Chapitre 2

Calcul différentiel dans $\mathbb{R}^n$

Nous allons donner d'abord les principales notions de "dérivation" d'une fonction de $\mathbb{R}^n \longrightarrow \mathbb{R}$. La norme utilisée désormais sera la norme euclidienne sauf mention contraire.

2.1 Notions de différentiabilité

Définition 2.1.1 *Soit U un ouvert de $\mathbb{R}^n$. Une fonction $f : U \longrightarrow \mathbb{R}$ est dite différentiable en un point $a \in U$ s'il existe une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}$, notée df_a, telle que*

$$\lim_{\|h\|\to 0} \frac{f(a+h) - f(a) - df_a(h)}{\|h\|} = 0 \tag{2.1}$$

L'application linéaire df_a s'appelle différentielle de f au point a, ou bien application linéaire tangente en a. La différentiabilité de f sur U signifie sa différentiabilité en tout point de U. En notant la fraction précédente par $\varepsilon(h)$, on peut réécrire (2.1) comme suit :

$$f(a+h) = f(a) + df_a(h) + \|h\|\varepsilon(h) \tag{2.2}$$

où $\lim\limits_{h\to 0} \varepsilon(h) = 0$. Rappelons que la linéarité de df_a s'écrit ainsi

$$df_a(h) = df_a(h_1, h_2, \ldots, h_n) = \lambda_1 h_1 + \lambda_2 h_2 + \cdots + \lambda_n h_n = \langle \lambda, h \rangle$$

où les λ_i sont des constantes qui dépendent de a et $\langle ., . \rangle$ est le produit scalaire usuel de $\mathbb{R}^n$.

Exemple 2.1.2 *Examinons un exemple simple. Déterminons la différentielle de $f(x, y) = xy$, définie sur $\mathbb{R}^2$, en un point (a, b). Il est facile de voir que*

$$f(a + h_1, b + h_2) = (a + h_1)(b + h_2) = ab + bh_1 + ah_2 + h_1h_2$$

Deux candidats à la formule (2.2) s'offrent aisément : $df_{(a,b)}(h_1, h_2) = bh_1 + ah_2$ et $\varepsilon(h_1, h_2) = \dfrac{h_1h_2}{\sqrt{h_1^2 + h_2^2}}$ si on choisit de travailler avec la norme euclidienne. Comme $\dfrac{|h_1h_2|}{h_1^2 + h_2^2} \leq \dfrac{1}{2}$, alors $|\varepsilon(h_1, h_2)| \leq \dfrac{1}{\sqrt{2}}\sqrt{|h_1h_2|}$ qui assure que $\lim\limits_{h\to 0} \varepsilon(h) = 0$.

Définition 2.1.3 *(Dérivation dans le sens d'un vecteur)*
Soit $f : U \longrightarrow \mathbb{R}$ et $a \in U$. Fixons un vecteur v dans $\mathbb{R}^n$. On appelle dérivée de f au point a dans le sens du vecteur v la limite suivante, si elle existe,

$$\lim_{t \to 0} \frac{f(a+tv) - f(a)}{t} = (D_v f)(a) \tag{2.3}$$

(remarquez que t est un réel).

Exemple 2.1.4 *Examinons le cas de $f(x,y) = x^2 - y^2$ avec $a = (1,-1)$ et $v = (1,1)$. On a*

$$f(1+t, -1+t) = (1+t)^2 - (-1+t)^2 = 4t \quad et \quad f(a) = 0$$

d'où $(D_v f)(a) = 4$.

Définition 2.1.5 *(Dérivation partielle)*
Soit $e_j = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$ le jème vecteur de la base canonique de $\mathbb{R}^n$. On appelle dérivée partielle de f par rapport à x_j la dérivée de f dans le sens du vecteur e_j i.e,

$$\frac{\partial f}{\partial x_j}(a) = (D_{e_j} f)(a) = \lim_{t \to 0} \frac{f(a_1, \ldots, a_j + t, \ldots, a_n) - f(a)}{t} \tag{2.4}$$

On appelle gradient de f le vecteur défini par

$$(\nabla f)(x) = (\vec{\text{grad}} f)(x) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{pmatrix}$$

Les deux notations sont consacrées. Le symbole ∇ se lit "nabla".

Exemple 2.1.6 *Considérons $f(x,y,z) = xy + xz + yz$. Alors*

$$\frac{\partial f}{\partial x}(x,y,z) = y + z \quad , \quad \frac{\partial f}{\partial y}(x,y,z) = x + z \quad , \quad \frac{\partial f}{\partial z}(x,y,z) = x + y$$

Proposition 2.1.7 *Si f est différentiable au point a, alors $\forall j = 1, \ldots, n$; $\dfrac{\partial f}{\partial x_j}(a)$ existe, de plus*

$$df_a(h) = \sum_{j=1}^{n} h_j \frac{\partial f}{\partial x_j}(a) = \langle h, \nabla f(a) \rangle \tag{2.5}$$

Démonstration : On a par hypothèse

$$f(a+h) = f(a) + \underbrace{\sum_{j=1}^{n} \lambda_j h_j}_{df_a(h)} + \|h\| \varepsilon(h)$$

Prenons alors $h = te_j$, alors

$$f(a + te_j) = f(a) + t\lambda_j + |t|\varepsilon(te_j)$$

Et d'après (2.4) on aura $\dfrac{\partial f}{\partial x_j}(a) = \lambda_j$. ■

Remarque 2.1.8 *La réciproque est en général fausse. Examinons l'exemple suivant :*

$$f(x,y) = \begin{cases} \dfrac{xy}{\sqrt{x^2+y^2}} & si \quad (x,y) \neq (0,0) \\ 0 & si \quad (x,y) = (0,0) \end{cases}$$

Il est facile de voir que $\dfrac{\partial f}{\partial x}(0,0) = 0 = \dfrac{\partial f}{\partial y}(0,0)$. *Si* f *était différentiable, on aurait forcément* $df_{(0,0)}(h) = 0$. *Or*

$$\frac{f(h_1,h_2) - f(0,0) - df_{(0,0)}(h)}{\|h\|} = \frac{h_1h_2}{h_1^2 + h_2^2}$$

ne tend manifestement pas vers 0 *quand* h *tend vers* $(0,0)$*, elle n'a d'ailleurs pas de limite.*

Proposition 2.1.9 *Si* f *est différentiable en* a*, alors elle est continue en ce point. De plus elle est dérivable dans le sens de n'importe quel vecteur* v *et on a*

$$(D_v f)(a) = \langle (\nabla f)(a), v \rangle$$

Démonstration : La différentiabilité en a s'écrit

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

Ce qui donne aisément $\lim\limits_{h \to 0} f(a+h) = f(a)$. Pour la deuxième affirmation, il suffit de prendre $h = tv$. ■

Dans la remarque précédente nous avons vu que l'existence des dérivées partielles premières n'implique pas en général la différentiabilité. Dans le prochain théorème, nous allons rajouter l'hypothèse de continuité de ces dérivées partielles afin d'obtenir la différentiabilité.

Théorème 2 *Si toutes les dérivées partielles* $\dfrac{\partial f}{\partial x_j}(x)$ *existent dans un voisinage de* a *et sont continues en* a*, alors* f *est différentiable au point* a*.*

Démonstration : Dans cette démonstration, on utilisera la norme $\|x\|_1 = \sum\limits_{i=1}^{n} |x_i|$ pour sa commodité. Posons

$$g(x) = f(x) - \sum_{i=1}^{n} x_i \frac{\partial f}{\partial x_j}(a)$$

Nous allons montrer que $g(x+h)-g(a) = \|h\|_1\varepsilon(h)$, c'est-à-dire que g est différentiable en a de différentielle nulle, ce qui donnera le résultat escompté pour f. On a d'abord

$$\frac{\partial g}{\partial x_j}(x) = \frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(a)$$

La continuité des dérivées partielles prise comme hypothèse peut s'exprimer ainsi :

$$\forall \varepsilon > 0 \quad \exists \alpha_\varepsilon > 0 \quad \forall x \in B(a, \alpha_\varepsilon) \quad \forall j = 1, \dots, n \qquad |\frac{\partial g}{\partial x_j}(x)| \leq \varepsilon.$$

Posons

$$\begin{aligned} y_0 &= (a_1, a_2, \dots, a_n) = a \\ y_1 &= (x_1, a_2, \dots, a_n) \\ y_2 &= (x_1, x_2, a_3, \dots, a_n) \\ \vdots &\quad \vdots \\ y_n &= (x_1, x_2, \dots, x_n) = x \end{aligned}$$

