

Calcul Différentiel III

STEP, MINES ParisTech*

9 décembre 2020 (#a46c5a3)

Table des matières

Objectifs d'apprentissage	3
Matrice hessienne et différentielle d'ordre 2	3
Dérivées partielles d'ordre 2	3
Matrice hessienne	4
Continue différentiabilité d'ordre 2	4
Différentielle d'ordre 2	5
Linéarité de $d(x \mapsto df(x) \cdot h_1)(x)$ par rapport à h_1	5
Applications linéaires d'ordre 2	5
Différentiation automatique d'ordre 2	6
Différentielle d'ordre 2 et matrice hessienne	7
Continue différentiabilité et différentiabilité d'ordre 2	8
Développement limité du gradient	9
Symétrie de la différentielle d'ordre 2	9
Développement limité à l'ordre 2	10
Développement de Taylor d'ordre 1 avec reste intégral	11
Différentielle d'ordre supérieur	12
Tenseur d'ordre n	12
Scalars, vecteurs, matrices	12
Les tenseurs avec NumPy	13
Applications linéaires d'ordre supérieur	14
Applications multilinéaires	14
Produit tensoriel	15
Produits tensoriels classiques	15
Produit tensoriel avec NumPy	15

*Ce document est un des produits du projet  boisgera/CDIS, initié par la collaboration de (S)ébastien Boisgérault (CAOR), (T)homas Romary et (E)milie Chautru (GEOSCIENCES), (P)auline Bernard (CAS), avec la contribution de Gabriel Stoltz (Ecole des Ponts ParisTech, CERMICS). Il est mis à disposition selon les termes de la licence Creative Commons “attribution – pas d'utilisation commerciale – partage dans les mêmes conditions” 4.0 internationale.

Différentielle d'ordre k	17
Dérivées partielles d'ordre k	17
Calcul des dérivées partielles d'ordre k	17
Différentielle d'ordre k et tenseur	18
Stratification	18
Symétrie des différentielles d'ordre supérieur	19
Dérivées partielles d'ordre supérieur et multi-indices	19
Puissance symbolique	20
Développement limité d'ordre supérieur	20
Développement de Taylor avec reste intégral I	21
Développement de Taylor avec reste intégral II	22
Annexe	22
Variation d'ordre 1 et 2	22
Variation d'ordre 2 et matrice hessienne	22
Exercices complémentaires	23
Convexité	23
Différentiation en chaîne à l'ordre 2	24
Différentiation matricielle	24
Solutions	25
Exercices essentiels	25
Convexité	28
Différentiation en chaîne à l'ordre 2	29
TODO – Différentiation matricielle	30
Références	32

Objectifs d'apprentissage

Cette section s'efforce d'expliciter et de hiérarchiser les acquis d'apprentissages associés au chapitre. Ces objectifs sont organisés en paliers :

(o) Prérequis (●) Fondamental (●●) Standard (●●●) Avancé (●●●●) Expert

Sauf mention particulière, les objectifs “Expert”, les démonstrations du document¹ et les contenus en annexe ne sont pas exigibles (“hors-programme”).

Matrice hessienne et différentiabilité d'ordre 2

- ● connaître les définitions et notations de la matrice hessienne et des dérivées partielles d'ordre 2.
- ● savoir définir et relier existence de la matrice hessienne, différentiabilité d'ordre 2 et continue différentiabilité d'ordre 2.
- ● savoir à quelle condition la matrice hessienne est symétrique ; savoir exploiter cette symétrie.
- ●●● comprendre le concept de fonctions linéaires d'ordre 2 ; savoir manipuler les notations associées (“→”, “.”, avec associativité à droite et à gauche respectivement).
- ● savoir définir la différentielle d'ordre 2 de f en x .
- ● savoir déterminer $d^2f(x)$ en fonction de $H_f(x)$ et réciproquement.
- ●● connaître et savoir exploiter les développements limités et avec reste intégral faisant intervenir la matrice Hessienne.

Différentielle d'ordre supérieur TODO

Matrice hessienne et différentielle d'ordre 2

Définition – Dérivées partielles d'ordre 2

Soit U un ouvert de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}$ et $x \in U$. Si la j_1 -ème dérivée partielle de f est définie sur U , et que la j_2 -ème dérivée partielle de $\partial_{j_1}f$ en x existe, on note

$$\partial_{j_2 j_1}^2 f(x) := \partial_{j_2}(\partial_{j_1} f)(x) \in \mathbb{R}.$$

sa *dérivée partielle d'ordre 2 par rapport aux j_1 -ème et j_2 -ème variables*.

1. L'étude des démonstrations du cours peut toutefois contribuer à votre apprentissage, au même titre que la résolution d'exercices.

Définition – Matrice hessienne

Soient U un ouvert de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}$ et x un point de U . Si toutes les dérivées partielles au premier ordre de f existent sur U et que toutes leurs dérivées partielles au premier ordre existent en x , on définit la *matrice hessienne* $H_f(x)$ de f en x par

$$[H_f(x)]_{j_1 j_2} = \partial_{j_2 j_1}^2 f(x) \in \mathbb{R}^{n \times n},$$

c'est-à-dire

$$H_f(x) = J_{\nabla f}(x) = \begin{bmatrix} \partial_{11}^2 f(x) & \partial_{21}^2 f(x) & \cdots & \partial_{n1}^2 f(x) \\ \partial_{12}^2 f(x) & \partial_{22}^2 f(x) & \cdots & \partial_{n2}^2 f(x) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_{1n}^2 f(x) & \partial_{2n}^2 f(x) & \cdots & \partial_{nn}^2 f(x) \end{bmatrix}.$$

Exercice – Laplacien et matrice hessienne (o) Soit U un ouvert de \mathbb{R}^n , $x \in U$ et $f : U \rightarrow \mathbb{R}$ une fonction dont la matrice hessienne en x est bien définie. Exprimer le laplacien de f en x , $\Delta f(x) := \sum_{i=1}^n \partial_{ii}^2 f(x)$, en fonction de $H_f(x)$. (Solution p. 25.)

Exercice – Matrice hessienne d'un monôme (•) Soit $f : (x_1, x_2) \in \mathbb{R}^2 \rightarrow \mathbb{R}$ la fonction définie par $f(x_1, x_2) = x_1 x_2^2$. Montrer que la matrice $H_f(x)$ est définie en tout point $x \in \mathbb{R}^2$ et la calculer. (Solution p. 25.)

Exercice – Matrice hessienne d'un lagrangien (•) Soit U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$ et $g : U \rightarrow \mathbb{R}$ deux applications dont les matrices hessiennes sont définies sur U . Soit $c \in \mathbb{R}$ une constante et $L : U \times \mathbb{R} \rightarrow \mathbb{R}$ la fonction telle que $L(x, \lambda) = f(x) + \lambda(g(x) - c)$. Calculer $H_L(x, \lambda)$. (Solution p. 26.)

Exercice – Matrice hessienne diagonale (••) Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction dont la matrice hessienne est définie en tout point. Montrer que sa matrice hessienne H_f est diagonale en tout point de \mathbb{R}^2 si et seulement si $f(x_1, x_2) = g(x_1) + h(x_2)$ où $g : \mathbb{R} \rightarrow \mathbb{R}$ et $h : \mathbb{R} \rightarrow \mathbb{R}$ sont des fonctions deux fois dérivables. (Solution p. 26.)

Définition – Continue différentiabilité d'ordre 2

Soit U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$. La fonction f est *deux fois continûment différentiable* si pour tout $j_1 \in \{1, \dots, n\}$ et tout $j_2 \in \{1, \dots, n\}$, la dérivée partielle d'ordre deux $\partial_{j_2 j_1}^2 f : U \rightarrow \mathbb{R}$ existe et est continue.

Alternativement, la fonction f est deux fois continûment différentiable si la fonction $x \in U \mapsto H_f(x) \in \mathbb{R}^{n \times n}$ est définie et continue.

Définition – Différentielle d'ordre 2

Soit U un ouvert de \mathbb{R}^n , $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ et $x \in U$. On dira que f est *deux fois différentiable en x* si f est différentiable sur U et si pour tout vecteur h_1 de \mathbb{R}^n , la fonction $x \in U \mapsto df(x) \cdot h_1$ est différentiable en x . La *différentielle d'ordre 2 de f en x* , notée $d^2f(x)$, est définie comme l'application linéaire telle que pour tout h_1 dans \mathbb{R}^n ,

$$d^2f(x) \cdot h_1 := d(x \mapsto df(x) \cdot h_1)(x),$$

c'est-à-dire pour tout vecteur h_2 de \mathbb{R}^n ,

$$(d^2f(x) \cdot h_1) \cdot h_2 = d(x \mapsto df(x) \cdot h_1)(x) \cdot h_2.$$

On dit que f est *deux fois différentiable (sur U)* si elle est deux fois différentiable en tout point x de U .

Dans cette définition, le caractère linéaire de $d^2f(x)$ repose sur le lemme technique suivant :

Lemme – Linéarité de $d(x \mapsto df(x) \cdot h_1)(x)$ par rapport à h_1

Soit U un ouvert de \mathbb{R}^n , $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ et $x \in U$. Si f est différentiable sur U et si pour tout vecteur h_1 de \mathbb{R}^n , la fonction $x \in U \mapsto df(x) \cdot h_1$ est différentiable en x , alors la fonction $d(x \mapsto df(x) \cdot h_1)(x)$ dépend linéairement de $h_1 \in \mathbb{R}^n$.

Démonstration Pour tous $h_1, k_1 \in \mathbb{R}^n$, par additivité de $df(x)$ et de la différentiation,

$$\begin{aligned} d(x \mapsto df(x) \cdot (h_1 + k_1))(x) &= d(x \mapsto (df(x) \cdot h_1 + df(x) \cdot k_1))(x) \\ &= d((x \mapsto df(x) \cdot h_1) + (x \mapsto df(x) \cdot k_1))(x) \\ &= d(x \mapsto df(x) \cdot h_1)(x) + d(x \mapsto df(x) \cdot k_1)(x) \end{aligned}$$

et pour tout $\alpha \in \mathbb{R}$, par homogénéité de $df(x)$ et de la différentiation,

$$\begin{aligned} d(x \mapsto df(x) \cdot (\alpha h_1))(x) &= d(x \mapsto \alpha df(x) \cdot h_1)(x) \\ &= d(\alpha(x \mapsto df(x) \cdot h_1))(x) = \alpha d(x \mapsto df(x) \cdot h_1)(x). \end{aligned}$$

■

Remarque – Applications linéaires d'ordre 2

Par construction, le terme $d(x \mapsto df(x) \cdot h_1)(x)$ est une application linéaire de \mathbb{R}^n dans \mathbb{R}^m , donc la fonction $d^2f(x)$ associe linéairement à un vecteur de \mathbb{R}^n une application linéaire de \mathbb{R}^n dans \mathbb{R} .

