II.4 Dérivées d'ordre supérieur

Cette section porte sur les dérivées d'ordre supérieur. Nous nous focalisons au départ sur la différentielle seconde d'une fonction scalaire, et sur la notion de *matrice hessienne* qui permet de la représenter dans une base orthonormée, puis nous présentons un cadre plus abstrait permettant de généraliser ces notions à des applications à valeurs dans un espace multidimensionnel, et de définir une notion de dérivation à un ordre arbitraire.

II.4.1 Dérivées partielles d'ordre supérieur pour les fonctions scalaires

Définition II.4.1. (Dérivées partielles d'ordre 2)
Soit f une application d'un ouvert U de $\mathbb{R}^n$ dans $\mathbb{R}$. On suppose que f admet des dérivées partielles $\partial f/\partial x_i$ continues sur U (f est donc continûment différentiable d'après la proposition II.1.15). Si chacune de ces dérivées partielles est dérivable en x par rapport à chacune des variables, on appelle dérivées partielles d'ordre 2 les quantités correspondantes, notées

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

Définition II.4.2. (Matrice hessienne)
Dans le cadre de la définition précédente, on appelle *matrice hessienne* en x, et l'on note $H_f(x)$ (ou plus simplement $H(x)$ s'il n'y a pas d'ambigüité) la matrice carrée dont les éléments sont les dérivées partielles d'ordre 2

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

La proposition qui suit, capitale, établit que, si les dérivées secondes sont définies au voisinage d'un point x, et sont continues en ce point, alors la matrice hessienne est symétrique.

Théorème II.4.3. (Schwarz)
Soit f une application d'un ouvert U de $\mathbb{R}^n$ dans $\mathbb{R}$, et $x \in U$. On suppose que f admet des dérivées partielles d'ordre 2 dans un voisinage de x, et que ses dérivées partielles sont *continues* en x. Alors la matrice hessienne en x est *symétrique*, i.e.

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

Démonstration. On considère une fonction de 2 variables seulement (on peut se ramèner à ce cas-là en gelant $n-2$ variables). L'idée est d'écrire de deux manières la quantité

$$f(x_1+h_2, x_2+h_2) - f(x_1, x_2),$$

en suivant deux chemins différents entre (x_1, x_2) et (x_1+h_1, x_2+h_2). On a en premier lieu

$$f(x_1+h_1, x_2+h_2) - f(x_1, x_2) = f(x_1+h_1, x_2+h_2) - f(x_1+h_1, x_2) + f(x_1+h_1, x_2) - f(x_1, x_2).$$

Les 2 derniers termes s'écrivent

$$f(x_1+h_1,x_2)-f(x_1,x_2)=h_1\partial_1 f(x_1,x_2)+\frac{h_1^2}{2}\partial_{11}f(x_1+\theta_1 h_1,x_2),$$

avec $\theta_1 \in]0,1[$. La première différence du membre de droite s'écrit elle

$$f(x_1+h_1,x_2+h_2)-f(x_1+h_1,x_2)=h_2\partial_2 f(x_1+h_1,x_2)+\frac{h_2^2}{2}\partial_{22}f(x_1+h_1,x_2+\theta_2' h_2)$$

Si l'on écrit maintenant la même quantité $f(x_1+h_1,x_2+h_2)-f(x_1,x_2)$ de la façon suivante

$$f(x_1+h_1,x_2+h_2)-f(x_1,x_2)=f(x_1+h_1,x_2+h_2)-f(x_1,x_2+h_2)+f(x_1,x_2+h_2)-f(x_1,x_2),$$

que l'on utilise des développement de Taylor-Lagrange comme précédemment, et que l'on identifie les deux écritures, on obtient

$$\begin{aligned}
0 &= h_2 h_1\left(\frac{\partial_2 f(x_1+h_1,x_2)-\partial_2 f(x_1,x_2)}{h_1}-\frac{\partial_2 f(x_1,x_2+h_2)-\partial_2 f(x_1,x_2)}{h_2}\right)\\
&+ \frac{h_2^2}{2}\left(\partial_{22}f(x_1+h_1,x_2+\theta_2' h_2)-\partial_{22}f(x_1,x_2+\theta_2 h_2)\right)\\
&+ \frac{h_1^2}{2}\left(\partial_{11}f(x_1+\theta_1' h_1,x_2+h_2)-\partial_{11}f(x_1+\theta_1 h_1,x_2)\right).
\end{aligned}$$