Définissons n fonctions auxiliaires d'une variable par

$$\begin{array}{llll} g_k : & [a_k, x_k] & \longrightarrow \mathbb{R} & \\ & t & \longrightarrow & g_k(t) = g(x_1, x_2, \dots, x_{k-1}, t, a_{k-1}, \dots, a_n) \end{array}$$

L'intervalle de définition peut tout aussi être $[x_k, a_k]$. Remarquer que $g_1(t) = g(t, a_2, \dots, a_n)$ et $g_n(t) = g(x_1, x_2, \dots, x_{n-1}, t)$. D'après les données, les fonctions g_k vérifient les hypothèses du théorème classique des accroissements finis (à une variable). Donc on peut écrire

$$g_k(x_k) - g_k(a_k) = (x_k - a_k)g_k'(c_k) = (x_k - a_k)\frac{\partial g}{\partial x_k}(x_1, x_2, \dots, x_{k-1}, c_k, a_{k-1}, \dots, a_n)$$

$$\Longrightarrow |g_k(x_k) - g_k(a_k)| \leq \varepsilon |x_k - a_k|$$

On a par ailleurs $g_k(x_k) = g(y_k)$ et $g_k(a_k) = g(y_{k-1})$. De là

$$g(x) - g(a) = g(y_n) - g(y_0) = \sum_{k=1}^{n} g(y_k) - g(y_{k-1})$$

$$\Longrightarrow |g(x) - g(a)| \leq \varepsilon \|x - a\|_1$$

ou encore

$$|g(a+h) - g(a)| \leq \varepsilon \|h\|_1$$

c'est-à-dire

$$\lim_{\|h\|_1 \to 0} \frac{g(a+h) - g(a)}{\|h\|_1} = 0$$

■

2.2 Dérivées partielles d'ordre supérieur

Les dérivées partielles premières $\dfrac{\partial f}{\partial x_j}$ sont à leur tour des fonctions de $\mathbb{R}^n$ dans $\mathbb{R}$, susceptibles d'être dérivables. Leurs dérivées partielles, notées $\dfrac{\partial}{\partial x_i}\left(\dfrac{\partial f}{\partial x_j}\right) := \dfrac{\partial^2 f}{\partial x_i \partial x_j}$, s'appellent dérivées partielles secondes, et ainsi de suite. Une question naturelle se pose alors : peut-on avoir $\dfrac{\partial^2 f}{\partial x_i \partial x_j} = \dfrac{\partial^2 f}{\partial x_j \partial x_i}$ pour $i \neq j$? Avec seulement la dérivabilité à l'ordre 2 la réponse est non en général (voir exercice 4 de la série de T.D N°2). Mais avec la dérivabilité à l'ordre 2 et la continuité de ces dérivées deuxièmes la réponse est oui, c'est le théorème de Hermann Schwarz (1843-1921). On dira qu'une fonction est de classe C^p si elle admet des dérivées partielles continues jusqu'à l'ordre p inclu. Sans le théorème de Schwarz, le nombre de dérivées d'ordre p est n^p. Avec le théorème de Schwarz, ce nombre se réduit à $\dfrac{(n+p-1)!}{p!(n-1)!}$. Nous allons énoncer et démontrer le théorème de Schwarz pour des fonctions de deux variables, cela suffit pour le cas général.

Théorème 3 *(de Schwarz)*
Soit U un ouvert de $\mathbb{R}^2$, $a \in U$ et $f : U \longrightarrow \mathbb{R}$ telle que $\dfrac{\partial^2 f}{\partial x \partial y}(x,y)$ et $\dfrac{\partial^2 f}{\partial y \partial x}(x,y)$ existent dans U et soient continues au point a. Alors

$$\frac{\partial^2 f}{\partial x \partial y}(a) = \frac{\partial^2 f}{\partial y \partial x}(a)$$

Démonstration : Soit $h = (h_1, h_2)$ de norme petite. Posons

$$A(h) = f(a_1 + h_1, a_2 + h_2) - f(a_1 + h_1, a_2) - f(a_1, a_2 + h_2) + f(a_1, a_2)$$

Considérons la fonction d'une variable $\varphi(t) = f(t, a_2 + h_2) - f(t, a_2)$. Alors par application du théorème des accroissements finis on a

$$A(h) = \varphi(a_1 + h_1) - \varphi(a_1) = h_1 \varphi'(a_1 + \theta_1 h_1), \quad \text{avec} \quad 0 < \theta_1 < 1$$

et

$$\varphi'(a_1 + \theta_1 h_1) = \frac{\partial f}{\partial x}(a_1 + \theta_1 h_1, a_2 + h_2) - \frac{\partial f}{\partial x}(a_1 + \theta_1 h_1, a_2)$$

On peut appliquer encore une fois le théorème des accroissements finis pour la deuxième variable de la fonction $\dfrac{\partial f}{\partial x}(a_1 + \theta_1 h_1, .)$. Cela donne

$$A(h) = h_1 h_2 \frac{\partial^2}{\partial y \partial x}(a_1 + \theta_1 h_1, a_2 + \theta_2 h_2), \quad \text{avec} \quad 0 < \theta_2 < 1$$

On considère ensuite la fonction $\psi(t) = f(a_1 + h_1, t) - f(a_1, t)$ et on refait le même travail. On obtient

$$A(h) = h_1 h_2 \frac{\partial^2}{\partial x \partial y}(a_1 + \tilde{\theta}_1 h_1, a_2 + \tilde{\theta}_2 h_2)$$

et donc

$$\frac{\partial^2}{\partial y \partial x}(a_1 + \theta_1 h_1, a_2 + \theta_2 h_2) = \frac{\partial^2}{\partial x \partial y}(a_1 + \tilde{\theta}_1 h_1, a_2 + \tilde{\theta}_2 h_2)$$

On fait tendre alors h vers 0 et la continuité permet de conclure. ■

Exemple 2.2.1 *Considérons la fonction*

$$f(x,y) = \begin{cases} y^2 \sin(\frac{x}{y}) & si \;\; y \neq 0 \\ 0 & si \;\; y = 0 \end{cases}$$

Il est facile de voir que f est continue partout. Calculons les dérivées partielles premières.

$$\frac{\partial f}{\partial x}(x,y) = \begin{cases} y \cos(\frac{x}{y}) & si \;\; y \neq 0 \\ 0 & si \;\; y = 0 \end{cases}$$

et

$$\frac{\partial f}{\partial y}(x,y) = \begin{cases} 2y \sin(\frac{x}{y}) - x\cos(\frac{x}{y}) & si \;\; y \neq 0 \\ 0 & si \;\; y = 0 \end{cases}$$

Le calcul des dérivées premières en $(x,0)$ se fait en appliquant la définition (voir (2.4)). Toujours selon la définition on a

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = \lim_{t \to 0} \frac{\frac{\partial f}{\partial y}(t,0) - \frac{\partial f}{\partial y}(0,0)}{t} = 0$$

et

$$\frac{\partial^2 f}{\partial y \partial x}(0,0) = \lim_{t \to 0} \frac{\frac{\partial f}{\partial x}(0,t) - \frac{\partial f}{\partial x}(0,0)}{t} = 1$$

D'autre part

$$\frac{\partial^2 f}{\partial y \partial x}(x,y) = \cos(\frac{x}{y}) + \frac{x}{y}\sin(\frac{x}{y}) \;\; si \;\; y \neq 0$$

Il est évident que $\frac{\partial^2 f}{\partial y \partial x}$ n'est pas continue en $(0,0)$ puisqu'elle n'a même pas de limite en ce point. Il n'est pas nécessaire d'examiner l'autre dérivée mixte. Ceci montre l'importance du théorème de Schwarz.

2.3 Formules de Taylor

Nous allons commencer par une formule de dérivation très utile.