Notons $A \rightarrow B$ l'ensemble des applications de A dans B ; on a donc

$$d^2 f(x) \in \mathbb{R}^n \rightarrow (\mathbb{R}^n \rightarrow \mathbb{R}),$$

ce qui se décline successivement en

$$d^2 f(x) \cdot h_1 \in \mathbb{R}^n \rightarrow \mathbb{R} \text{ et } (d^2 f(x) \cdot h_1) \cdot h_2 \in \mathbb{R}^m.$$

Pour simplifier les notations, on conviendra que dans ce contexte le symbole “ \rightarrow ” associe à droite :

$$\mathbb{R}^n \rightarrow \mathbb{R}^n \rightarrow \mathbb{R} := \mathbb{R}^n \rightarrow (\mathbb{R}^n \rightarrow \mathbb{R}).$$

La convention associée veut que lors de l'application d'une fonction linéaire, le symbole “ \cdot ” associe à gauche :

$$d^2 f(x) \cdot h_1 \cdot h_2 := (d^2 f(x) \cdot h_1) \cdot h_2.$$

L'usage du “ \cdot ” doit nous rappeler que les dépendances de $df(x) \cdot h_1$ en h_1 et de $d^2 f(x) \cdot h_1 \cdot h_2$ en h_2 sont linéaires². Cela signifie pour la linéarité de $d^2 f(x)$ que $d^2 f(x) \cdot (\alpha h_1) = \alpha d^2 f(x) \cdot h_1$ et $d^2 f(x) \cdot (h_1 + k_1) = d^2 f(x) \cdot h_1 + d^2 f(x) \cdot k_1$; il s'agit d'égalités entre fonctions de $\mathbb{R}^n \rightarrow \mathbb{R}$, que nous devons donc interpréter comme :

$$\begin{aligned} d^2 f(x) \cdot (\alpha h_1) \cdot h_2 &= \alpha d^2 f(x) \cdot h_1 \cdot h_2 \\ d^2 f(x) \cdot (h_1 + k_1) \cdot h_2 &= d^2 f(x) \cdot h_1 \cdot h_2 + d^2 f(x) \cdot k_1 \cdot h_2 \end{aligned}$$

La linéarité de $d^2 f(x) \cdot h_1$ conduit quant à elle à

$$\begin{aligned} d^2 f(x) \cdot h_1 \cdot (\beta h_2) &= \beta d^2 f(x) \cdot h_1 \cdot h_2 \\ d^2 f(x) \cdot h_1 \cdot (h_2 + k_2) &= d^2 f(x) \cdot h_1 \cdot h_2 + d^2 f(x) \cdot h_1 \cdot k_2 \end{aligned}$$

L'expression $d^2 f(x) \cdot h_1 \cdot h_2$ est donc linéaire par rapport à h_1 et h_2 pris isolément ; on parlera de dépendance *bilinéaire* dans le couple (h_1, h_2) .

Remarque – Différentiation automatique d'ordre 2

La bibliothèque autograd nous a déjà permis de calculer automatiquement le gradient de fonctions scalaires et la matrice jacobienne de fonctions scalaires ou vectorielles.

```
import autograd as ag
from autograd import numpy as np
```

```
def grad(f):
    def grad_f(*x):
        n = len(x)
```

² la notation “générique” serait sinon $d^2 f(x)(h_1)$ et $d^2 f(x)(h_1)(h_2)$, voire $(d^2 f(x))(h_1)$ et $((d^2 f(x))(h_1))(h_2)$ si l'on souhaitait être tout à fait explicite.

```

        return np.array([ag.grad(f, i)(*x) for i in range(n)])
    return grad_f

def J(f):
    def J_f(*x):
        n = len(x)
        di_f_x = [ag.jacobian(f, i)(*x) for i in range(n)]
        return np.array(di_f_x).T
    return J_f

```

Autograd permet également le calcul des dérivées partielles d'ordre supérieur. Concrètement, on peut appliquer à nouveau un opérateur différentiel sur une fonction qui est issue d'un calcul fait par autograd. L'implémentation de la fonction qui calcule la matrice hessienne d'une fonction scalaire est donc particulièrement simple :

```

def H(f):
    return J(grad(f))

```

Un exemple d'usage :

```

def f(x1, x2):
    return np.exp(-0.5 * (x1 * x1 + x2 * x2))

```

exercé de la façon suivante :

```

>>> H_f = H(f)
>>> H_f(1.0, 2.0)
array([[0.          , 0.16417 ],
       [0.16417   , 0.246255]])

```

Proposition – Différentielle d'ordre 2 et matrice hessienne

Soit U un ouvert de \mathbb{R}^n , $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ et $x \in U$. La fonction f est deux fois différentiable en x si et seulement si elle est différentiable sur U et que son gradient ∇f est différentiable en x . Sa matrice hessienne est alors définie en x et pour tous $h_1, h_2 \in \mathbb{R}^n$,

$$d^2 f(x) \cdot h_1 \cdot h_2 = \langle h_1, H_f(x) \cdot h_2 \rangle = h_1^\top \cdot H_f(x) \cdot h_2 = \sum_{j_1=1}^n \sum_{j_2=1}^n [H_f(x)]_{j_1 j_2} h_{1j_1} h_{2j_2}.$$

En particulier

$$[H_f(x)]_{j_1 j_2} = d^2 f(x) \cdot e_{j_1} \cdot e_{j_2}.$$

Démonstration Si la fonction f est deux fois différentiable en x , la fonction f est différentiable donc son gradient existe. Pour tout $h_1 \in \mathbb{R}^n$, la fonction $x \mapsto df(x) \cdot h_1$ est également différentiable en x donc en particulier, pour

tout $j_1 \in \{1, \dots, n\}$, $(\nabla f(x))_{j_1} = \langle \nabla f(x), e_{j_1} \rangle = df(x) \cdot e_{j_1}$; le gradient de f est différentiable composante par composante et donc différentiable. Réciproquement, si f est différentiable et que son gradient est différentiable en x , pour tout $h \in \mathbb{R}^n$ on a

$$df(x) \cdot h_1 = df(x) \cdot \left(\sum_{j_1=1}^n h_{1j_1} e_{j_1} \right) = \sum_{j_1=1}^n h_{1j_1} df(x) \cdot e_{j_1} = \sum_{j_1=1}^n h_{1j_1} (\nabla f(x))_{j_1};$$

la fonction $x \mapsto (df(x) \cdot h)$ est donc différentiable en x comme combinaison linéaire de fonction différentiables en x .

Par définition, $[H_f(x)]_{j_1 j_2}(x) = \partial_{j_2 j_1}^2 f(x) = \partial_{j_2}(\partial_{j_1} f)(x)$ et donc

$$[H_f(x)]_{j_1 j_2}(x) = \partial_{j_2}(x \mapsto df(x) \cdot e_{j_1})(x) = d(x \mapsto df(x) \cdot e_{j_1})(x) \cdot e_{j_2},$$

c'est-à-dire $[H_f(x)]_{j_1 j_2}(x) = d^2 f(x) \cdot e_{j_1} \cdot e_{j_2}$. Pour prouver l'égalité restante, on exploite la linéarité de $d^2 f(x) \cdot h_1 \cdot h_2$ par rapport à h_1 et à h_2 :

$$\begin{aligned} d^2 f(x) \cdot h_1 \cdot h_2 &= d^2 f(x) \cdot \left(\sum_{j_1=1}^n h_{1j_1} e_{j_1} \right) \cdot \left(\sum_{j_2=1}^n h_{2j_2} e_{j_2} \right) \\ &= \sum_{j_2=1}^n h_{2j_2} \left(d^2 f(x) \cdot \left(\sum_{j_1=1}^n h_{1j_1} e_{j_1} \right) \cdot e_{j_2} \right) \\ &= \sum_{j_1=1}^n \sum_{j_2=1}^n h_{1j_1} h_{2j_2} (d^2 f(x) \cdot e_{j_1} \cdot e_{j_2}) \\ &= \sum_{j_1=1}^n \sum_{j_2=1}^n [H_f(x)]_{j_1 j_2} h_{1j_1} h_{2j_2}. \end{aligned}$$

■

Proposition – Continue différentiabilité et différentiabilité d'ordre 2

Soit U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$. Si f est deux fois continûment différentiable, alors f est deux fois différentiable.

Démonstration La fonction f est différentiable à l'ordre 2 si elle est différentiable et que son gradient est également différentiable (p. 7). Or, si f est deux fois continûment différentiable, tous les dérivées partielles à l'ordre 1 de ∇f existent et sont elles-mêmes partiellement dérivables, de dérivées partielles continues. Donc, le gradient de f est continûment différentiable et donc différentiable. En particulier, il est continu, la fonction f est donc continûment différentiable et donc différentiable. ■

Proposition – Développement limité du gradient

Soit U un ouvert de \mathbb{R}^n , $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ et $x \in U$. Si la fonction f est deux fois différentiable en x alors

$$\nabla f(x+h) = \nabla f(x) + H_f(x) \cdot h + \varepsilon(h)\|h\|$$

où $\lim_{h \rightarrow 0} \varepsilon(h) = 0$.