Si l'on prend maintenant h_1 et h_2 égaux à ε, et que l'on fait tendre ε vers 0, les deux derniers termes sont des $o(\varepsilon^2)$ par continuité de la dérivée seconde. Le premier terme doit donc lui même être un $o(\varepsilon^2)$, ce qui impose que la quantité entre parenthèse converge vers 0 avec ε, d'où le résultat. □

Définition II.4.4. (Continue différentiabilité)
Soit f une application d'un ouvert U de $\mathbb{R}^n$ dans $\mathbb{R}$. On dit que f est deux fois continûment différentiable sur U, et l'on écrit $f \in C^2(U)$, si toutes les dérivées partielles d'ordre 2 de f existent et sont continues sur U, ce qui est équivalent à dire que f admet une matrice hessienne $H(x)$ en tout point x de U, et que la correspondance $x \mapsto H(x)$ est continue.

Remarque II.4.5. En toute rigueur (voir à la fin de la section pour plus de détail), mais au prix de certaines définitions abstraites que nous avons choisi d'écarter, nous devrions définir la différentielle seconde comme l'application différentielle de la différentielle : $d^2 f = d(df)$, c'est à dire comme une application linéaire de $\mathbb{R}^n$ dans l'espace des applications linéaires de $\mathbb{R}^n$ dans $\mathbb{R}$. Et ensuite dire que l'application est C^2 si cette correspondance est continue, indépendamment des dérivées partielles premières ou secondes afférentes à une base particulière. On peut néanmoins montrer, dans l'esprit de la proposition II.1.15 pour les différentielles d'ordre 1, que la continuité de toutes les dérivées partielles secondes implique le caractère C^2. Il est donc licite de fonder la définition précédente sur la caractérisation basée sur les dérivées partielles.

Proposition II.4.6. (Développement limité du gradient)
Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans $\mathbb{R}$, deux fois continûment différentiable sur U, et h tel que le segment $[x,x+h]=\{x+\theta h\,,\ \theta\in[0,1]\}$ soit inclus dans U. On a alors

$$\nabla f(x+h)=\nabla f(x)+H(x)\cdot h+\varepsilon(h)\,\|h\|\,.$$

Démonstration. Pour tout $i = 1, \ldots, N$, la fonction $y \longmapsto \partial_i f(y)$ est continûment différentiable sur U, et l'on a

$$\begin{aligned}
\partial_i f(x+h) &= \partial_i f(x) + \langle \nabla \partial_i f(x) \,|\, h \rangle + \varepsilon(h) \, \|h\| \\
&= \partial_i f(x) + \sum_{j=1}^{N} \partial_j \partial_i f(x) h_j + \varepsilon(h) \, \|h\| \\
&= \partial_i f(x) + H \cdot h + \varepsilon(h) \, \|h\| \, ,
\end{aligned}$$

qui est l'identité annoncée. □

Proposition II.4.7. (Développement limité à l'ordre 2)
Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans $\mathbb{R}$, deux fois continûment différentiable sur U, et h tel que le segment $[x, x+h] = \{x + \theta h \, , \; \theta \in [0,1]\}$ soit inclus dans U. On a alors

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

Démonstration. On introduit la fonction

$$h \longmapsto g(h) = f(x+h) - f(x) - \langle \nabla f(x) \,|\, h \rangle - \frac{1}{2} \langle h \,|\, H(x) \cdot h \rangle.$$

On a

$$\nabla g(h) = \nabla f(x+h) - \nabla f(x) - H(x) \cdot h.$$

D'après la proposition II.4.6 ci-dessus, cette quantité est un $o(h)$. Pour tout $\varepsilon > 0$, il existe donc $\eta > 0$ tel que, pour tout h tel que $\|h\| \leq \eta$,

$$\|\nabla g(h)\| \leq \varepsilon \, \|h\| \, .$$

On applique à présent le théorème des accroissements finis II.1.18 :

$$\|g(h)\| = \|g(h) - g(0)\| \leq \sup_{h' \in [0,h]} \|\nabla f(x+h')\| \, \|h\| \leq \varepsilon \, \|h\|^2 \, ,$$

avec $g(h) = f(x+h) - f(x) - \langle \nabla f(x) \,|\, h \rangle - \frac{1}{2} \langle h \,|\, H(x) \cdot h \rangle$. □