Proposition 2.3.1 *Soit $f : U \longrightarrow \mathbb{R}^n$ différentiable et $\varphi : I \longrightarrow U$ une application d'une variable t continûment dérivable sur un intervalle ouvert réel I*

$$\varphi(t) = (\varphi_1(t), \varphi_2(t), \dots, \varphi_n(t))$$

Posons $\psi(t) = f(\varphi(t))$. Alors

$$\psi'(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\varphi(t)).\varphi_i'(t) \tag{2.6}$$

Démonstration : La formule de Taylor-Lagrange appliquée à l'ordre 1 pour chaque φ_i permet d'écrire pour t fixé

$$\varphi(t+\tau) = \varphi(t) + \tau\varphi'(t) + \tau\theta_t(\tau)$$

où $\lim_{\tau\to 0} \theta_t(\tau) = 0$. Maintenant la différentiabilité de f donne

$$\psi(t+\tau) = \psi(t) + \langle (\nabla f)(\varphi(t)), \tau\varphi'(t) + \tau\theta_t(\tau)\rangle + \|\tau\varphi'(t) + \tau\theta_t(\tau)\| \varepsilon(\tau\varphi'(t) + \tau\theta_t(\tau))$$

puis

$$\frac{\psi(t+\tau) - \psi(t)}{\tau} = \langle (\nabla f)(\varphi(t)), \varphi'(t) + \theta_t(\tau)\rangle + \frac{|\tau|}{\tau}\|\varphi'(t) + \theta_t(\tau)\| \varepsilon(\tau\varphi'(t) + \tau\theta_t(\tau))$$

En passant à la limite quand τ tend vers 0, on obtient la formule annoncée. ∎

Maintenant on est en mesure de donner les formules de Taylor avec reste de Lagrange et avec reste intégral.

Théorème 4 *(Formule de Taylor avec reste de Lagrange)*
Soit $U \subset \mathbb{R}^n$ un ouvert et $f : U \longrightarrow \mathbb{R}$ une application de classe C^{p+1}. Alors, pour h de norme petite, on a

$$f(x+h) = \sum_{k=0}^{p} \frac{1}{k!} \left(\sum_{i_1,\dots,i_k=1}^{n} h_{i_1}\dots h_{i_k} \frac{\partial^k f(x)}{\partial x_{i_1}\dots\partial x_{i_k}} \right) + \\ + \frac{1}{(p+1)!} \sum_{i_1,\dots,i_{p+1}=1}^{n} h_{i_1}\dots h_{i_{p+1}} \frac{\partial^{p+1} f(x+\theta h)}{\partial x_{i_1}\dots\partial x_{i_{p+1}}} \tag{2.7}$$

avec $0 < \theta < 1$.

Démonstration : Définissons une fonction d'une variable t par $\eta(t) = f(x+th)$. Pour $\|h\|$ petite, on peut prendre t dans $[0,1]$. On écrit tout simplement la formule de Taylor avec reste de Lagrange à l'ordre p pour la fonction η :

$$\eta(t) = \sum_{k=0}^{p} \frac{\eta^{(k)}(0)}{k!} t^k + \frac{t^{p+1}}{(p+1)!} \eta^{(p+1)}(\theta t)$$

La formule annoncée s'obtient en posant $t = 1$. Reste à évaluer les dérivées de η. On utilise la formule (2.6) :

$$\begin{aligned}
\eta'(t) &= \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(x + th) \\
\eta''(t) &= \sum_{i,j=1}^{n} h_i h_j \frac{\partial^2 f}{\partial x_i \partial x_j}(x + th) \\
&\vdots \\
\eta^{(k)}(t) &= \sum_{i_1,\dots,i_k=1}^{n} h_{i_1} \dots h_{i_k} \frac{\partial^k f}{\partial x_{i_1} \dots \partial x_{i_k}}(x + th)
\end{aligned}$$

Il suffit de prendre $t = 0$ quand $k = 0, 1, \dots, p$, puis $t = \theta$ pour $k = p + 1$ dans les expressions précédentes. ■

Théorème 5 *(Formule de Taylor avec reste intégral)*
Sous les mêmes hypothèses que le théorème précédent, on a

$$\begin{aligned}
f(x+h) = \sum_{k=0}^{p} \frac{1}{k!} \left(\sum_{i_1,\dots,i_k=1}^{n} h_{i_1} \dots h_{i_k} \frac{\partial^k f(x)}{\partial x_{i_1} \dots \partial x_{i_k}} \right) + \\
+ \frac{1}{(p+1)!} \sum_{i_1,\dots,i_{p+1}=1}^{n} h_{i_1} \dots h_{i_{p+1}} \int_0^1 (1-\xi)^p \frac{\partial^{p+1} f(x+\xi h)}{\partial x_{i_1} \dots \partial x_{i_{p+1}}} \, d\xi
\end{aligned} \tag{2.8}$$

La démonstration est pareille que la précédente en ce sens qu'elle utilise la formule de Taylor avec reste intégral pour la fonction d'une variable $\eta(.)$. A titre indicatif, nous allons écrire la formule de Taylor-Lagrange à l'ordre 2 pour des fonctions de deux et trois variables. Pour le cas $n = 2, p = 2$, on a

$$\begin{aligned}
f(x+h_1, y+h_2) = f(x,y) + h_1 \frac{\partial f(x,y)}{\partial x} + h_2 \frac{\partial f(x,y)}{\partial y} + \\
+ \frac{1}{2} \left\{ h_1^2 \frac{\partial^2 f(x,y)}{\partial x^2} + 2h_1 h_2 \frac{\partial^2 f(x,y)}{\partial x \partial y} + h_2^2 \frac{\partial^2 f(x,y)}{\partial y^2} \right\} + \mathcal{R}_{Lag}(h)
\end{aligned} \tag{2.9}$$

avec

$$\begin{aligned}
\mathcal{R}_{Lag}(h) = \frac{1}{6} \Bigg\{ h_1^3 \frac{\partial^3 f(x+\theta h_1, y+\theta h_2)}{\partial x^3} + 3h_1^2 h_2 \frac{\partial^3 f(x+\theta h_1, y+\theta h_2)}{\partial x^2 \partial y} + \\
+3h_1 h_2^2 \frac{\partial^3 f(x+\theta h_1, y+\theta h_2)}{\partial x \partial y^2} + h_2^3 \frac{\partial^3 f(x+\theta h_1, y+\theta h_2)}{\partial y^3} \Bigg\}
\end{aligned}$$

Notez que nous avons utilisé le théorème de Schwarz. Notez aussi que la bornitude des dérivées troisièmes implique que $|\mathcal{R}_{Lag}(h)| \leq M \|h\|^3$.

Pour $n = 3, p = 2$, on aura

$$\begin{aligned} f(x+h_1, y+h_2, z+h_3) &= f(x,y,z) + h_1\frac{\partial f(x,y,z)}{\partial x} + h_2\frac{\partial f(x,y,z)}{\partial y} + h_3\frac{\partial f(x,y,z)}{\partial z} \\ &+ \frac{1}{2}\left\{h_1^2\frac{\partial^2 f(x,y,z)}{\partial x^2} + h_2^2\frac{\partial^2 f(x,y,z)}{\partial y^2} + h_3^2\frac{\partial^2 f(x,y,z)}{\partial z^2} + \right. \\ &\left. +2h_1h_2\frac{\partial^2 f(x,y,z)}{\partial x\partial y} + 2h_1h_3\frac{\partial^2 f(x,y,z)}{\partial x\partial z} + 2h_2h_3\frac{\partial^2 f(x,y,z)}{\partial y\partial z}\right\} + \mathcal{R}_{Lag}(h) \end{aligned} \tag{2.10}$$

On peut condenser les écritures précédentes en introduisant la notion de *hessienne.*

Définition 2.3.2 *On appelle hessienne la matrice des dérivées secondes définie par*

$$Hess_f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Remarquons que grâce au théorème de Schwarz la matrice hessienne est symétrique (par rapport à sa diagonale principale). On peut réécrire à présent la formule de Taylor-Lagrange à l'ordre 2 pour n quelconque d'ailleurs ainsi :

$$f(x+h) = f(x) + \langle \nabla f(x), h\rangle + \frac{1}{2}\langle Hess_f(x)h, h\rangle + \mathcal{R}_{Lag}(h) \tag{2.11}$$

2.4 Différentiabilité de fonctions vectorielles

On entend par fonction vectorielle, une fonction à valeurs dans $\mathbb{R}^p$ i.e,

$$\begin{array}{rccl} f: & \mathbb{R}^n & \longrightarrow & \mathbb{R}^p \\ & x & \longrightarrow & f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_p(x) \end{pmatrix} \end{array}$$

.