Démonstration D'après la proposition “Différentielle d'ordre 2 et matrice hessienne” (p. 7), ∇f existe et est différentiable en x . Par conséquent, ∇f admet un développement limité au 1er ordre en x :

$$\nabla f(x+h) = \nabla f(x) + J_{\nabla f}(x) \cdot h + \varepsilon(h)\|h\|.$$

D'après la définition de la matrice hessienne (p. 4), $H_f(x) = J_{\nabla f}(x)$ d'où l'égalité de l'énoncé. ■

Théorème – Symétrie de la différentielle d'ordre 2

Soit $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction deux fois différentiable en un point x de U . Pour tout couple de vecteurs h_1 et h_2 de \mathbb{R}^n , on a

$$d^2 f(x) \cdot h_1 \cdot h_2 = d^2 f(x) \cdot h_2 \cdot h_1,$$

ou de façon équivalente, la matrice hessienne de f en x est symétrique

$$H_f(x)^\top = H_f(x),$$

c'est-à-dire, pour tous $j_1, j_2 \in \{1, \dots, n\}$,

$$\partial_{j_2 j_1}^2 f(x) = \partial_{j_1 j_2}^2 f(x).$$

Démonstration Notons au préalable que

$$\begin{aligned} \Delta^2 f(x, h_1, h_2) &:= (f(x+h_2+h_1) - f(x+h_2)) - (f(x+h_1) - f(x)) \\ &= f(x+h_1+h_2) - f(x+h_1) - f(x+h_2) + f(x) \\ &= (f(x+h_2+h_1) - f(x+h_1)) - (f(x+h_2) - f(x)) \\ &= \Delta^2 f(x, h_2, h_1). \end{aligned}$$

La variation d'ordre 2 de f en x est donc symétrique par rapport à ses arguments h_1 et h_2 . On peut alors exploiter la relation entre variation d'ordre 2 et différentielle d'ordre 2 (p. 22) en notant que

$$\begin{aligned} \|d^2 f(x) \cdot h_1 \cdot h_2 - d^2 f(x) \cdot h_2 \cdot h_1\| &\leq \\ \| \Delta^2 f(x, h_1, h_2) - d^2 f(x) \cdot h_1 \cdot h_2 \| &+ \| \Delta^2 f(x, h_2, h_1) - d^2 f(x) \cdot h_1 \cdot h_2 \|. \end{aligned}$$

On obtient pour tout $\varepsilon > 0$ et quand h_1 et h_2 sont suffisamment petits,

$$\|d^2 f(x) \cdot h_1 \cdot h_2 - d^2 f(x) \cdot h_2 \cdot h_1\| \leq 2\varepsilon(\|h_1\| + \|h_2\|)^2.$$

Si h_1 et h_2 sont arbitraires, en substituant th_1 à h_1 et th_2 à h_2 pour un $t > 0$ suffisamment petit pour que l'inégalité ci-dessus soit valable, comme

$$d^2 f(x) \cdot th_1 \cdot th_2 - d^2 f(x) \cdot th_2 \cdot th_1 = t^2 \times (d^2 f(x) \cdot h_1 \cdot h_2 - d^2 f(x) \cdot h_2 \cdot h_1)$$

et

$$2\varepsilon(\|th_1\| + \|th_2\|)^2 = t^2 \times 2\varepsilon(\|h_1\| + \|h_2\|)^2,$$

on voit que l'inégalité est en fait valable pour des h_1 et h_2 arbitraires. On en déduit que $d^2 f(x) \cdot h_1 \cdot h_2 - d^2 f(x) \cdot h_2 \cdot h_1 = 0$. ■

Exercice – Analyse vectorielle (•) Soit U un ouvert de \mathbb{R}^3 et $f : U \rightarrow \mathbb{R}^3$. On note (quand les expressions ont du sens)

$$\operatorname{div} f(x) := \partial_1 f_1(x) + \partial_2 f_2(x) + \partial_3 f_3(x) \quad \text{et} \quad \operatorname{rot} f(x) := \begin{bmatrix} \partial_2 f_3(x) - \partial_3 f_2(x) \\ \partial_3 f_1(x) - \partial_1 f_3(x) \\ \partial_1 f_2(x) - \partial_2 f_1(x) \end{bmatrix}.$$

Soient $f : U \rightarrow \mathbb{R}^3$ et $g : U \rightarrow \mathbb{R}$ des fonctions deux fois différentiable en $x \in U$. Calculer $\operatorname{div}(\operatorname{rot} f)(x)$ et $\operatorname{rot}(\nabla f)(x)$ (Solution p. 26.)

Exercice – Gradient unitaire (•) Soit U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$ une fonction deux fois différentiable. Montrer que si $\|\nabla f\| = 1$, alors $H_f \cdot \nabla f = 0$. (Solution p. 27.)

Proposition – Développement limité à l'ordre 2

Soit U un ouvert de \mathbb{R}^n , $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ et $x \in U$. Si la fonction f est deux fois différentiable en x alors

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

où $\lim_{h \rightarrow 0} \varepsilon(h) = 0$.

Démonstration Il s'agit de montrer que pour tout $\varepsilon > 0$, on peut trouver un seuil $r > 0$ tel que si $\|h\| \leq r$, alors

$$\left\| f(x+h) - f(x) - \langle \nabla f(x), h \rangle - \frac{1}{2} \langle h, H_f(x) \cdot h \rangle \right\| \leq \varepsilon \|h\|^2.$$

La fonction $g : h \mapsto f(x+h) - f(x) - \langle \nabla f(x), h \rangle - \frac{1}{2} \langle h, H_f(x) \cdot h \rangle \in \mathbb{R}$ est différentiable, de gradient en h

$$\nabla g(h) = \nabla f(x+h) - \nabla f(x) - \left(\frac{H_f(x) + H_f(x)^\top}{2} \right) \cdot h,$$

c'est-à-dire, comme la matrice hessienne est symétrique (p. 9),

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

Compte tenu du développement limité du gradient de f en x (p. 9), il existe un seuil $r > 0$ tel que pour tout k tel que $\|k\| \leq r$,

$$\|\nabla g(k)\| = \|\nabla f(x+k) - \nabla f(x) - H_f(x) \cdot k\| \leq \varepsilon \|k\|.$$

Par l'inégalité des accroissements finis, quand $\|h\| \leq r$, on a donc

$$\begin{aligned} \|g(h)\| &= \|g(h) - g(0)\| \leq \sup_{k \in [0, h]} \|dg(k)\| \times \|h\| \\ &= \sup_{k \in [0, h]} \|\nabla g(k)\| \times \|h\| \\ &\leq \sup_{k \in [0, h]} \varepsilon \|k\| \times \|h\| \\ &\leq \varepsilon \|h\|^2. \end{aligned}$$

■

Proposition – Développement de Taylor d'ordre 1 avec reste intégral

Soit U un ouvert de \mathbb{R}^n , $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$, $x \in U$ et $h \in \mathbb{R}^n$ tel que $[x, x+h] \subset U$. Si la fonction f est deux fois continûment différentiable, alors

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

Démonstration Définissons la fonction $\phi : [0, 1] \rightarrow \mathbb{R}$ par $\phi(t) = f(x+th)$. Par le théorème fondamental du calcul, puis par intégration par parties, on obtient

$$\begin{aligned} \phi(1) &= \phi(0) + \int_0^1 \phi'(t) dt \\ &= \phi(0) + [\phi'(t)(t-1)]_0^1 - \int_0^1 \phi''(t) \times (t-1) dt \\ &= \phi(0) + \phi'(0) + \int_0^1 \phi''(t) \times (1-t) dt \end{aligned}$$

Or, on a $\phi(0) = f(x)$, $\phi(1) = f(x+h)$; par la règle de dérivation en chaîne on obtient également $\phi'(t) = df(x+th) \cdot h$ et $\phi''(t) = d^2f(x+th) \cdot h \cdot h$. Par conséquent

$$f(x+h) = f(x) + df(x) \cdot h + \int_0^1 (d^2f(x+th) \cdot h \cdot h) \times (1-t) dt,$$

ce qui est équivalent à l'équation recherchée. ■

Exercice – Accroissements finis d'ordre 2 (•) Montrer que si $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est deux fois continûment différentiable et que la norme d'opérateur de H_f est bornée par la constante M sur \mathbb{R}^n , alors pour tous $x, h \in \mathbb{R}^n$,

$$\|f(x+h) - f(x) - \langle \nabla f(x), h \rangle\| \leq M \frac{\|h\|^2}{2}.$$

(Solution p. 27.)

Différentielle d'ordre supérieur

Définition – Tenseur d'ordre n

On appellera *tenseur d'ordre $n \in \mathbb{N}$* et de *type $(m_1, m_2, \dots, m_n) \in \mathbb{N}^n$* un élément de $\mathbb{R}^{m_1 \times m_2 \times \dots \times m_n}$ c'est-à-dire un réel

$$A_{i_1 i_2 \dots i_n} \in \mathbb{R},$$

indexé par n indices

$$(i_1, i_2, \dots, i_n) \in \{1, \dots, m_1\} \times \{1, \dots, m_2\} \times \dots \times \{1, \dots, m_n\}.$$

Remarque – Scalaires, vecteurs, matrices

Le concept de tenseur englobe et généralise les scalaires, vecteurs et matrices :

1. Les scalaires sont les tenseurs d'ordre 0 (ils ne dépendent d'aucun indice). Il n'existe qu'un type de tenseur d'ordre 0 : $()$ (l'unique 0-uplet).
2. les vecteurs sont les tenseurs d'ordre 1 ; le type d'un vecteur de \mathbb{R}^m est (m) (le 1-uplet contenant m).
3. les matrices les tenseurs d'ordre 2 ; le type d'une matrice de $\mathbb{R}^{m \times n}$ est (m, n) (la paire contenant m et n).

Les tenseurs d'ordre $n \geq 3$ généralisent ces constructions.

