Calcul Différentiel III

STEP, MINES ParisTech*

8 octobre 2020 (`#bd0f014`)

Table des matières

*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).

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 :

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

Sauf mention particulière, les objectifs "Expert", les démonstrations du document [1] et les contenus en annexe ne sont pas exigibles ("hors-programme").

TODO

- matrice hessienne & "formules" directement liées à la définition: $H_f = J_{\nabla f}$, $\partial_{j_1}\partial_{j_2} = \partial^2_{j_1 j_2}$, $[H_f]_{j_1 j_2} = \partial^2_{j_1 j_2} f$,
- diff d'ordre 2, cont. diff d'ordre 2
- "formules" supposant la diff d'ordre 2: $d^2 f(x) \cdot h_1 \cdot h_2 = h_1^\top \cdot H_f(x) \cdot h_2$, $\nabla f(x+h) = \nabla f(x) + H_f(x) \cdot h + \varepsilon(h)\|h\|$.
- symétrie de la matrice hessienne,
- dvlpt limité d'ordre 2

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

TODO

Evaluer stratégie

- diff d'ordre deux d'une fonction (multivariable) scalaire et tout ce qu'on peut faire à ce niveau, avec de façon concrête la matrice hessienne au centre de tout ça (comme le jacobien l'était dans le chapitre 1).
- puis, dans un second temps seulement, introduction des tenseurs et "déblocage" : de la différentielle d'ordre 2 de fonction vectorielles, puis de la différentielle d'ordre n.
- Tenseur d'ordre (0, 1, 2 et) 3. Structure d'espace vectoriel normé. Contraction tensorielle, lien avec les applis n-linéaires (ouch).
- Différentielle et matrices (surtout à *valeurs* matricielles ; il va s'agir de différencier $f'(x)$. Mais on peut en profiter pour avoir des variables matricielles aussi ... D'autant que si on veut utiliser la chain rule, pour avoir une "chain rule d'ordre 2", on voudrait utilser la chain rule d'ordre 1 à travers le produit matriciel $(A, B) \to A \cdot B$ donc tout ça est lié.
- Tenseur des dérivées d'ordre 3.
- Différentiabilité d'ordre 2, fct 2 fois continument différentiable.
- Th fcts implicite version C^n, C^n difféo ?

Exercices :

- Fcts quadratique, Gaussienne, etc.
- Courbure (dans le plan ?)
- "bordered hessian" (optim.)
- formules d'analyse vectorielle (div de rot, div $f\vec{u}$, etc.)
- exemples calcul de DIFFERENTIAL CALCULUS, TENSOR PRODUCTS AND THE IMPORTANCE OF NOTATION (JONATHAN H. MANTON).

Matrice hessienne et différentielle d'ordre 2

Définition – Dérivées partielles d'ordre 2

Soit U un ouvert de $\mathbb{R}^m$, $f : U \to \mathbb{R}$ 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

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

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

Définition – Matrice hessienne

Soient U un ouvert de $\mathbb{R}^n$, $f : U \to \mathbb{R}^m$ 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

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

c'est-à-dire

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

Exercice – Matrice hessienne d'un monôme (•) Soit $f : (x_1, x_2) \in \mathbb{R}^2 \to \mathbb{R}$ 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. 17.)

Exercice – Matrice hessienne d'un lagrangien (••) Soit U un ouvert de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$ et $g : U \to \mathbb{R}$ deux applications dont les matrices hessiennes sont définies sur U. Soit $c \in \mathbb{R}$ un constante et $L : U \times \mathbb{R} \to \mathbb{R}$ la fonction telle que $L(x, \lambda) = f(x) + \lambda(g(x) - c)$. Calculer $H_L(x, \lambda)$. (Solution p. 17.)