Définition 2.4.1 *Une fonction $f : U \longrightarrow \mathbb{R}^p$ est dite différentiable en $a \in U$ s'il existe une application linéaire, notée $df_a : \mathbb{R}^n \longrightarrow \mathbb{R}^p$, telle que*

$$\lim_{\|h\|\to 0} \frac{\|f(a+h) - f(a) - df_a(h)\|_{\mathbb{R}^p}}{\|h\|_{\mathbb{R}^n}} = 0 \tag{2.12}$$

On omettra les indices dans les normes quand il n'y a pas de confusion. On pourra écrire aussi la différentiabilité comme suit :

$$f(a+h) = f(a) + df_a(h) + \|h\|\varepsilon(h) \tag{2.13}$$

où $\varepsilon(.)$ est une fonction vectorielle telle que $\lim\limits_{\|h\|\to 0} \|\varepsilon(h)\| = 0$.

La linéarité de df_a s'exprime par (se référer au cours d'Algèbre Linéaire)

$$df_a(h) = \begin{pmatrix} \langle \nabla f_1(a), h\rangle \\ \langle \nabla f_2(a), h\rangle \\ \vdots \\ \langle \nabla f_p(a), h\rangle \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1(a)}{\partial x_1} & \frac{\partial f_1(a)}{\partial x_2} & \cdots & \frac{\partial f_1(a)}{\partial x_n} \\ \frac{\partial f_2(a)}{\partial x_1} & \frac{\partial f_2(a)}{\partial x_2} & \cdots & \frac{\partial f_2(a)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_p(a)}{\partial x_1} & \frac{\partial f_p(a)}{\partial x_2} & \cdots & \frac{\partial f_p(a)}{\partial x_n} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{pmatrix} \tag{2.14}$$

La matrice $p \times n$ précédente représentant df_a s'appelle *jacobienne* de f en a. Elle est aussi notée

$$J_f(a) = \left(\frac{\partial f_i(a)}{\partial x_j}\right)_{1\le i\le p\ ,\ 1\le j\le n} \tag{2.15}$$

Exemple 2.4.2 *Si* $f(x,y,z) = \begin{pmatrix} xy+xz+yz \\ xyz \end{pmatrix}$ *alors*

$$J_f(x,y,z) = \begin{pmatrix} y+z & x+z & x+y \\ yz & xz & xy \end{pmatrix}$$

La différentielle (et par "ricochet" la jacobienne J_f) est linéaire par rapport à f i.e, $d(\lambda f + \mu g) = \lambda df + \mu dg$ (ou bien $J_{\lambda f+\mu g} = \lambda J_f + \mu J_g$). Ceci est le résultat évident de la linéarité de la dérivation partielle. On donne maintenant la règle, tout aussi utile, de la dérivation des fonctions composées.

Proposition 2.4.3 *(Dérivation des fonctions composées)*
Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ *et* $g : \mathbb{R}^p \longrightarrow \mathbb{R}^q$ *différentiables. Alors*

$$d(g\circ f)_a = dg_{f(a)} \circ df_a \qquad \text{ou bien} \qquad J_{g\circ f}(a) = J_g(f(a)).J_f(a) \tag{2.16}$$

Pour ne pas se tromper dans la multiplication des matrices jacobiennes, il faut se rappeler qu'on multiplie une matrice $q \times p$ (qui doit être écrite à gauche) par une matrice $p \times n$ (écrite à droite) pour obtenir une matrice $q \times n$.

Démonstration : Posons $y = f(x)$ et $z = g(y)$ pour distinguer les variables et les valeurs des deux fonctions. Alors pour $1 \le i \le q$ et $1 \le j \le n$

$$\frac{\partial (g\circ f)_i}{\partial x_j} = \frac{\partial}{\partial x_j} g_i(f_1(x), f_2(x)), \ldots, f_p(x) = \sum_{k=1}^{p} \frac{\partial g_i(f(x))}{\partial y_k}\,\frac{\partial f_k(x)}{\partial x_n}$$

où nous avons fait usage de (2.6). C'est exactement la formule annoncée. ∎

2.5 Inversion locale

Nous allons présenter une généralisation d'un résultat abordé en première année (Analyse 1), celui de l'inversibilité locale de fonctions C^1. En effet, si pour une fonction d'une variable de classe C^1 on a $\varphi'(a) \neq 0$ alors elle est bijective d'un voisinage de a vers un voisinage de $\varphi(a)$. De plus la réciproque est de classe C^1 et $(\varphi^{-1})' = \dfrac{1}{\varphi' \circ \varphi^{-1}}$. La condition $\varphi'(a) \neq 0$ sera généralisée par l'inversibilité de la différentielle. L'outil essentielle dans la démonstration qui suivra est celui du théorème du point fixe de Banach. On peut l'énoncer dans le cas particulier de $\mathbb{R}^n$. Soit $T : \mathcal{F} \longrightarrow \mathcal{F}$ une application contractante d'un fermé $\mathcal{F}$ dans lui même i.e,

$$\exists \rho \in [0,1[\qquad \text{tel que} \qquad \forall x, x' \in \mathcal{F} \qquad \|T(x) - T(x')\| \leq \rho \|x - x'\|$$

Alors T admet dans $\mathcal{F}$ un point fixe unique, c'est-à-dire il existe un unique point $x_* \in \mathcal{F}$ tel que $T(x_*) = x_*$. Ce résultat a été largement discuté dans le cours de Topologie.

Théorème 6 *Soit U un ouvert de $\mathbb{R}^n$ et soit $f : U \longrightarrow \mathbb{R}^n$ une application de classe C^1. Si en un point a de U la différentielle df_a est inversible (det $(J_f(a) \neq 0)$), alors il existe V_a un voisinage de a et $W_{f(a)}$ un voisinage de $f(a)$ tels que $f : V_a \longrightarrow W_{f(a)}$ est bijective. L'application réciproque f^{-1} est aussi de classe C^1, de plus*

$$\left(df^{-1}\right)_{f(x)} = \left(df_x\right)^{-1}$$

pour tout $x \in V_a$.

Démonstration : Notons d'abord que dans ce théorème la jacobienne est une matrice carrée $n \times n$. C'est pourquoi son inversibilité a un sens et se vérifie par le fait que son déterminant est non nul. Notons par $\mathcal{M}_n(\mathbb{R})$ l'espace vectoriel des matrices réelles d'ordre n qui est de dimension n^2. Étant de dimension finie, toutes les normes qu'on peut y définir sont équivalentes. Nous allons travailler ici avec la norme d'opérateur définie par

$$\forall A \in \mathcal{M}_n(\mathbb{R}), \qquad \|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

car elle vérifie

$$\forall x \in \mathbb{R}^n, \qquad \|Ax\| \leq \|A\| \, \|x\|$$

Nous avons utilisé la même notation pour la norme d'une matrice et d'un vecteur, mais il n'y a pas de confusion possible. Selon les hypothèses, on a pour la différentiabilité en a

$$\forall \alpha > 0, \ \exists \delta_1 > 0, \quad \|x - a\| \leq \delta_1 \implies \|f(x) - f(a) - df_a(x - a)\| \leq \alpha \|x - a\| \tag{2.17}$$

et pour la continuité des dérivées partielles premières

$$\forall \alpha > 0, \ \exists \delta_2 > 0, \quad \|x - a\| \leq \delta_2 \implies \|J_f(x) - J_f(a)\| \leq \alpha \tag{2.18}$$

On écrira indifféremment df_a ou $J_f(a)$. Considérons $\delta_3 = \min(\delta_1, \delta_2)$ afin que les deux inégalités précédentes soit vérifiées dans la même boule fermée de centre a. Pour montrer la bijection, nous allons montrer que pour y fixé proche de $f(a)$, il existe un unique x proche de a tel que $y = f(x)$. Définissons la fonction

$$g_y(x) = x + [J_f(a)]^{-1}(y - f(x))$$

Si $\|y - f(a)\| \leq \beta$, alors de

$$\|g_y(x) - a\| = \| [J_f(a)]^{-1} \{y - f(a) - (f(x) - f(a) - J_f(a)(x-a))\} \|$$

on déduit que

$$\|g_y(x) - a\| \leq \| [J_f(a)]^{-1} \| (\beta + \alpha\delta_3) \leq \delta_3$$

avec un choix adéquat des paramètres libres α et β. Par exemple $\alpha < \dfrac{1}{\| [J_f(a)]^{-1} \|}$ puis $\beta \leq \delta_3 \left(\dfrac{1}{\| [J_f(a)]^{-1} \|} - \alpha \right)$. Cette dernière inégalité affirme que la fonction (manifestement continue) g_y envoie la boule fermée $\overline{B}(a, \delta_3)$ dans elle même. Montrons qu'on peut affiner le choix des paramètres pour qu'elle soit contractante. En effet

$$g_y(x) - g_y(x') = x - x' + [J_f(a)]^{-1}(f(x') - f(x))$$

peut-être arrangée comme suit

$$g_y(x) - g_y(x') = [J_f(a)]^{-1} \{(f(x') - f(x) - J_f(x)(x' - x)) + (J_f(x) - J_f(a))(x' - x)\}$$

et donc

$$\|g_y(x) - g_y(x')\| \leq 2\alpha \| [J_f(a)]^{-1} \| \|x - x'\|$$

On n'a plus qu'à exiger en outre que

$$\rho = 2\alpha \| [J_f(a)]^{-1} \| < 1 \Longrightarrow \alpha < \frac{1}{2\| [J_f(a)]^{-1} \|}$$