Remarque – Les tenseurs avec NumPy

Les tenseurs sont des tableaux n -dimensionnels ; ils sont donc représentés comme des instances du type `array` de NumPy. Leur ordre est donnée par la méthode `ndim` (nombre de dimensions), leur type par la méthode `shape`. Ainsi, avec

```
>>> T0 = np.array(1.0)
>>> T1 = np.array([1.0, 2.0, 3.0])
>>> T2 = np.array([[1.0, 2.0, 3.0], [4.0, 5.0, 6.0]])
>>> T3 = np.array([[[1.0], [2.0], [3.0]], [[4.0], [5.0], [6.0]]])
```

on a

```
>>> (T0.ndim, T1.ndim, T2.ndim, T3.ndim)
(0, 1, 2, 3)
```

et

```
>>> T0.shape
()
>>> T1.shape
(3,)
>>> T2.shape
(2, 3)
>>> T3.shape
(2, 3, 1)
```

Les coefficients d'un tenseur `T` s'obtiennent au moyen du crochet `T[]` (méthode `__getitem__`)³, mais avec un indexation commençant à 0 et non 1 comme la convention mathématique classique. Ainsi :

```
>>> T1[1]
2.0
>>> T2[1,2]
6.0
>>> T3[1,2,0]
6.0
```

3. La notation `T2[(1,2)]` (n -uplet explicite) est équivalente à `T2[1,2]` (n -uplet implicite). Cette remarque est utile pour accéder au contenu des tenseurs d'ordre 0, car la notation `T0[]` n'est pas acceptée :

```
>>> T0[()]
1.0
>>> T0[]
Traceback (most recent call last):
...
SyntaxError: invalid syntax
```

Remarque – Applications linéaires d'ordre supérieur

La raison d'être des matrices de $\mathbb{R}^{m \times n}$ est de représenter concrètement les applications linéaires de $\mathbb{R}^n \rightarrow \mathbb{R}^m$, dont l'espace est noté $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Si l'on désigne par A une telle l'application linéaire et par $[a_{ij}]_{ij}$ la matrice associée,

$$a_{ij} = (A \cdot e_j)_i \text{ et } A \cdot x = \sum_i \left(\sum_j a_{ij} x_j \right) e_i$$

pour tout $x \in \mathbb{R}^n$. (Par abus de notation, e_j désigne le j -ème vecteur de la base canonique de \mathbb{R}^p quel que soit p). Cette correspondance légitime l'identification fréquemment opérée entre l'application linéaire A et la matrice $[a_{ij}]_{ij}$.

Une correspondance similaire existe pour les tenseurs d'ordre supérieur à 2. Ainsi, à l'ordre 3, on peut mettre en correspondance un tenseur $(t_{ijk})_{ijk}$ de type (m, n, p) et une application linéaire T de \mathbb{R}^p dans l'espace des applications linéaires de \mathbb{R}^n dans \mathbb{R}^m , c'est-à-dire établir la correspondance

$$(t_{ijk})_{ijk} \in \mathbb{R}^{m \times n \times p} \longleftrightarrow T \in \mathcal{L}(\mathbb{R}^p, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$$

de la façon suivante :

$$t_{ijk} = ((T \cdot e_k) \cdot e_j)_i \text{ et } (T \cdot x) \cdot y = \sum_i \left(\sum_j \left(\sum_k t_{ijk} x_k \right) y_j \right) e_i$$

(sous-entendu, pour tout $x \in \mathbb{R}^p$ et $y \in \mathbb{R}^n$). Cette représentation du tenseur d'ordre 3 par une application linéaire est dite d'ordre supérieur car les valeurs des applications (linéaires) en question sont elles-mêmes des applications (linéaires).

Le processus décrit dans ce paragraphe se généralise à des tenseurs d'ordre supérieur à 3.

Remarque – Applications multilinéaires

Les tenseurs vus comme des tableaux permettent de représenter d'autres objets mathématiques, équivalents aux applications linéaires (d'ordre supérieur). A titre d'exemple, si l'on considère les tenseurs d'ordre 2, une matrice $[a_{ij}]_{ij} \in \mathbb{R}^{m \times n}$ correspond à une application $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ mais également à une forme bilinéaire $Q \in \mathcal{L}_2(\mathbb{R}^n \times \mathbb{R}^m, \mathbb{R})$, c'est-à-dire une fonction de deux variables dans \mathbb{R}^n et \mathbb{R}^m , linéaire par rapport à chacune de ces variables indépendamment et à valeurs dans \mathbb{R} . Cette forme bilinéaire B est donnée par

$$B(x, y) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_i y_j.$$

Dans le cas général, un tenseur d'ordre n correspond à une forme n -linéaire.

Définition – Produit tensoriel

Soient A et B des tenseurs de type respectifs $(m_1, m_2, \dots, m_n) \in \mathbb{N}^n$ et $(p_1, p_2, \dots, p_q) \in \mathbb{N}^q$. Si $m_n = p_1$, le produit de A et B est le tenseur de type $(m_1, \dots, m_{n-1}, p_2, \dots, p_q) \in \mathbb{N}^{n+q-2}$ noté $A \cdot B$ défini par

$$(A \cdot B)_{i_1 \dots i_{n-1} i_{n+1} \dots i_{n+q}} = \sum_{i_n=1}^{m_n} A_{i_1 i_2 \dots i_n} B_{i_n i_{n+1} \dots i_{n+q}}.$$

Remarque – Produits tensoriels classiques

Pour $x, y \in \mathbb{R}^n$, on a

$$x \cdot y = \sum_{i=1}^n x_i y_i \in \mathbb{R}.$$

Le produit tensoriel de deux vecteurs est bien défini et coïncide avec leur produit scalaire⁴. Si de plus $A \in \mathbb{R}^{m \times n}$ et $B \in \mathbb{R}^{n \times p}$,

$$A \cdot x \in \mathbb{R}^m, (A \cdot x)_i = \sum_{j=1}^n A_{ij} x_j$$

$$A \cdot B \in \mathbb{R}^{m \times p} \text{ et } (A \cdot B)_{ik} = \sum_{j=1}^n A_{ij} B_{jk}.$$

Autrement dit, les produits tensoriels matrices-vecteurs et matrices-matrices coïncident avec les produits classiques de l'algèbre linéaire.

Remarque – Produit tensoriel avec NumPy

Si A et B sont deux tenseurs de type compatibles pour le produit (la dernière dimension de A égale à la première dimension de B) représentés par les tableaux n -dimensionnels `A` et `B`, **et tant que l'ordre de B est inférieur ou égal à 2**, on peut calculer le produit tensoriel de A et B au moyen de la méthode `dot`. Par exemple, avec :

```
>>> x = np.array([0.0, 1.0])
>>> y = np.array([2.0, 4.0])
>>> A = np.array([[1.0, 2.0], [3.0, 4.0]])
>>> B = np.array([[5.0, 6.0], [7.0, 8.0]])
>>> T = np.array([[1.0, 2.0], [3.0, 4.0]],
...               [[5.0, 6.0], [7.0, 8.0]])
```

4. Souvenons-nous à l'inverse que si l'on interprète “.” comme un produit matriciel et que l'on représente implicitement x et y comme deux vecteurs-colonnes de $\mathbb{R}^{n \times 1}$, l'expression $x \cdot y$ n'a pas de sens ; il alors considérer $x^\top \cdot y$ à la place, puis assimiler ensuite le résultat – qui est une matrice 1×1 – à un nombre réel. Les conventions du calcul tensoriel ont donc ici une action simplificatrice.

on obtient des produits tensoriels variés par les appels :

```
>>> x.dot(y)
4.0
>>> A.dot(x)
array([2., 4.])
>>> A.dot(B)
array([[19., 22.],
       [43., 50.]])
>>> T.dot(A)
array([[[ 7., 10.],
        [15., 22.]],
       <BLANKLINE>
       [[23., 34.],
        [31., 46.]]])
```

Par contre, si l'ordre de B est 3 ou plus, on ne pourra pas utiliser cette méthode pour calculer le produit tensoriel $A \cdot B$ car son résultat diffère du produit tensoriel tel que nous l'avons défini (cf. documentation de `numpy.dot`).

```
>>> x.dot(T) # Not what we expect here!
array([[3., 4.],
       [7., 8.]])
```

Une option consiste alors (dans ce cas particulier ou systématiquement) à utiliser la fonction NumPy `tensor_dot` avec l'option `axes=1` :

```
>>> def dot(A, B): # Let's define our own tensor product
...     return np.tensordot(A, B, axes=1)
>>> dot(x, T) # Problem solved!
array([[5., 6.],
       [7., 8.]])
```

Pour un contrôle plus fin des opérations tensorielles, on pourra également avoir recours à la fonction `numpy.einsum`, une fonction qui exploite la *convention de sommation (des indices répétés) d'Einstein*. Pour calculer $(x \cdot T)_{jk} = \sum_i x_i T_{ijk}$, comme nous le souhaitons :

```
>>> np.einsum("i, ijk -> jk", x, T)
array([[5., 6.],
       [7., 8.]])
```

Ici le tenseur calculé par `x.dot(T)` correspond à $S_{jk} = \sum_i x_i T_{jik}$; si c'est ce que l'on souhaite, on peut aussi l'obtenir par :

```
>>> np.einsum("i, jik -> jk", x, T)
array([[3., 4.],
       [7., 8.]])
```

Armés de la notion d'application linéaire d'ordre supérieure, et de sa représentation concrète comme tenseur, nous pouvons désormais généraliser la notion de

différentielle à un ordre $k \geq 2$ arbitraire, pour des fonctions à valeurs scalaires ou vectorielles de \mathbb{R}^m , par induction sur l'ordre de la différentielle.

Définition – Différentielle d'ordre k

Soit $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction différentiable à l'ordre $k - 1$ dans un voisinage d'un point x de U . On dira que f est k fois différentiable en x si pour tous vecteurs h_1, \dots, h_{k-1} de \mathbb{R}^n , la fonction

$$x \mapsto d^{k-1}f(x) \cdot h_1 \cdot h_2 \cdot \dots \cdot h_{k-1}$$

est différentiable en x . La *différentielle d'ordre k de f en x* , notée $d^k f(x)$ est définie comme l'application linéaire telle que pour tout h_1, \dots, h_{k-1} de \mathbb{R}^n ,

$$d^k f(x) \cdot h_1 \cdot h_2 \cdot \dots \cdot h_{k-1} := d(x \mapsto d^{k-1}f(x) \cdot h_1 \cdot h_2 \cdot \dots \cdot h_{k-1})(x)$$

ou de façon équivalente

$$d^k f(x) \cdot h_1 \cdot h_2 \cdot \dots \cdot h_{k-1} \cdot h_k := d(x \mapsto d^{k-1}f(x) \cdot h_1 \cdot h_2 \cdot \dots \cdot h_{k-1})(x) \cdot h_k$$

Les dérivées partielles d'ordre supérieur – qui se définissent par récurrence – permettent vont permettant d'explicitier les différentielles d'ordre supérieures comme une tenseur.