Définition – Continue différentiabilité d'ordre 2

Soit U un ouvert de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$. 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 $\partial^2_{j_2 j_1} f : U \to \mathbb{R}$ 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 $\mathbb{R}^n$ et $f : U \subset \mathbb{R}^n \to \mathbb{R}$. 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 $d^2 f(x)$, est définie [2] comme l'application linéaire telle que pour tout h_1 dans $\mathbb{R}^n$,

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

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

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

2. On peut vérifier que le terme $d(x \mapsto df(x) \cdot h_1)(x)$ dépend bien linéairement de h_1, ce qui justifie l'assertion que $d^2 f(x)$ est linéaire et donc l'usage du "$\cdot$" lorsqu'elle est appliquée à un argument h_1.

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

Remarque – Notations

Par construction, le terme $d(x \mapsto df(x) \cdot h_1)(x)$ est une application linéaire de $\mathbb{R}^n \to \mathbb{R}^m$, donc la fonction $d^2 f(x)$ associe linéairement à un vecteur de $\mathbb{R}^n$ une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}$. Autrement dit, si l'on note $A \to B$ l'ensemble des fonctions de A dans B, on a

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

ce qui se décline successivement en

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

On conviendra que dans ce contexte, le symbole "$\to$" associe à droite :

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

La convention associée – utilisée dans la définition de la différentielle d'ordre 2 – veut que lors de l'application d'une fonction linéaire, le symbole "$\cdot$" associe à gauche :

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

Proposition – Différentielle d'ordre 2 et matrice hessienne

Soit U un ouvert de $\mathbb{R}^n$, $f : U \subset \mathbb{R}^n \to \mathbb{R}$ 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 $h_1, h_2 \in \mathbb{R}^n$,

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

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

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

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

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

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

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 :

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

■

Proposition – Continue différentiabilité et différentiabilité d'ordre 2

Soit U un ouvert de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$. 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. 5). 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 $\mathbb{R}^n$, $f : U \subset \mathbb{R}^n \to \mathbb{R}$ 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 $\lim_{h \to 0} \varepsilon(h) = 0$.

Démonstration D'après la proposition "Différentielle d'ordre 2 et matrice hessienne" (p. 5), ∇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. 3), $H_f(x) = J_{\nabla f}(x)$ d'où l'égalité de l'énoncé. ■

Théorème – Symétrie de la différentielle d'ordre 2

Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}$ 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 $j_1, j_2 \in \{1, \dots, n\}$,

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

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. 14) en notant que

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

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 $d^2 f(x) \cdot h_1 \cdot h_2 - d^2 f(x) \cdot h_2 \cdot h_1 = 0$. ■

Proposition – Développement limité à l'ordre 2

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

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

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

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

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

La fonction $g : h \mapsto f(x+h) - f(x) - \langle \nabla f(x), h \rangle - h^\top \cdot H_f(x) \cdot h \in \mathbb{R}$ 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. 7),

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

Compte tenu du développement limité du gradient de f en x (p. 7), 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

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

■

Différentielle d'ordre supérieur

TODO notation i bof ; prendre m ?

Définition – Tenseur d'ordre n

On appelera *tenseur d'ordre* n un élément de $\mathbb{R}^{i_1 \times i_2 \times \cdots \times i_n}$ où $(i_1, i_2, \ldots, i_n) \in \mathbb{N}^n$, c'est-à-dire une application A de la forme

$$(i_1, i_2, \ldots, i_n) \mapsto A_{i_1 i_2 \ldots i_n} \in \mathbb{R},$$

ou encore, un tableau n-dimensionnel de réels.

Le concept de tenseur généralise la notion de scalaire de $\mathbb{R}$ (un tenseur d'ordre 0), de vecteur de $\mathbb{R}^n$ (un tenseur d'ordre 1) et de matrice $\mathbb{R}^{m \times n}$ (un tenseur d'ordre 2).

TODO.

- Identification tenseur application n-linéaire.
- Contraction entre tenseurs (taille compatible),
- Contraction d'ordre p (quelle convention et notation ?),
- Décomposer produit de tenseurs et contraction d'indice (pour UN tenseur) ou combiner ? Indices nommés ?
- Coller au plus près de NumPy et donner des exemples avec NumPy (et einsum ?). Regarder aussi dot, tensordot, outer, etc. Voir ce qui fait le job ... Ca serait bien de pouvoir se limiter à `dot` ... Regarder les 3 use cases: diff d'ordre n, chain rule d'ordre 2, determinant et/ou diff de fct matricielles (valeurs et/ou args).