D'après le théorème du point fixe de Banach, il existe un unique $x_* \in \overline{B}(a, \delta_3)$ tel que $g_y(x_*) = x_* \Longrightarrow y = f(x_*)$ i.e, f est bijective de $\overline{B}(a, \delta_3)$ vers $\overline{B}(f(a), \beta)$. Notons f^{-1} sa fonction réciproque. Prenons $y = f(a)$ dans la définition de g_y et deux points quelconques x, x' dans $\overline{B}(a, \delta_3)$. Alors

$$x - x' = g_{f(a)}(x) - g_{f(a)}(x') + [J_f(a)]^{-1}(f(x) - f(x'))$$

et donc

$$\|x - x'\| \leq \rho\|x - x'\| + \| [J_f(a)]^{-1} \| \|f(x) - f(x')\|$$

$$\Longrightarrow \|x - x'\| \leq \frac{\| [J_f(a)]^{-1} \|}{1 - \rho} \|f(x) - f(x')\|$$

$$\Longrightarrow \|f^{-1}(y) - f^{-1}(y')\| \leq \frac{\| [J_f(a)]^{-1} \|}{1 - \rho} \|y - y'\|$$

Ceci assure la continuité de f^{-1}. Reste à montrer sa différentiabilité.
Notons $\gamma = \dfrac{\| [J_f(a)]^{-1} \|}{1-\rho}$. Rappelons qu'on peut s'inspirer des idées développées dans l'exercice 3 de la série de T.D.N°1, pour montrer que le choix précédent de α permet de dire que pour tout $x \in \overline{B}(a, \delta_3)$ la jacobienne $J_f(x)$ est inversible car elle est suffisamment proche de $J_f(a)$. Fixons $y = f(x)$ dans $\overline{B}(f(a), \beta)$. Prenons w de norme petite pour que $y + w \in \overline{B}(f(a), \beta)$ et posons $h = f^{-1}(y + w) - f^{-1}(x)$. Alors $\|h\| \le \gamma\|w\|$. Aussi $y + w = f(x + h)$. D'où

$$\begin{aligned} f^{-1}(y+w) - f^{-1}(y) - (J_f(x))^{-1}(w) &= h - (J_f(x))^{-1}(w) \\ &= -(J_f(x))^{-1} \left(f(x+h) - f(x) - (J_f(x))(h)\right) \end{aligned}$$

On sait que

$$\|f(x+h) - f(x) - (J_f(x))(h)\| = \|h\| \|\varepsilon(h)\|$$

avec $\lim\limits_{\|h\|\to 0} \|\varepsilon(h)\| = 0$. De là

$$\|f^{-1}(y+w) - f^{-1}(y) - (J_f(x))^{-1}(w)\| \le \gamma \|(J_f(x))^{-1}\| \|w\| \|\varepsilon(h)\|$$

Ceci montre bien que f^{-1} est différentiable en y avec $(df^{-1})_y = (J_f(x))^{-1}$. ■

Exemple 2.5.1 *Considérons la fonction*

$$f(x_1, x_2) = \begin{pmatrix} x_1 + x_2 \\ x_1 x_2 \end{pmatrix}$$

qui est $C^\infty(\mathbb{R}^2)$. Il est facile de calculer sa jacobienne

$$J_f(x_1, x_2) = \begin{pmatrix} 1 & 1 \\ x_2 & x_1 \end{pmatrix}$$

et det$(J_f(x_1, x_2)) = x_1 - x_2$. Si on se met au voisinage d'un point $a = (a_1, a_2)$ avec $a_1 \neq a_2$, par exemple une boule euclidienne de centre a et de rayon $r < \dfrac{|a_1 - a_2|}{\sqrt{2}}$ qui est la distance euclidienne de a à la droite Δ d'équation $x_1 - x_2 = 0$, alors f y est bijective. Contrairement au cas général où les calculs explicites sont très rares, on peut dans ce cas particulier pousser les calculs un peu plus loin. Par exemple si $y_1 = x_1 + x_2$ et $y_2 = x_1 x_2$, alors au voisinage de a avec $a_1 > a_2$, on aura

$$x_1 = \frac{y_1 + \sqrt{y_1^2 - 4y_2}}{2} \qquad , \qquad x_2 = \frac{y_1 - \sqrt{y_1^2 - 4y_2}}{2}$$

et

$$J_{f^{-1}}(y_1, y_2) = (J_f)^{-1} = \frac{1}{x_1 - x_2} \begin{pmatrix} x_1 & -1 \\ -x_2 & 1 \end{pmatrix} = \frac{1}{\sqrt{y_1^2 - 4y_2}} \begin{pmatrix} \dfrac{y_1 + \sqrt{y_1^2 - 4y_2}}{2} & -1 \\ \dfrac{-y_1 + \sqrt{y_1^2 - 4y_2}}{2} & 1 \end{pmatrix}$$

2.6 Théorème des fonctions implicites

La question abordée dans la section précédente concerne la résolution de l'équation $f(x) = y$. C'est en réalité un système de n équations à n inconnues i.e,

$$\begin{cases} f_1(x_1, x_2, \ldots, x_n) = y_1 \\ f_2(x_1, x_2, \ldots, x_n) = y_2 \\ \vdots \\ f_n(x_1, x_2, \ldots, x_n) = y_n \end{cases}$$

Nous allons dans cette section examiner des systèmes de p équations à n indéterminées avec $p < n$ i.e,

$$\begin{cases} f_1(x_1, x_2, \ldots, x_n) = y_1 \\ f_2(x_1, x_2, \ldots, x_n) = y_2 \\ \vdots \\ f_p(x_1, x_2, \ldots, x_n) = y_p \end{cases}$$

Il est clair qu'il faudra choisir parmi les n variables x_i, p variables par rapport auxquelles il sera possible de résoudre, le reste seront considérées comme paramètres. La condition d'existence et d'unicité de la solution, comme on peut s'y attendre, est celle donnée dans la section précédente à savoir l'inversibilité de la sous-jacobienne $p \times p$ formées avec les dérivées des f_i par rapports aux variables choisies. Pour énoncer le résultat suivant, nous allons privilégier les variables $x_1, x_2, \ldots, x_p$.

Théorème 7 *Identifions d'abord $\mathbb{R}^n$ à $\mathbb{R}^p \times \mathbb{R}^{n-p}$. Soit $f : \mathbb{R}^p \times \mathbb{R}^{n-p} \longrightarrow \mathbb{R}^p$ de classe C^1 et $a = (a', a'') \in \mathbb{R}^p \times \mathbb{R}^{n-p}$ avec $b = f(a', a'')$. Supposons que la matrice carrée*

$$J_p^f = \left(\frac{\partial f_i(a)}{\partial x_j} \right)_{1 \le i,j \le p}$$

soit inversible. Alors il existe $U \subset \mathbb{R}^p$ voisinage de a', $V \subset \mathbb{R}^{n-p}$ voisinage de a'' et $W \subset \mathbb{R}^p$ voisinage de b, et une fonction $\varphi : V \times W \longrightarrow U$, tels que

$$\forall\, x'' \in V,\ \forall\, y \in W \qquad x' = \varphi(x'', y)$$

est l'unique solution du système $f(x', x'') = y$; autrement dit $f(\varphi(x'', y), x'') = y$. On dit alors que $x' = \varphi(x'', y)$ est une fonction <u>implicite</u> définie par le système $f(x) = y$.