Définition – Dérivées partielles d'ordre k

Soient U un ouvert de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}^m$ et $x \in U$. Soient i_1, i_2, \dots, i_k des indices de $\{1, \dots, n\}$; lorsque la dérivée partielle $\partial_{i_2 \dots i_k}^{k-1} f$ est définie en tout point de U et est différentiable par rapport à la i_1 -ème variable en x , on définit

$$\partial_{i_1 \dots i_k}^k f(x) := \partial_{i_1}(\partial_{i_2 \dots i_k}^{k-1} f)(x).$$

Proposition – Calcul des dérivées partielles d'ordre k

Soient U un ouvert de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}^m$ et $x \in U$. Si f est k fois différentiable en x , alors pour tout i_1, i_2, \dots, i_k dans $\{1, \dots, n\}$,

$$\partial_{i_1 i_2 \dots i_k}^k f(x) = d^k f(x) \cdot e_{i_k} \cdot \dots \cdot e_{i_2} \cdot e_{i_1}.$$

Démonstration A l'ordre 1, $\partial_{i_1}^1 f(x) = \partial_{i_1} f(x)$; l'égalité $\partial_{i_1} f(x) = df(x) \cdot e_{i_1}$ a été démontrée dans le chapitre “Calcul Différentiel I”. Supposons que l'égalité

soit vraie à l'ordre $k - 1$. Alors,

$$\begin{aligned}\partial_{i_1 \dots i_k}^k f(x) &= \partial_{i_1}(\partial_{i_2 \dots i_k}^{k-1} f)(x) \\ &= \partial_{i_1}(x \mapsto d^{k-1} f(x) \cdot e_{i_k} \cdot \dots \cdot e_{i_2}) \\ &= d(x \mapsto d^{k-1} f(x) \cdot e_{i_k} \cdot \dots \cdot e_{i_2}) \cdot e_{i_1} \\ &= d^k f(x) \cdot e_{i_k} \cdot \dots \cdot e_{i_2} \cdot e_{i_1}\end{aligned}$$

et l'égalité est également vraie à l'ordre k . ■

Remarque – Différentielle d'ordre k et tenseur

On a

$$d^k f(x) \in \overbrace{\mathbb{R}^n \rightarrow \mathbb{R}^n \rightarrow \dots \rightarrow \mathbb{R}^n}^{k \text{ termes}} \rightarrow \mathbb{R}^m,$$

chaque application dans la chaîne étant linéaire. La différentielle $d^k f(x)$ peut donc être représentée concrètement par un tenseur T d'ordre $k + 1$ et de type (m, n, \dots, n) :

$$T_{i_1 i_2 \dots i_{k+1}} := (\partial^k f_{i_2 \dots i_{k+1}}(x)) \cdot e_{i_1} = (d^k f(x) \cdot e_{i_{k+1}} \cdot \dots \cdot e_{i_2}) \cdot e_{i_1}.$$

Par linéarité par rapport à chacun des h_i , on a :

$$d^k f(x) \cdot h_{k+1} \cdot \dots \cdot h_2 = \sum_{i_1, i_2, \dots, i_{k+1}} (\partial^k f_{i_1 i_2 \dots i_{k+1}}(x) \times h_{k+1 i_{k+1}} \times \dots \times h_{2 i_2}) e_{i_1}.$$

Lemme – Stratification

Si $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ est une fonction k fois différentiable en un point x de U , pour tous vecteurs h_1, h_2, \dots, h_k de \mathbb{R}^n , et tout $p \in \{0, \dots, k\}$, on a

$$d^k f(x) \cdot h_1 \cdot \dots \cdot h_k = d^{k-p}(x \mapsto d^p f(x) \cdot h_1 \cdot \dots \cdot h_p)(x) \cdot h_{p+1} \cdot \dots \cdot h_k.$$

Démonstration Faisons l'hypothèse que le théorème est satisfait lorsque la fonction est j fois différentiable pour tout $j \leq k$. C'est de toute évidence le cas pour $k = 0, 1, 2$; montrons qu'il est encore vrai pour $j = k + 1$.

Notons tout d'abord que si $p = 0$, le résultat est évident; on supposera donc dans la suite que $p \in \{1, \dots, k + 1\}$. Par définition des différentielles d'ordre supérieur (p. 17),

$$d^{k+1} f(x) \cdot h_1 \cdot \dots \cdot h_{k+1} = d(d^k f(x) \cdot h_1 \cdot \dots \cdot h_k) \cdot h_{k+1}.$$

Or, par l'hypothèse de récurrence à l'ordre k ,

$$d^k f(x) \cdot h_1 \cdot \dots \cdot h_k = d^{k-p}(d^p f(x) \cdot h_1 \cdot \dots \cdot h_p) \cdot h_{p+1} \cdot \dots \cdot h_k$$

donc si l'on pose $g(x) = d^p f(x) \cdot h_1 \cdot \dots \cdot h_p$ et que l'on applique l'hypothèse de récurrence à l'ordre $k+1-p$ (un nombre compris entre 0 et k), on obtient

$$\begin{aligned} d^{k+1} f(x) \cdot h_1 \cdot \dots \cdot h_{k+1} &= d(d^{k-p} g(x) \cdot h_{p+1} \cdot \dots \cdot h_k) \cdot h_{k+1} \\ &= d^{k+1-p} g(x) \cdot h_{p+1} \cdot \dots \cdot h_k \cdot h_{k+1} \end{aligned}$$

et donc au final

$$d^{k+1} f(x) \cdot h_1 \cdot \dots \cdot h_{k+1} = d^{k+1-p}(d^p f(x) \cdot h_1 \cdot \dots \cdot h_p) \cdot h_{p+1} \cdot \dots \cdot h_k \cdot h_{k+1}.$$

L'hypothèse de récurrence est donc prouvée au rang $k+1$, ce qui établit le résultat. ■

Proposition – Symétrie des différentielles d'ordre supérieur

Soit $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction k fois différentiable en un point x de U . Pour toute permutation σ de $\{1, \dots, n\}$ et pour tous vecteurs h_1, h_2, \dots, h_k de \mathbb{R}^n , on a :

$$d^k f(x) \cdot h_{\sigma(1)} \cdot \dots \cdot h_{\sigma(i)} \cdot \dots \cdot h_{\sigma(k)} = d^k f(x) \cdot h_1 \cdot \dots \cdot h_i \cdot \dots \cdot h_k.$$

Démonstration Toute permutation peut être décomposée en une succession de transpositions τ_{ij} , où $\tau_{ij}(i) = j$, $\tau_{ij}(j) = i$ et $\tau_{ij}(k) = k$ si k diffère de i et de j . Il suffit donc d'établir le résultat quand σ est une transposition. Nous procédons par récurrence sur k . Le résultat dans le cas $k=2$ résulte de la symétrie de la différentielle d'ordre 2 (p. 9). Supposons désormais le résultat établi au rang $k \geq 2$. En utilisant la stratification de $d^{k+1} f(x) \cdot h_1 \cdot \dots \cdot h_k \cdot h_{k+1}$ pour $p=1$ et $p=k$ (p. 18), on peut établir le résultat si i et j appartiennent tous les deux à $\{2, \dots, k+1\}$ ou à $\{1, \dots, k\}$. Dans l'unique cas restant, on peut décomposer $\tau_{1(k+1)}$ en $\tau_{2(k+1)} \circ \tau_{12} \circ \tau_{2(k+1)}$ et se ramener au cas précédent. ■

Remarque – Dérivées partielles d'ordre supérieur et multi-indices

Pour compacter la notation $\partial_{i_1 \dots i_k}^k f(x)$, on peut exploiter le fait que si f est k fois différentiable en x ,

$$\partial_{i_1 \dots i_k}^k f(x) = d^k f(x) \cdot e_{i_1} \cdot \dots \cdot e_{i_k}.$$

Compte tenu de la symétrie de $d^k f(x)$, peu importe l'ordre de i_1, \dots, i_k , seul le nombre de fois où un indice apparaît compte. Cette remarque fonde une notation basée sur les multi-indices $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$ où α_i détermine le nombre de fois où l'indice i apparaît. Formellement, le symbole $\partial^\alpha f(x)$ désigne $f(x)$ si $\alpha = (0, \dots, 0)$ et dans le cas contraire :

$$\partial^{(\alpha_1, \dots, \alpha_i+1, \dots, \alpha_n)} f(x) = \partial_i(\partial^\alpha f)(x).$$

Puissance symbolique

Comme les différentielles d'ordre supérieure sont fréquemment évaluées lorsque les termes h_1, h_2, \dots , sont égaux, on adoptera la notation (purement syntaxique) suivante :

$$(\cdot h)^k := \overbrace{\cdot h \cdot \dots \cdot h}^{k \text{ termes}}.$$

Théorème – Développement limité d'ordre supérieur

Soit $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction j fois différentiable au point $x \in U$. Alors

$$f(x+h) = \sum_{i=0}^j \frac{d^i f(x)}{i!} (\cdot h)^i + \varepsilon(h) \times \|h\|^j.$$

où $\varepsilon(h) \rightarrow 0$ quand $h \rightarrow 0$.

Démonstration Le résultat est clair pour $j = 0$. Supposons le vrai à un rang $j - 1$ arbitraire pour toute fonction $j - 1$ fois différentiable et supposons que f est j fois différentiable. Formons le reste d'ordre j associé à f :

$$r(h) = f(x+h) - \sum_{i=0}^j \frac{d^i f(x)}{i!} (\cdot h)^i.$$

Il nous faut montrer que $r(h)$ est de la forme $\varepsilon(h) \times \|h\|^j$ où $\varepsilon(h) \rightarrow 0$ quand $h \rightarrow 0$, ce qui nous allons accomplir en établissant que $\|dr(h)\| = \varepsilon'(h) \times \|h\|^{j-1}$ avec $\varepsilon'(h) \rightarrow 0$ quand $h \rightarrow 0$. En effet, si c'est le cas, $dr(h) = E(h)\|h\|^{j-1}$ où l'application linéaire E tend vers 0 quand h tend vers 0, et pour tout $\varepsilon > 0$ et h assez proche de 0 on a $\|E(h)\| \leq \varepsilon$ et donc par l'inégalité des accroissements finis,

$$\|r(h)\| = \|r(h) - r(0)\| \leq \varepsilon \|h\|^{j-1} \times \|h\| = \varepsilon \|h\|^j,$$

ce qui établit que $r(h) = \varepsilon(h) \times \|h\|^j$ avec $\varepsilon(h) \rightarrow 0$ quand $h \rightarrow 0$.

