy - 2$ et $f_y = x^2 - 1$. On voit que les points critiques sont $(1,1)$ et $(-1,-1)$. La matrice hessienne est donnée par

$$H_f(x,y) = \begin{pmatrix} 2y & 2x \\ 2x & 0 \end{pmatrix}$$

Quelque soit $x \neq 0$, on voit que $H_f(x,y) = -2x^2 < 0$, d'où les deux points critiques sont des points de selle.

3. Soit $f(x,y,z) = \frac{1}{2}(x^2 + y^2 + z^2 - 2xy - 2xz - 2yz)$. On voit que le seul point critique est l'origine (exercice). La matrice hessienne est donnée par

$$H_f = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$$

dont le déterminant est -4. Il est possible que soit toutes les valeurs propres sont négatives (maximum), soit deux sont positives et une est négative (point de selle). On calcule les valeurs propres :

$$\det \begin{pmatrix} 1-\lambda & -1 & -1 \\ -1 & 1-\lambda & -1 \\ -1 & -1 & 1-\lambda \end{pmatrix} = (1-\lambda)(\lambda-2)(\lambda+2)$$

d'où les valeurs propres sont $1, 2$ et 2. Il s'ensuit que l'origine est un point de selle.

Références 1. J. Dieudonné, Foundations of Modern Analysis, Academic Press 1969.
2. J. Lelong-Ferrand et J. M. Arnaudiès, Cours de Mathématiques Tome 2 : Analyse, 4ème édition, Dunod Université 1977.