Démonstration : Il est d'abord clair, d'après les notations, que $x = (x', x'')$, c'est-à-dire $x' = (x_1, x_2, \ldots, x_p)$ et $x'' = (x_{p+1}, x_{p+2}, \ldots, x_n)$. Pour la démonstration, on se ramène au théorème de l'inversion locale en complétant le système de la façon suivante :

$$\begin{cases} f_1(x_1, x_2, \dots, x_n) = y_1 \\ f_2(x_1, x_2, \dots, x_n) = y_2 \\ \vdots \\ f_p(x_1, x_2, \dots, x_n) = y_p \\ x_{p+1} = z_{p+1} \\ x_{p+2} = z_{p+2} \\ \vdots \\ x_n = z_n \end{cases}$$

Les membres de gauche définissent une fonction $\tilde{f} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ de classe C^1. Sa jacobienne s'écrit (triangulaire par blocs)

$$J_{\tilde{f}} = \begin{pmatrix} \left(\dfrac{\partial f_i}{\partial x_j}\right)_{1\le i,j\le p} & \left(\dfrac{\partial f_i}{\partial x_j}\right)_{1\le i\le p\ ,\ p+1\le j\le n} \\ 0 & I_{n-p,n-p} \end{pmatrix}$$

On a noté par $I_{n-p,n-p}$ la matrice identité $(n-p) \times (n-p)$. Il facile de voir que $\det(J_{\tilde{f}}) = \det(J_p^f)$, et donc le théorème de l'inversion locale s'applique. La réciproque $\tilde{f}^{-1}$ existe dans un voisinage de (y, z) et $(x', x'') = \tilde{f}^{-1}(y, z)$. La fonction φ est formée des p premières composantes. Remarquez aussi que $z = x''$. ■

Exemple 2.6.1 *Considérons l'équation $x^2 + y^2 + z^2 = 1$ de la sphère euclidienne de centre $(0, 0, 0)$ et de rayon 1. Plaçons-nous par exemple au voisinage du "pôle nord" $(0, 0, 1)$. C'est une équation à trois inconnues i.e, $f(x, y, z) = x^2 + y^2 + z^2 = 1$. On a $f(0, 0, 1) = 1$ et $\dfrac{\partial f(0,0,1)}{\partial z} = 2z\ |_{(0,0,1)} = 2 \neq 0$. Donc dans un petit voisinage de $(0, 0, 1)$, z est une fonction de (x, y). Dans ce cas particulier on peut même expliciter cette fonction : $z = \varphi(x, y) = \sqrt{1 - x^2 - y^2}$. Remarquons qu'au voisinage du même point x (ou bien y) n'est pas fonction de (y, z). En effet, il existe deux solutions $x = \pm\sqrt{1 - y^2 - z^2}$ qui sont toutes les deux possibles, ce qui exclut la bijection (l'unicité de la solution). On peut aussi le vérifier par $\dfrac{\partial f(0,0,1)}{\partial x} = 2x\ |_{(0,0,1)} = 0$, non inversibilité de la sous-jacobienne.*

Exemple 2.6.2 *Regardons le système*

$$\begin{cases} f_1(x, y, z) = & x^2 - y^2 + z^2 & = 1 \\ f_2(x, y, z) = & xyz & = 1 \end{cases}$$

au voisinage du point $(1, 1, 1)$ qui est solution particulière. La jacobienne est

$$J_f(x, y, z) = \begin{pmatrix} 2x & -2y & 2z \\ yz & xz & xy \end{pmatrix}$$

Si on privilégie les variables (x, y), il faut considérer la sous-jacobienne

$$J^f_{(x,y)} = \begin{pmatrix} 2x & -2y \\ yz & xz \end{pmatrix}$$

dont le déterminant est $det(J^f_{(x,y)}) = 2z(x^2 + y^2) = 4$ *au point* $(1, 1, 1)$. *Donc au voisinage de ce point, x et y sont fonctions de z i.e,*

$$x = \varphi_1(z) \qquad , \qquad y = \varphi_2(z)$$

On laisse au lecteur la tâche (un peu lourde !) de calculer les expressions de ces deux fonctions.

2.7 Extremums locaux

Dans ce qui suit nous allons examiner la question des maximums et minimums (extremums) locaux de fonctions de $\mathbb{R}^n \longrightarrow \mathbb{R}$. Pour une variable ($n = 1$), il est bien connu que si f est C^1 et admet un extremum en $x = a$, alors $f'(a) = 0$. La réciproque est en général fausse, à savoir si a est un point critique ($f'(a) = 0$) alors on n'a pas forcément un extremum. Voir par exemple $f(x) = x^3$ qui n'a ni maximum, ni minimum en $x = 0$. Ce point s'appelle un point col (ou parfois point plat). Si le point critique n'est pas dégénéré ($f''(a) \neq 0$), alors dans ce cas on aura un maximum (respect. minimum) si $f''(a) < 0$ (respect. $f''(a) > 0$). Ce sont ces résultats que nous allons généraliser à $\mathbb{R}^n$.

Définition 2.7.1 *Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$. On dit qu'elle admet un maximum local au point a, s'il existe V_a un voisinage de a tel que*

$$\forall x \in V_a \qquad f(x) \leq f(a)$$

Un maximum est dit strict si en plus on a $\forall x \in V_a \setminus \{a\} \quad f(x) < f(a)$.
Pour un minimum on a

$$\forall x \in V_a \qquad f(x) \geq f(a)$$

Et pour le minimum strict, l'inégalité est stricte pour $x \neq a$.

Proposition 2.7.2 *Si f admet en a un extremum alors $(\nabla f)(a) = 0$, ou encore* $\forall i = 1, 2, \ldots, n \quad \dfrac{\partial f(a)}{\partial x_i} = 0$.

Démonstration : Une manière simple de le prouver est de se ramener à une variable. En effet, si on pose $\varphi(t) = f(a_1, \ldots, a_{i-1}, a_i + t, a_{i+1}, \ldots, a_n)$, alors φ admet en 0 un extremum, d'où $\varphi'(0) = 0$. C'est le résultat escompté. ■

Définition 2.7.3 *Si $(\nabla f)(a) = 0$, on dit que a est un point critique, et qu'il est non dégénéré si* $det(Hess_f(a)) \neq 0$.

La précédente proposition dit en particulier qu'il faut chercher les extremums parmi les points critiques.

Quelques rappels d'algèbre linéaire

Une matrice carrée $A = (a_{ij})_{i,j=1}^n$ est dite symétrique si $\forall i, j = 1, \ldots, n \quad a_{ij} = a_{ji}$. On a alors $\forall x, y \in \mathbb{R}^n \quad \langle Ax, y\rangle = \langle x, Ay\rangle$. Une matrice carrée A est dite positive si $\forall x \in \mathbb{R}^n \quad \langle Ax, x\rangle \geq 0$, et elle est dite définie positive si elle est déjà positive avec en plus $\langle Ax, x\rangle = 0 \Longrightarrow x = 0$.

Une matrice A symétrique est diagonalisable dans une base orthogonale i.e,

$$\exists P \text{ telle que } P^tP = PP^t = I \quad \text{et} \quad A = PDP^t$$

où P^t désigne la matrice transposée de P. La condition sur P exprime le caractère orthogonal. La matrice D est diagonale

$$D = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}$$

où les λ_i sont les valeurs propres de A. La positivité de A est équivalente à la positivité des valeurs propres. La définie positivité est, elle, équivalente à la stricte positivité des valeurs propres. On peut montrer, sans grande difficulté, que l'encadrement suivant a lieu

$$\lambda_1\|x\|^2 \leq \langle Ax, x\rangle \leq \lambda_n\|x\|^2 \tag{2.19}$$

si $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Vérifier qu'une matrice symétrique est définie positive à l'aide de la définition ou bien grâce aux valeurs propres peut s'avérer une tâche très difficile. C'est pourquoi le critère suivant est de loin le plus pratique pour cette vérification. Donnons avant, la définition des mineurs principaux d'une matrice. On appelle mineurs principaux d'une matrice $A = (a_{ij})_{i,j=1}^n$ les déterminants suivants :

$$\Delta_1 = a_{11} \ , \ \Delta_2 = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \ , \ \Delta_3 = \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \ , \ldots, \Delta_n = \det(A)$$

Critère de définie positivité : une matrice symétrique A est définie positive si et seulement si tous ses mineurs principaux sont strictement positifs.

Exemple : $A = \begin{pmatrix} 4 & -1 & 3 & 0 \\ -1 & 2 & -1 & 1 \\ 3 & -1 & 5 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$ est définie positive car

$$\Delta_1 = 4 \ , \ \Delta_2 = 7 \ , \ \Delta_3 = 19 \ , \ \Delta_4 = 8$$

Enfin une matrice symétrique est négative (resp. définie négative) si $-A$ est positive (resp. définie positive). Elle est non définie si elle n'est ni positive ni négative i.e, elle admet des valeurs propres positives et négatives.