Etablissons donc ce résultat. Les termes $d^i f(x) \cdot h_1 \cdot \dots \cdot h_i$ sont linéaires par rapport à chacun des h_j , donc pour tout vecteur k , compte tenu de la symétrie de $d^i f(x)$,

$$d^i f(x)(\cdot (h+k))^i = d^i f(x)(\cdot h)^i + i d^i f(x)(\cdot h)^{i-1} \cdot k + \varepsilon(k) \|k\|.$$

La différentielle de $h \mapsto d^i f(x)(\cdot h)^i$ vaut donc $i d^i f(x)(\cdot h)^{i-1}$ et

$$dr(h) \cdot k = df(x+h) \cdot k - df(x) \cdot k - d^2 f(x) \cdot h \cdot k - \dots - \frac{d^i f(x)}{(i-1)!} (\cdot h)^{i-1} \cdot k.$$

Par le lemme de stratification (p. 18) et la symétrie des différentielles d'ordre supérieur (p. 19), on obtient

$$\begin{aligned} dr(h) \cdot k &= df(x+h) \cdot k - df(x) \cdot k \\ &\quad - d(x \mapsto df(x) \cdot k)(x) \cdot h - \dots - \frac{d^{i-1}(x \mapsto df(x) \cdot k)(x)}{(i-1)!}(\cdot h)^{i-1}. \end{aligned}$$

soit en posant $\phi(x) = df(x) \cdot k$,

$$dr(h) \cdot k = \phi(x+h) - \phi(x) - d\phi(x) \cdot h - \dots - \frac{d^{i-1}\phi(x)}{(i-1)!}(\cdot h)^{i-1}.$$

L'hypothèse de récurrence nous garantit donc que $dr(h) \cdot k = \varepsilon(h)\|h\|^{j-1}$ à k fixé, ce qui, combiné avec la linéarité de $dr(h)$, fournit $\|dr(h)\| = \varepsilon(h)\|h\|^{j-1}$. ■

Développement de Taylor avec reste intégral I

Soit $f : [a, a+h] \rightarrow \mathbb{R}^m$ où $a \in \mathbb{R}$, $h \in [0, +\infty[$. Si f est $j+1$ fois dérivable sur $[a, a+h]$,

$$f(a+h) = \sum_{i=0}^n \frac{f^{(i)}(a)}{i!} h^i + \int_a^{a+h} \frac{f^{(j+1)}(t)}{j!} (a+h-t)^j dt.$$

Démonstration A l'ordre $j=0$, la relation à prouver est

$$f(a+h) = f(a) + \int_a^{a+h} f'(t) dt$$

qui n'est autre que le théorème fondamental du calcul. Si l'on suppose la relation vérifiée à l'ordre j , et f $j+2$ fois dérivable, par intégration par parties, on obtient

$$\begin{aligned} \int_a^{a+h} f^{(j+1)}(t) \frac{(a+h-t)^j}{j!} dt &= \\ &\quad \left[f^{(j+1)}(t) \times \left(-\frac{(a+h-t)^{j+1}}{(j+1)!} \right) \right]_a^{a+h} \\ &\quad - \int_a^{a+h} f^{(j+2)}(t) \left(-\frac{(a+h-t)^{j+1}}{(j+1)!} \right) dt, \end{aligned}$$

soit

$$\begin{aligned} \int_a^{a+h} f^{(j+1)}(t) \frac{(a+h-t)^j}{j!} dt &= \\ &\quad f^{(j+1)}(a) \times \frac{h^{j+1}}{(j+1)!} + \int_a^{a+h} f^{(j+2)}(t) \frac{(a+h-t)^{j+1}}{(j+1)!} dt, \end{aligned}$$

ce qui achève la preuve par récurrence. ■

Développement de Taylor avec reste intégral II

Si $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ est $j + 1$ fois différentiable et $[a, a + h] \subset U$,

$$f(a + h) = \sum_{i=0}^j \frac{df^{(i)}(a)}{i!} (\cdot h)^i + \int_0^1 \frac{df^{(j+1)}(a + th)}{j!} (\cdot h)^{j+1} (1-t)^j dt.$$

Démonstration La démonstration découle directement du développement de Taylor avec reste intégral dans le cas d'une fonction d'une variable réelle (p. 21), appliqué à la fonction $\phi : t \in [0, 1] \mapsto f(a + th) \in \mathbb{R}^m$. Il nous suffit de montrer que ϕ est $j + 1$ fois différentiable et que pour tout entier i inférieur ou égal à $j + 1$, $\phi^{(i)}(t) = df^{(i)}(a + th)(\cdot h)^i$.

Cette relation est évidemment satisfaite pour $i = 0$. Supposons qu'elle soit vérifiée au rang $i \leq j$. La fonction f étant $i + 1$ fois différentiable, la fonction $g : x \in U \mapsto df^{(i)}(x)(\cdot h)^i$ est différentiable, et

$$dg(x) \cdot h = df^{(i+1)}(x)(\cdot h)^{i+1}.$$

Par dérivation en chaîne, la fonction $t \mapsto df^{(i)}(a + th)(\cdot h)^i$ est donc dérivable, de dérivée $dg(a + th) \cdot h$, soit $df^{(i+1)}(a + th)(\cdot h)^{i+1}$. ■

Annexe

Définition – Variation d'ordre 1 et 2

Soient U un ouvert de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}^m$ et $x \in U$. Quand cette expression est définie, on appelle *variation d'ordre 1* de f en x associée à la variation h_1 de l'argument la grandeur

$$\Delta f(x, h_1) := f(x + h_1) - f(x)$$

et *variation d'ordre 2* de f en x , associée aux variations h_1 et h_2 de l'argument, la grandeur

$$\begin{aligned} \Delta^2 f(x, h_1, h_2) &:= \Delta(x \mapsto \Delta f(x, h_1))(x, h_2) \\ &= \Delta f(x + h_2, h_1) - \Delta f(x, h_1). \end{aligned}$$

Lemme – Variation d'ordre 2 et matrice hessienne

Soient U un ouvert de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}^m$ et $x \in U$. Si f est deux fois différentiable en x , pour tout $\varepsilon > 0$, il existe un $r > 0$ tel que si $\|h_1\| \leq r$ et $\|h_2\| \leq r$, alors

$$\|\Delta^2 f(x, h_1, h_2) - h_1^\top \cdot H_f(x) \cdot h_2\| \leq \varepsilon (\|h_1\| + \|h_2\|)^2.$$

Démonstration Considérons des vecteurs h_1 et h_2 tels que $x + h_1$, $x + h_2$ et $x + h_1 + h_2$ soient dans le domaine de définition de f . La différence e entre $\Delta^2 f(x, h_1, h_2)$ et $d^2 f(x) \cdot h_1 \cdot h_2$ vaut

$$\begin{aligned} e &= (f(x + h_1 + h_2) - f(x + h_2)) - (f(x + h_1) - f(x)) - d^2 f(x) \cdot h_1 \cdot h_2 \\ &= (f(x + h_1 + h_2) - f(x + h_1) - h_1^\top \cdot H_f(x) \cdot h_2 \\ &\quad - (f(x + h_2) - f(x) - 0^\top \cdot H_f(x) \cdot h_2) \end{aligned}$$

Par conséquent, si l'on définit g par

$$g(u) = f(x + u + h_2) - f(x + u) - u^\top \cdot H_f(x) \cdot h_2,$$

la différence vaut $e = g(h_1) - g(0)$. Cette différence peut être majorée par l'inégalité des accroissements finis : g est différentiable sur le segment $[0, h_1]$ et

$$\nabla g(u) = \nabla f(x + u + h_2) - df(x + u) - H_f(x) \cdot h_2.$$

Comme

$$\begin{aligned} \nabla g(u) &= (\nabla f(x + u + h_2) - \nabla f(x) - H_f(x) \cdot (u + h_2)) \\ &\quad - (\nabla f(x + u) - \nabla f(x) - H_f(x) \cdot u), \end{aligned}$$

par le développement limité du gradient de f (p. 9), pour $\varepsilon > 0$ quelconque, comme $\|u + h_2\| \leq \|h_1\| + \|h_2\|$ et $\|u\| \leq \|h_1\|$, on peut trouver un $r > 0$ tel que si $\|h_1\| < r$ et $\|h_2\| < r$, alors

$$\|\nabla g(u)\| \leq \frac{\varepsilon}{2}(\|h_1\| + \|h_2\|) + \frac{\varepsilon}{2}\|h_1\|.$$

Par conséquent, l'inégalité des accroissements finis fournit

$$\begin{aligned} \|e\| = \|\nabla g(u) - \nabla g(0)\| &\leq \left(\frac{\varepsilon}{2}(\|h_1\| + \|h_2\|) + \frac{\varepsilon}{2}\|h_1\| \right) \|h_1\| \\ &\leq \varepsilon(\|h_1\| + \|h_2\|)^2. \end{aligned}$$

■

Exercices complémentaires

Convexité

Soit U un ensemble ouvert et convexe de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$ une fonction deux fois différentiable.

Question 0 Calculer le développement limité à l'ordre 2 de $f(x + 2h) - 2f(x + h) + f(x)$. (Solution p. 28.)

Question 1 Montrer que si f est convexe, c'est-à-dire si pour tous $x, y \in U$ et $\lambda \in [0, 1]$,

$$f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y),$$

alors pour tout $x \in U$ et $h \in \mathbb{R}^n$,

$$d^2 f(x)(\cdot h)^2 = h^\top \cdot H_f(x) \cdot h \geq 0.$$

(Solution p. 28.)

Question 2 Montrer la réciproque de ce résultat. (Solution p. 28.)