Proposition II.4.8. (Développement de Taylor avec reste intégral)
Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans $\mathbb{R}$, deux fois continûment différentiable sur U, et h tel que le segment $[x, x+h] = \{x + \theta h \, , \; \theta \in [0,1]\}$ soit inclus dans U. On a alors

$$f(x+h) = f(x) + \langle \nabla f(x) \,|\, h \rangle + \int_0^1 \langle H(x+th) \cdot h \,|\, h \rangle (1-t) \, dt.$$

Démonstration. On pose $\Phi(t) = f(x+th)$. On a, par intégration par parties,

$$\begin{aligned}
\Phi(1) &= \Phi(0) + \int_0^1 \Phi'(t) \, dt = \Phi(0) + \left[\Phi'(t)(1-t)\right]_0^1 + \int_0^1 \Phi''(t)(1-t) \, dt \\
&= \Phi(0) + \Phi'(0) + \int_0^1 \Phi''(t)(1-t) \, dt.
\end{aligned}$$

On a

$$\Phi(t) = f(x+th) \, , \; \Phi'((t) = \langle \nabla f(x+th) \,|\, h \rangle \, , \; \Phi''(t) = \langle H(x+th) \cdot h \,|\, h \rangle,$$

qui conduit à la formule annoncée. □

Définition II.4.9. (Laplacien, opérateur laplacien)
Soit f une fonction définie sur un ouvert U de $\mathbb{R}^n$, à valeurs dans $\mathbb{R}$. On suppose que la matrice hessienne est définie en $x \in U$. On appelle laplacien de f en x, la quantité

$$\Delta f = \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2} = \operatorname{tr}(H)$$

(trace de la hessienne de f). Pour les fonctions telles que cette quantité est définie sur U, on appelle opérateur laplacien l'application qui à la fonction f associe la fonction Δf.

II.4.2 Différentielles d'ordre supérieur pour les fonctions de $\mathbb{R}^n$ dans $\mathbb{R}^m$

La notion de différentielle seconde découle de celle de la différentielle. Comme pour les fonction de $\mathbb{R}$ dans $\mathbb{R}$, la différentielle seconde sera simplement la différentielle de la différentielle. Pour une fonction de départ de $\mathbb{R}$ dans $\mathbb{R}^m$, cette différentielle est une application de $\mathbb{R}^n$ dans $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ (définition II.1.2). Si nous souhaitons dériver cette différentielle, nous avons besoin d'une définition un peu plus générale, qui porte sur des applications à valeurs dans un espace vectoriel normé E.

Définition II.4.10. (Différentielle (•))
Soit f une application définie d'un ouvert $U \subset \mathbb{R}^n$ dans un espace vectoriel normé E. On dit que f est différentiable en $x \in U$ s'il existe une application linéaire de $\mathbb{R}^n$ dans E, notée $df(x)$, telle que

$$f(x+h) = f(x) + df(x) \cdot h + \varepsilon(h) \|h\| \tag{II.4.1}$$

où $\varepsilon(h)$ est une application de $\mathbb{R}^n$ dans E, telle que $\|\varepsilon(h)\|$ tend vers 0 quand h tend vers 0.

Définition II.4.11. (Différentielle seconde (• • •))
Soit f une application différentiable dans un voisinage U d'un point $x \in \mathbb{R}^n$, à valeurs dans $\mathbb{R}^m$. On dit que f est deux fois différentiable en $x \in U$ si l'application $x \mapsto df(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ (muni de la norme d'opérateur canonique) est différentiable en x. La différentielle de df en x, notée $d^2 f(x)$, est une application linéaire de $\mathbb{R}^n$ dans $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.