Théorème 8 *Soit* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *de classe* C^2 *et* a *un point critique. Alors*

1. *Si* f *admet en* a *un minimum local, alors la matrice hessienne* $Hess_f(a)$ *est positive.*

2. *Si* f *admet en* a *un maximum local, alors la matrice hessienne* $Hess_f(a)$ *est négative*

Inversement, en supposant de plus qu'il est non dégénéré,

1. *Si* $Hess_f(a)$ *est définie positive, alors* f *admet en* a *un minimum strict.*

2. *Si* $Hess_f(a)$ *est définie négative, alors* f *admet en* a *un maximum strict.*

Si la hessienne est non définie, on dit que a *est un col.*

Démonstration : Tout est basé sur la formule de Taylor-Young

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

Comme a est un point critique, alors en prenant $h = t.u$, avec u un vecteur unitaire ($\|u\| = 1$),

$$f(a+tu) - f(a) = t^2\left(\frac{1}{2}\langle Hess_f(a).u, u\rangle + \varepsilon(t.u)\right)$$

Faisons la démonstration pour un minimum. Pour $|t|$ petite, on a $f(a+tu) - f(a) \geq 0$, donc $\frac{1}{2}\langle Hess_f(a).u, u\rangle + \varepsilon(t.u) \geq 0$. En faisant tendre t vers 0 on aura la positivité de la hessienne (à partir de u on revient à n'importe quel vecteur en posant $u = \frac{\xi}{\|\xi\|}$).
Inversement, si la hessienne est définie positive, alors sa plus petite valeur propre est strictement positive, $\lambda_1 > 0$. En utilisant (2.19) on aura

$$f(a+h) - f(a) \geq \|h\|^2\left(\frac{1}{2}\lambda_1 + \varepsilon(h)\right)$$

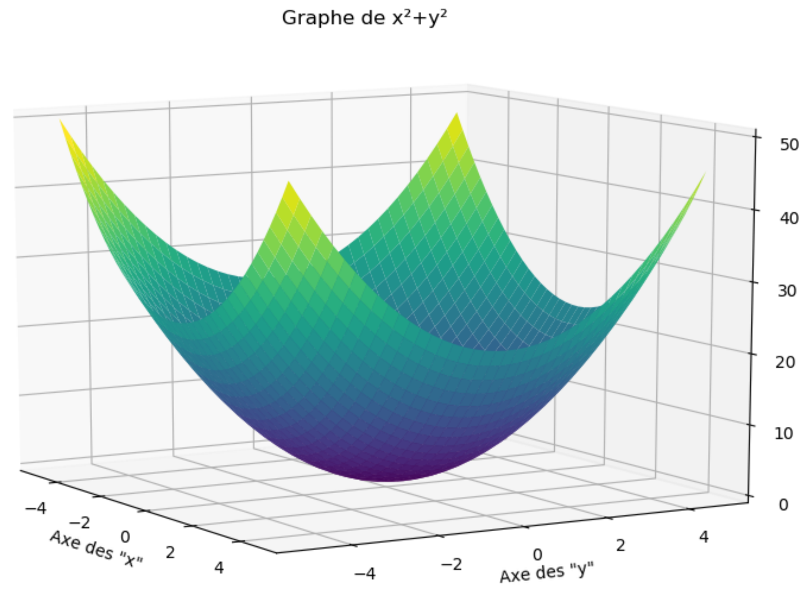
D'où le résultat en prenant $\|h\|$ suffisamment petite pour que $\frac{1}{2}\lambda_1 + \varepsilon(h) > 0$. ■

Exemple 2.7.4 *Soit* $f(x,y) = x^2 + y^2$. *Alors*

$$\nabla f(x,y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix} \qquad et \qquad Hess_f(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Il y a un seul point critique $(0,0)$ *et il est non dégénéré puisque partout* $\det(Hess_f) = 4 \neq 0$. *Aussi la hessienne est définie positive. Donc en* $(0,0)$ *il y a un minimum strict (ce qui est par ailleurs facilement détectable.)*

Voici une illustration par le dessin du graphe dans un voisinage de $(0,0)$

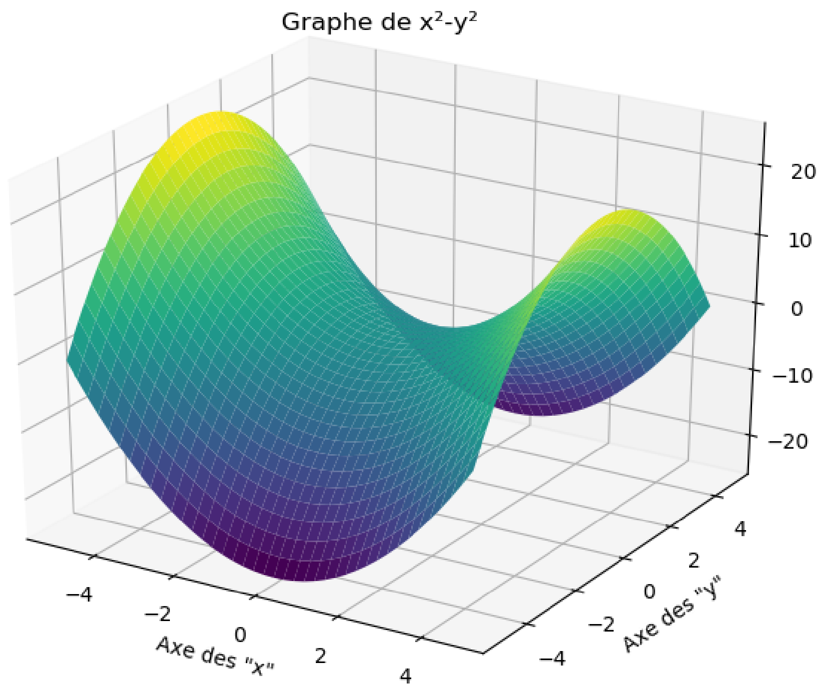


Exemple 2.7.5 *Soit $g(x,y) = x^2 - y^2$. Alors*

$$\nabla g(x,y) = \begin{pmatrix} 2x \\ -2y \end{pmatrix} \qquad et \qquad Hess_g(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

Il y a un seul point critique $(0,0)$ et il est non dégénéré puisque partout $\det(Hess_g) = -4 \neq 0$. Dans ce cas la hessienne est non définie. Donc en $(0,0)$ il y a un point "col" ou bien "selle".

Voici le graphe dans un voisinage de $(0,0)$.



On voit que suivant un certain chemin sur cette surface $(0,0,0)$ apparaît comme un maximum, et suivant un autre chemin, il apparaît comme un minimum.

2.8 Extrema liés

Dans cette dernière section nous allons présenter la notion d'extrema liés (le mot extrema est une forme du pluriel d'extremum). C'est la recherche de maximum ou de minimum d'une fonction $f(x)$ avec en plus une condition (contrainte) sur les variables ($g(x) = 0$). On peut même envisager plusieurs contraintes. Le nombre de contraintes p est dans tous les cas strictement inférieur à n ($p < n$). Posons

$$\mathcal{C} = \{x \in \mathbb{R}^n \diagup g(x) = 0\}$$

Le problème est d'étudier l'existence de $\min_{x \in \mathcal{C}} f(x)$ ou $\max_{x \in \mathcal{C}} f(x)$.

Proposition 2.8.1 *On suppose que les fonctions f est g sont de classe C^1. Si f admet en a un extremum lié alors il existe un unique nombre réel λ tel que*

$$(\nabla f)(a) = \lambda(\nabla g)(a)$$

Ce nombre s'appelle multiplicateur de Lagrange.

Démonstration : Soit $\gamma : J \longrightarrow \mathcal{C}$ une application différentiable définie sur un petit intervalle ouvert J contenant 0 et telle que $\gamma(0) = a$. D'après les hypothèses, l'application d'une variable réelle $f(\gamma(t))$ admet un extremum en 0, donc sa dérivée en 0 s'annule i.e,

$$\langle (\nabla f)(a), \gamma'(0) \rangle = 0$$

D'autre part, $g(\gamma(t)) = 0$, d'où

$$\langle (\nabla g)(a), \gamma'(0) \rangle = 0$$

On apprend dans le cours de géométrie que le vecteur $\gamma'(0)$ est tangent à l'hypersurface $\mathcal{C}$, et donc parcourt un sous-espace vectoriel de $\mathbb{R}^n$ de dimension $n-1$. Les deux vecteurs $(\nabla f)(a)$ et $(\nabla g)(a)$ sont donc orthogonaux à un même sous-espace de dimension $n-1$. Il sont de ce fait colinéaires. ■

Posons $F(x, \lambda) = f(x) - \lambda g(x)$. D'après cette proposition chercher les extremums liés revient à résoudre le système

$$\begin{cases} (\nabla_x F)(x, \lambda) = 0 \\ g(x) = 0 \end{cases}$$

par rapport à x et λ. C'est un système de $n + 1$ équations à $n + 1$ inconnues.