Différentiation en chaîne à l'ordre 2

Soit U et V des ouverts de \mathbb{R}^n et de \mathbb{R}^m , $f : U \rightarrow \mathbb{R}^m$ et $g : V \rightarrow \mathbb{R}$ deux applications deux fois différentiables telles que $f(U) \subset V$.

Question 1 (••) Montrer que $g \circ f$ est deux fois différentiable sur U . (Solution p. 29.)

Question 2 (•••) Montrer que pour tout $x \in U$,

$$H_{g \circ f}(x) = J_f(x)^\top \cdot H_g(f(x)) \cdot J_f(x) + \sum_{k=1}^m \partial_k g(f(x)) H_{f_k}(x).$$

(Solution p. 30.)

Différentiation matricielle

Source: (Tao 2013)

Dans cet exercice :

1. Une fonction $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^{m \times p}$ à valeurs matricielles est différentiable si chacune de ses composantes $F_{ij} : U \rightarrow \mathbb{R}$ est différentiable. La différentielle de F est alors définie par $[dF]_{ij} = dF_{ij}$.
2. Une fonction $f : U \subset \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^p$ dont l'argument X est matriciel est différentiable si la fonction $g : \pi(U) \subset \mathbb{R}^{mn} \rightarrow \mathbb{R}^p$ caractérisée par

$$g(x) := f \left(\begin{bmatrix} X_{11} & \cdots & X_{1n} \\ \vdots & \vdots & \vdots \\ X_{m1} & \cdots & X_{mn} \end{bmatrix} \right)$$

avec

$$x = \pi(X) := (X_{11}, \dots, X_{1n}, \dots, X_{m1}, \dots, X_{mn})$$

est différentiable. On définit alors pour tout $H \in \mathbb{R}^{m \times n}$

$$df(X) \cdot H = dg(x) \cdot h \text{ avec } x = \pi(X), h = \pi(H).$$

La construction se généralise sans difficulté aux fonctions dépendant de plusieurs matrices.

3. Il est possible de combiner les deux cas précédents pour définir la différentielle de fonctions d'argument et de valeur matriciels.

Question 1 Montrer que l'application $\det : A \in \mathbb{R}^{n \times n} \rightarrow \det A \in \mathbb{R}$ est différentiable en l'identité ($A = I$) et calculer cette différentielle. (Solution p. 30.)

Question 2 L'identité de Weinstein–Aronszajn $\det(I + AB) = \det(I + BA)$ vaut pour toutes les matrices carrées A et B de même dimension. En déduire une identité concernant $\text{tr } AB$ et $\text{tr } BA$. (Solution p. 31.)

Question 3 Montrer que l'application $A \mapsto A^{-1}$ est définie dans un voisinage ouvert de l'identité, est différentiable en ce point et calculer cette différentielle. (Solution p. 31.)

Solutions

Exercices essentiels

Laplacien et matrice hessienne Le laplacien de f en x est la somme des coefficients diagonaux – donc la trace – de la matrice hessienne de f en x :

$$\Delta f(x) = \text{tr } H_f(x).$$

Matrice hessienne d'un monôme Le gradient de f est défini en tout point de \mathbb{R}^2 et vaut

$$\nabla f(x_1, x_2) = \begin{bmatrix} \partial_1(x_1 x_2^2) \\ \partial_2(x_1 x_2^2) \end{bmatrix} = \begin{bmatrix} x_2^2 \\ 2x_1 x_2 \end{bmatrix}.$$

Toutes les dérivées partielles des composantes de ∇f sont définies ; on a donc

$$H_f(x) = J_{\nabla f}(x_1, x_2) = \begin{bmatrix} \partial_1(x_2^2) & \partial_2(x_2^2) \\ \partial_1(2x_1 x_2) & \partial_2(2x_1 x_2) \end{bmatrix} = \begin{bmatrix} 0 & 2x_2 \\ 2x_2 & x_1 x_2 \end{bmatrix}.$$

Matrice hessienne d'un lagrangien Le gradient de L en (x, λ) vaut

$$\nabla L(x, \lambda) = \begin{bmatrix} \nabla_x(f(x) + \lambda(g(x) - c)) \\ \partial_\lambda(f(x) + \lambda(g(x) - c)) \end{bmatrix} = \begin{bmatrix} \nabla f(x) + \lambda \nabla g(x) \\ g(x) - c \end{bmatrix},$$

par conséquent

$$H_L(x, \lambda) = J_{\nabla L}(x, \lambda) = \begin{bmatrix} H_f(x) + \lambda H_g(x) & \nabla g(x) \\ \nabla g(x)^\top & 0 \end{bmatrix}.$$

Matrice hessienne diagonale Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction dont la matrice hessienne est partout définie. Si $f(x_1, x_2) = g(x_1) + h(x_2)$ où $g : \mathbb{R} \rightarrow \mathbb{R}$ et $h : \mathbb{R} \rightarrow \mathbb{R}$ sont des fonctions deux fois dérivables, alors f est différentiable et

$$\nabla f(x_1, x_2) = \begin{bmatrix} g'(x_1) \\ h'(x_2) \end{bmatrix}.$$

Ce gradient admet bien une matrice jacobienne et

$$H_f(x_1, x_2) = J_{\nabla f}(x_1, x_2) = \begin{bmatrix} g''(x_1) & 0 \\ 0 & h''(x_2) \end{bmatrix}.$$

Cette matrice est bien diagonale pour tout $(x_1, x_2) \in \mathbb{R}^2$. Réciproquement, si H_f est diagonale, alors $\partial_{12}^2 f = \partial_1 \partial_2 f = 0$. Par le théorème fondamental du calcul on a pour tout $(x_1, x_2) \in \mathbb{R}^2$,

$$\partial_2 f(x_1, x_2) = \partial_2 f(0, x_2) + \int_0^{x_1} \partial_{12}^2 f(y_1, x_2) dy_1 = \partial_2 f(0, x_2).$$

La fonction $\partial_2 f(0, x_2)$ étant dérivable par rapport à x_2 , elle est continue par rapport à x_2 et à nouveau par le théorème fondamental du calcul, on obtient

$$f(x_1, x_2) = f(x_1, 0) + \int_0^{x_2} \partial_2 f(x_1, y_2) dy_2 = f(x_1, 0) + \int_0^{x_2} \partial_2 f(0, y_2) dy_2.$$

Cette fonction est de la forme $f(x_1, x_2) = g(x_1) + h(x_2)$ avec

$$g(x_1) = f(x_1, 0) \text{ et } h(x_2) = \int_0^{x_2} \partial_2 f(0, y_2) dy_2.$$

Les fonctions f et g sont bien deux fois dérivables.

Analyse vectorielle Soient $f : U \rightarrow \mathbb{R}^3$ et $g : U \rightarrow \mathbb{R}$ des fonctions deux fois différentiable en $x \in U$. On a

$$\begin{aligned} \operatorname{div}(\operatorname{rot} f)(x) &= \operatorname{div} \begin{bmatrix} \partial_2 f_3(x) - \partial_3 f_2(x) \\ \partial_3 f_1(x) - \partial_1 f_3(x) \\ \partial_1 f_2(x) - \partial_2 f_1(x) \end{bmatrix} \\ &= \partial_1(\partial_2 f_3 - \partial_3 f_2)(x) + \partial_2(\partial_3 f_1 - \partial_1 f_3)(x) + \partial_3(\partial_1 f_2 - \partial_2 f_1)(x) \\ &= (\partial_{12}^2 f_3 - \partial_{21}^2 f_3)(x) + (\partial_{23}^2 f_1 - \partial_{32}^2 f_1)(x) + (\partial_{31}^2 f_2 - \partial_{13}^2 f_2)(x), \end{aligned}$$

et donc $\operatorname{div}(\operatorname{rot} f)(x) = 0$ par symétrie de la différentielle d'ordre 2 (p. 9). On a également

$$\begin{aligned}\operatorname{rot} \nabla g(x) &:= \begin{bmatrix} \partial_2(\nabla g)_3(x) - \partial_3(\nabla g)_2(x) \\ \partial_3(\nabla g)_1(x) - \partial_1(\nabla g)_3(x) \\ \partial_1(\nabla g)_2(x) - \partial_2(\nabla g)_1(x) \end{bmatrix} \\ &= \begin{bmatrix} \partial_2\partial_3g(x) - \partial_3\partial_2g(x) \\ \partial_3\partial_1g(x) - \partial_1\partial_3g(x) \\ \partial_1\partial_2g(x) - \partial_2\partial_1g(x) \end{bmatrix}\end{aligned}$$

et donc – à nouveau par symétrie de la différentielle d'ordre 2 (p. 9) – on obtient $\operatorname{rot} \nabla g(x) = 0$

Gradient unitaire Si $\|\nabla f\| = 1$, alors

$$\|\nabla f\|^2 = \langle \nabla f, \nabla f \rangle = \sum_i (\partial_i f)^2 = 1$$

et donc pour tout $j \in \{1, \dots, n\}$,

$$\partial_j \left(\sum_i \partial_i f^2 \right) = 2 \sum_i \partial_{ji}^2 f \times \partial_i f = 2H_f^\top \cdot \nabla f = 0.$$

Le résultat $H_f \cdot \nabla f = 0$ s'en déduit donc par symétrie de la différentielle d'ordre 2 (p. 9).