On peut de la même manière, si df^2 est définie dans un voisinage de x, définir la différentielle d'ordre 3 par $d^3 f = d(d^2 f)$, qui est une application de $\mathbb{R}^n$ dans $\mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$, et les différentielles d'ordre $k = 4, 5, \ldots$

II.5 Exercices

Exercice II.5.1. Calculer les matrices hessiennes des applications suivantes

$$f(x_1,x_2) = \frac{1}{2}(x_1^2 + x_2^2)\,,\ f(x,y) = x_1^p x_2^q.$$

Exercice II.5.2. Soit A une matrice carrée d'ordre n, et f l'application quadratique de $\mathbb{R}^n$ dans $\mathbb{R}$ définie par

$$x = (x_1,\ldots,x_n) \longmapsto f(x) = \langle A\cdot x \,|\, x\rangle.$$

a) Calculer la matrice Hessienne de f.

b) Dans quel cas cette matrice hessienne est-elle nulle ?

Exercice II.5.3. Soit f une fonction deux fois continûment différentiable au voisinage d'un point $x \in \mathbb{R}^n$, et h un vecteur de $\mathbb{R}^n$.

a) Quelle est la limite de

$$\frac{f(x-\varepsilon h) - 2f(x) + f(x+\varepsilon h)}{\varepsilon^2}$$

quand ε tend vers 0 ?

b) Soit f une fonction deux fois continûment différentiable sur un ouvert convexe U. On suppose de plus f convexe, c'est-à-dire telle que

$$f((1-\theta)x + \theta y) \leq (1-\theta)f(x) + \theta f(y) \quad \forall x\,,\ y \in U\,,\ \forall \theta \in]0,1[.$$

Montrer que pour tout x de U, la matrice H est positive, c'est à dire que

$$\langle H(x)\cdot h \,|\, h\rangle \geq 0 \quad \forall h \in \mathbb{R}^n.$$

c) Soit f une fonction deux fois continûment différentiable sur un ouvert convexe U. On suppose que f est λ-*convexe*, c'est-à-dire telle que

$$f((1-\theta)x + \theta y) \leq (1-\theta)f(x) + \theta f(y) - \frac{\lambda}{2}\theta(1-\theta)\,\|y-x\|^2 \quad \forall x\,,\ y \in U\,,\ \forall \theta \in]0,1[,$$

avec $\lambda \in \mathbb{R}$. Que peut-on en déduire sur la matrice $H(x)$, pour $x \in U$?

d) (Condition suffisante d'optimalité locale)
On considère f une fonction deux fois continûment différentiable sur un ouvert U. On considère un point $x \in U$ en lequel le gradient de f s'annule, et tel que les valeurs propres de $H(x)$ sont toutes strictement positives. Que peut on dire de x vis-à-vis de f ?

e) (Condition suffisante d'optimalité globale)
On suppose maintenant l'ouvert U convexe, et f convexe sur U. Montrer que x minimise f sur U, c'est à-dire-que

$$f(y) \geq f(x) \quad \forall y \in U.$$

f) (Condition nécessaire d'optimalité)
On considère pour finir f une fonction deux fois continûment différentiable sur un ouvert U.

On suppose que $x \in U$ est un minimiseur local de f. Montrer que $\nabla f(x) = 0$, et que $H(x)$ est une matrice positive.

Exercice II.5.4. Soit f une fonction deux fois continûment différentiable sur un ouvert U de $\mathbb{R}^n$, et telle que ∇f est de norme constante égale à un sur U. Montrer que

$$H(x) \cdot \nabla f(x) = 0.$$

pour tout x dans U.

Exercice II.5.5. On cherche ici à exprimer le fait que le laplacien quantifie l'écart entre la valeur ponctuelle d'une fonction et la moyenne des valeurs de la fonction au voisinage de ce point. En dimension 1, une telle propriété est données par le a) de l'exercice II.5.3. En dimension 2, cette propriété prend la forme exprimée ci-dessous.

Soit f une fonction à valeurs réelles deux fois continûment différentiable au voisinage d'un point $x \in \mathbb{R}^2$. On note e_θ le vecteur unitaire $(\cos\theta, \sin\theta)$. Montrer que

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \frac{1}{2\pi} \int_0^{2\pi} \left(f(x + \varepsilon e_\theta) - f(x) \right) \, d\theta = \frac{1}{4} \Delta f(x).$$