Supposons que (a_*, λ_*) soit un point critique de F, c'est-à-dire une solution du système précédent. Si on suppose en plus que f et g sont de classe C^2, alors la sous-hessienne de F au point (a_*, λ_*) qui concerne les variables x_i, renseignera sur la nature de ce point critique. Nous verrons plus clairement à travers les exemples qui suivront.

Exemple 2.8.2 *On veut résoudre par les méthodes développées ici, un problème bien connu en géométrie euclidienne plane. Il s'agit de trouver la distance euclidienne d'un* *point $A(a,b)$ à une droite (D) d'équation $\alpha x + \beta y + \gamma = 0$ ($\alpha^2 + \beta^2 \neq 0$). Soit $M(x,y) \in (D)$. La distance euclidienne de A à M est donnée par :*

$$d(A,M) = \sqrt{(x-a)^2 + (y-b)^2}$$

Pour bénéficier du caractère C^∞, nous travaillerons plutôt avec la fonction

$$f(x,y) = (x-a)^2 + (y-b)^2 = d(A,M)^2$$

car la distance est minimale si et seulement si f l'est. Ici la contrainte est manifestement $g(x,y) = \alpha x + \beta y + \gamma$. On a

$$F(x,y,\lambda) = (x-a)^2 + (y-b)^2 - \lambda\,(\alpha x + \beta y + \gamma)$$

et le système à résoudre

$$\begin{cases} \dfrac{\partial F}{\partial x} = & 2(x-a) - \lambda\alpha & = 0 \\[2ex] \dfrac{\partial F}{\partial y} = & 2(y-a) - \lambda\beta & = 0 \\[2ex] \dfrac{\partial F}{\partial \lambda} = & -(\alpha x + \beta y + \gamma) & = 0 \end{cases}$$

Des deux premières équations on aura $x = a + \dfrac{\lambda\alpha}{2}$ et $y = b + \dfrac{\lambda\beta}{2}$. En les remplaçant dans la troisième équation on a $\alpha a + \beta b + \gamma + \dfrac{\lambda}{2}(\alpha^2 + \beta^2) = 0$. La solution du système est donc

$$\begin{cases} \lambda_* = \dfrac{-2}{\alpha^2+\beta^2}(\alpha a + \beta b + \gamma) \\[2ex] x_* = a - \dfrac{\alpha}{\alpha^2+\beta^2}(\alpha a + \beta b + \gamma) \\[2ex] y_* = b - \dfrac{\beta}{\alpha^2+\beta^2}(\alpha a + \beta b + \gamma) \end{cases}$$

Pour montrer que nous avons effectivement un minimum, il suffit de calculer la sous-hessienne

$$\tilde{H}_F(x,y) = \begin{pmatrix} \dfrac{\partial^2 F}{\partial x^2} & \dfrac{\partial^2 F}{\partial x \partial y} \\ \dfrac{\partial^2 F}{\partial x \partial y} & \dfrac{\partial^2 F}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Elle est définie positive. La valeur de ce minimum est $f(x_*, y_*) = \dfrac{(\alpha a + \beta b + \gamma)^2}{\alpha^2 + \beta^2}$. Et donc la distance est

$$d(A,(D)) = \frac{|\alpha a + \beta b + \gamma|}{\sqrt{\alpha^2 + \beta^2}}$$

formule bien connue par ailleurs.

On laisse le soin au lecteur de reprendre tous ces calculs dans le cas d'un point $A = (a, b, c)$ de $\mathbb{R}^3$ et d'un plan $\mathcal{P}$ d'équation $\alpha x + \beta y + \gamma z + \delta = 0$. La formule attendue est

$$d(A, \mathcal{P}) = \frac{|\alpha a + \beta b + \gamma c + \delta|}{\sqrt{\alpha^2 + \beta^2 + \gamma^2}}$$

Exemple 2.8.3 *On se propose pour terminer ce chapitre, de traiter le problème de l'exemple précédent mais dans $\mathbb{R}^3$. Une droite dans l'espace est l'intersection de deux plans $(D) = \mathcal{P} \cap \mathcal{P}'$. Précisons les équations des deux plans*

$$\mathcal{P} : \qquad g_1(x, y, z) = \alpha x + \beta y + \gamma z + \delta = 0$$

$$\mathcal{P}' : \qquad g_2(x, y, z) = \alpha' x + \beta' y + \gamma' z + \delta' = 0$$

Donnons d'abord la condition sur les paramètres pour que les deux plans ne soient ni confondus ni parallèles et distincts. Posons $\vec{n} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}$ et $\vec{n'} = \begin{pmatrix} \alpha' \\ \beta' \\ \gamma' \end{pmatrix}$ pour désigner les vecteurs normaux à $\mathcal{P}$ et $\mathcal{P}'$ respectivement. L'intersection suivant une droite de ces deux plans est équivalente à l'indépendance linéaire des vecteurs normaux. Une manière de l'exprimer est d'utiliser le Gramien (déterminant de la matrice de Gram) :

$$G := det \begin{pmatrix} \langle \vec{n}, \vec{n} \rangle & \left\langle \vec{n}, \vec{n'} \right\rangle \\ \left\langle \vec{n}, \vec{n'} \right\rangle & \left\langle \vec{n'}, \vec{n'} \right\rangle \end{pmatrix} = \|\vec{n}\|^2 \, \|\vec{n'}\|^2 - \left\langle \vec{n}, \vec{n'} \right\rangle^2 \neq 0$$

En fait $G > 0$ par l'inégalité de Cauchy-Schwarz. Notons $X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ et $A = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ et considérons

$$F(X, \lambda_1, \lambda_2) = \|X - A\|^2 - \lambda_1 g_1(X) - \lambda_2 g_2(X)$$

Remarquons qu'on a $g_1(X) = \langle X, \vec{n} \rangle + \delta$ et $g_2(X) = \left\langle X, \vec{n'} \right\rangle + \delta'$, d'où

$$F(X, \lambda_1, \lambda_2) = \|X - A\|^2 - \lambda_1(\langle X, \vec{n} \rangle + \delta) - \lambda_2(\left\langle X, \vec{n'} \right\rangle + \delta')$$

De là

$$(\nabla_X)F = 2(X - A) - \lambda_1 \vec{n} - \lambda_2 \vec{n'} = 0$$

$$\Longrightarrow X = X_* = A + \frac{1}{2}\left(\lambda_1 \vec{n} + \lambda_2 \vec{n'}\right)$$

Pour déterminer λ_1 et λ_2 il suffit de remplacer la valeur trouvée de X dans les contraintes $g_1(X) = 0$ et $g_2(X) = 0$. On obtient le système

$$\begin{pmatrix} \langle \vec{n}, \vec{n} \rangle & \left\langle \vec{n}, \vec{n'} \right\rangle \\ \left\langle \vec{n}, \vec{n'} \right\rangle & \left\langle \vec{n'}, \vec{n'} \right\rangle \end{pmatrix} \begin{pmatrix} \frac{\lambda_1}{2} \\ \frac{\lambda_2}{2} \end{pmatrix} = \begin{pmatrix} -g_1(A) \\ -g_2(A) \end{pmatrix}$$

La solution est donnée par

$$\frac{1}{2}\begin{pmatrix} \lambda_{1,*} \\ \lambda_{2,*} \end{pmatrix} = \frac{1}{G}\begin{pmatrix} \left\langle \vec{n'}, \vec{n'} \right\rangle & -\left\langle \vec{n}, \vec{n'} \right\rangle \\ -\left\langle \vec{n}, \vec{n'} \right\rangle & \langle \vec{n}, \vec{n} \rangle \end{pmatrix} \begin{pmatrix} -g_1(A) \\ -g_2(A) \end{pmatrix}$$

On en déduit en définitive

$$d(A,(D)) = \|X - A\| = \frac{1}{\sqrt{G}}\|g_1(A)\vec{n'} - g_2(A)\vec{n}\|$$