Accroissements finis d'ordre 2 Le développement de Taylor avec reste intégral (p. 11) et l'inégalité triangulaire nous fournissent

$$\begin{aligned}\|f(x+h) - f(x) - \langle \nabla f(x), h \rangle\| &\leq \left\| \int_0^1 (h^\top \cdot H_f(x+th) \cdot h) \times (1-t) dt \right\| \\ &\leq \int_0^1 \|h^\top \cdot H_f(x+th) \cdot h\| (1-t) dt.\end{aligned}$$

L'inégalité de Cauchy-Schwarz et la définition de la norme d'opérateur donnent :

$$\begin{aligned}\|h^\top \cdot H_f(x+th) \cdot h\| &\leq \|h\| \times \|H_f(x+th) \cdot h\| \\ &\leq \|h\| \times \|H_f(x+th)\| \times \|h\| \\ &\leq M\|h\|^2.\end{aligned}$$

Le résultat cherché se déduit alors comme suit :

$$\begin{aligned} \int_0^1 \|h^\top \cdot H_f(x+th) \cdot h\| (1-t) dt &\leq \int_0^1 M \|h\|^2 (1-t) dt \\ &= M \|h\|^2 \int_0^1 (1-t) dt \\ &= M \frac{\|h\|^2}{2}. \end{aligned}$$

Convexité

Question 0 Le développement limité à l'ordre 2 de f en x (p. 20) fournit

$$f(x+h) = f(x) + df(x) \cdot h + \frac{d^2 f(x)}{2} (\cdot h)^2 + o(\|h\|^2)$$

et donc

$$f(x+2h) = f(x) + 2df(x) \cdot h + 4 \frac{d^2 f(x)}{2} (\cdot h)^2 + o(\|h\|^2).$$

Par conséquent,

$$f(x+2h) - 2f(x+h) + f(x) = d^2 f(x) (\cdot h)^2 + o(\|h\|^2).$$

Question 1 En considérant $y = x + 2h$ et $\lambda = 1/2$, on voit que l'hypothèse de convexité de f entraîne

$$f(x+h) \leq \frac{1}{2} f(x) + \frac{1}{2} f(x+2h),$$

soit

$$f(x+2h) - 2f(x+h) + f(x) \geq 0.$$

En utilisant le résultat de la question précédente, on obtient

$$d^2 f(x) (\cdot h)^2 + o(\|h\|^2) \geq 0$$

et donc, en substituant th à h et en faisant tendre t vers 0, $d^2 f(x) (\cdot h)^2 \geq 0$.

Question 2 Comme $f((1-\lambda)x + \lambda y) = f(x + \lambda(y-x))$, l'inégalité de Taylor avec reste intégral fournit

$$\begin{aligned} f((1-\lambda)x + \lambda y) &= f(x) + df(x) \cdot \lambda(y-x) \\ &\quad + \int_0^1 d^2 f(x + t\lambda(y-x)) (\cdot \lambda(y-x))^2 (1-t) dt. \end{aligned}$$

L'intégrale ci-dessus étant égale à

$$\lambda \int_0^1 d^2 f(x + t\lambda(y-x))(\cdot(y-x))^2 \left(1 - \frac{\lambda t}{\lambda}\right) \lambda dt,$$

par le changement de variable $t\lambda \rightarrow t$ elle est égale à

$$\lambda \int_0^\lambda d^2 f(x + t(y-x))(\cdot(y-x))^2 \left(1 - \frac{t}{\lambda}\right) dt.$$

En utilisant le développement de Taylor avec reste intégral pour $\lambda \in]0, 1]$ et $\lambda = 1$, on obtient donc

$$\begin{aligned} f((1-\lambda)x + \lambda y) - \lambda f(y) &= f(x) - \lambda f(x) + df(x) \cdot \lambda(y-x) - \lambda df(x) \cdot (y-x) \\ &\quad + \lambda \int_0^\lambda d^2 f(x + t(y-x))(\cdot(y-x))^2 \left(1 - \frac{t}{\lambda}\right) dt \\ &\quad - \lambda \int_0^1 d^2 f(x + t(y-x))(\cdot(y-x))^2 (1-t) dt, \end{aligned}$$

soit

$$f((1-\lambda)x + \lambda y) - \lambda f(y) - (1-\lambda)f(x) = \lambda \int_0^1 \phi_f(t) \psi_\lambda(t) dt$$

où $\phi_f(t) := d^2 f(x + t(y-x))(\cdot(y-x))^2$ est positive par hypothèse et

$$\psi_\lambda(t) := \begin{cases} t(1-1/\lambda) & \text{si } t \leq \lambda \\ (t-1) & \text{sinon.} \end{cases}$$

La fonction ψ_λ étant négative, on en conclut que $f((1-\lambda)x + \lambda y) - \lambda f(y) - f(x)$ est négative pour tout $\lambda \in]0, 1]$; cette inégalité est également trivialement satisfaite si $\lambda = 0$. La fonction f est donc convexe.

Différentiation en chaîne à l'ordre 2

Question 1 Nous savons par la règle de différentiation en chaîne que $g \circ f$ est différentiable et vérifie $d(g \circ f)(x) = dg(f(x)) \cdot df(x)$, ou encore

$$\nabla(g \circ f)(x) = \nabla f(x) \cdot [J_g(f(x))]^\top.$$

Les coefficients de J_g sont différentiables ainsi que les composants de f , par conséquent tous les coefficients de $J_g \circ f$ sont différentiables par la règle de différentiation en chaîne. Les composants de ∇f sont également différentiables; les composants de $\nabla(g \circ f)$ se déduisant de tous ces composants par des opérations différentiables – des produits et des sommes – ils sont tous différentiables. La fonction $\nabla(g \circ f)$ est donc différentiable et $g \circ f$ est deux fois différentiable.

Question 2 La règle de différentiation en chaîne donne pour tout indice $i \in \{1, \dots, n\}$

$$\partial_i(g \circ f)(x) = dg(f(x)) \cdot \partial_i f(x) = \sum_{k=1}^m \partial_k g(f(x)) \partial_i f_k(x).$$

Pour tout $j \in \{1, \dots, m\}$, on a donc

$$\begin{aligned} \partial_{ji}^2(g \circ f) &= \partial_j(\partial_i(g \circ f)) \\ &= \partial_j \left(\sum_{k=1}^m (\partial_k g) \circ f \times \partial_i f_k \right) \\ &= \sum_{k=1}^m \partial_j((\partial_k g) \circ f) \times \partial_i f_k + (\partial_k g) \circ f \times \partial_j(\partial_i f_k) \end{aligned}$$

Comme par la règle de différentiation en chaîne

$$\partial_j((\partial_k g) \circ f) = [d((\partial_k g) \circ f)]_j = [(d(\partial_k g) \circ f) \cdot df]_j = \sum_{\ell=1}^m \partial_\ell(\partial_k g) \circ f \times \partial_j f_\ell,$$

on en déduit que

$$\partial_{ji}^2(g \circ f) = \sum_{k=1}^m \left[\sum_{\ell=1}^m (\partial_{\ell k}^2 g) \circ f \times \partial_j f_\ell \times \partial_i f_k \right] + \sum_{k=1}^m (\partial_k g) \circ f \times \partial_{ji}^2 f_k,$$

soit

$$[H_{g \circ f}]_{ij} = \sum_{k=1}^m \sum_{\ell=1}^m [J_f^\top]_{ik} \times ([H_g]_{k\ell} \circ f) \times [J_f]_{\ell j} + \sum_{k=1}^m (\partial_k g) \circ f \times [H_{f_k}]_{ij},$$

ce qui prouve pour tout $x \in U$ la relation matricielle

$$H_{g \circ f}(x) = J_f(x)^\top \cdot H_g(f(x)) \cdot J_f(x) + \sum_{k=1}^m \partial_k g(f(x)) H_{f_k}(x).$$

TODO – Différentiation matricielle

Question 1 Soit $H \in \mathbb{R}^{n \times n}$, telle que

$$H = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & h_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ h_{n1} & h_{n2} & \dots & h_{nn} \end{bmatrix}.$$

En développant le déterminant selon la première colonne, on constate que

$$\begin{aligned}\det(I + H) &= \begin{vmatrix} 1 + h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & 1 + h_{22} & \dots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \dots & 1 + h_{nn} \end{vmatrix} \\ &= (1 + h_{11}) \begin{vmatrix} 1 + h_{22} & \dots & h_{2n} \\ \vdots & \ddots & \vdots \\ h_{n2} & \dots & 1 + h_{nn} \end{vmatrix} + o(\|H\|),\end{aligned}$$

une relation dont on tire par récurrence que

$$\begin{aligned}\det(I + H) &= \prod_{i=1}^n (1 + h_{ii}) + o(\|H\|) = \det I + \sum_{i=1}^n h_{ii} + o(\|H\|) \\ &= \det I + \operatorname{tr} H + o(\|H\|).\end{aligned}$$

La différentiel du déterminant existe donc en l'identité et $d\det(I) \cdot H = \operatorname{tr} H$.

Question 2 Pour tout réel ε et A, B matrices carrées de même taille, on a

$$\det(I + \varepsilon AB) = \det(I + \varepsilon BA).$$

Les deux membres de cette équation sont dérivables par rapport à ε en 0 par la règle de différentiation en chaîne et l'égalité de ces dérivées fournit

$$\operatorname{tr} AB = \operatorname{tr} BA.$$

Question 3 Le déterminant étant une application continue, si $A \in \mathbb{R}^{n \times n}$ est suffisamment proche de l'identité – dont le déterminant vaut 1 – son déterminant est positif; la matrice A est alors inversible.

Quand la matrice $A \in \mathbb{R}^{n \times n}$ est suffisamment proche de l'identité pour être inversible, la formule de Cramer établit

$$A^{-1} = \frac{1}{\det A} \operatorname{co}(A)^t.$$

Chaque coefficient de $\operatorname{co}(A)^t$ (la transposée de la comatrice de A) est une fonction polynomiale des coefficients a_{ij} de A ; chaque coefficient de $\operatorname{co}(A)^t$ est donc une fonction continûment différentiable des coefficients de A et donc différentiable en $A = I$. Par la règle du produit, chaque coefficient de A^{-1} est donc différentiable en $A = I$; l'application $A \mapsto A^{-1}$ est donc différentiable en $A = I$.

Notons $\operatorname{inv}(A) = A^{-1}$; comme $\operatorname{inv}(I + H) = I + d\operatorname{inv}(I) \cdot H + o(\|H\|)$, l'identité $(I + H)(I + H)^{-1} = I$ fournit :

$$(I + H)(I + d\operatorname{inv}(I) \cdot H + o(\|H\|)) = I + H + d\operatorname{inv}(I) \cdot H + o(\|H\|) = I,$$

et donc

$$d\operatorname{inv}(I) \cdot H = -H.$$

Références

Tao, Terence. 2013. “Matrix Identities as Derivatives of Determinant Identities.”
<https://terrytao.wordpress.com/2013/01/13/matrix-identities-as-derivatives-of-determinant-identities/>.