La notion de différentielle d'ordre 2 se généralise sans difficulté à un ordre plus élevé, par induction sur l'ordre de la différentielle.

TODO

Expliquer généralisation scalaire -> vectoriel et ordre k.

Définition – Différentielle d'ordre k

Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ 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

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

Remarque

On a

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

Lemme – Stratification

Si $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ est une fonction k fois différentiable en un point x de U, pour tous vecteurs h_1, h_2, ..., 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. 10),

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

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

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

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

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

et donc au final

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

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

Proposition – Symétrie des différentielles d'ordre supérieur

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

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

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. 7). Supposons désormais le résultat établi au rang $k \geq 2$. En utilisant la stratification de $d^{k+1} f(x) \cdot h_1 \cdot \ldots \cdot h_k \cdot h_{k+1}$ pour $p = 1$ et $p = k$ (p. 10), on peut établir le résultat si i et j appartiennent tous les deux à $\{2, \ldots, k+1\}$ ou à $\{1, \ldots, k\}$. 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

Les dérivées partielles d'ordre supérieur se définissent par récurrence, de manière similaire aux dérivées partielles d'ordre 2. Pour simplifier la notation $\partial^k_{i_1 \ldots i_k} f(x)$, on exploite le fait que si f est k fois différentiable en x,

$$\partial^k_{i_1 \ldots i_k} f(x) = d^k f(x) \cdot e_{i_1} \cdot \ldots \cdot e_{i_k}.$$

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

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

Puissance symbolique

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

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

Théorème – Développement limité d'ordre supérieur

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

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

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 un $o(\|h\|^j)$, ce qui nous allons accomplir en établissant que $\|dr(h)\| = o(\|h\|^{j-1})$. En effet, si $dr(h) = E(h)\|h\|^{j-1}$ où l'application linéaire E est un $o(1)$, alors 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 $r(h) = o(\|h\|^j)$.

Etablissons donc que $r(h)$ est un $o(\|h\|^j)$. Les termes $d^i f(x) \cdot h_1 \cdot \ldots \cdot h_i$ sont linéaires par rapport à chacun des h_j, donc pour tout vecteur k, compte tenu de la symétrie de $d^i f(x)$,

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

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

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

Par le lemme de stratification (p. 10) et la symétrie des différentielles d'ordre supérieur (p. 11), on obtient

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

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

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

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

Développement de Taylor avec reste intégral I

Soit $f : [a, a+h] \to \mathbb{R}^m$ 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 (p. **??**). Si l'on suppose la relation vérifiée à l'ordre j, et f $j+2$ fois dérivable, par intégration par parties (p. **??**), on obtient

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

soit

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

ce qui achève la preuve par récurrence. ■

Développement de Taylor avec reste intégral II

Si $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ 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. 13), 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 $\mathbb{R}^n$, $f : U \to \mathbb{R}^m$ 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 $\mathbb{R}^n$, $f : U \to \mathbb{R}^m$ 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

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

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

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

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

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

par le développement limité du gradient de f (p. 7), 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

Convexité

Soit U un ensemble ouvert et convexe de $\mathbb{R}^n$ et $f:U\to\mathbb{R}$ 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. 18.)

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. 18.)

Question 2 Montrer la réciproque de ce résultat. (Solution p. 18.)

Différentiation en chaîne à l'ordre 2

Soit U et V des ouverts de $\mathbb{R}^n$ et de $\mathbb{R}^m$, $f : U \to \mathbb{R}^m$ et $g : V \to \mathbb{R}$ 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. 19.)

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. 19.)

Différentiation matricielle

Source: (Tao 2013)

Dans cet exercice :

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

$$g(x) := f\left(\left[\begin{array}{ccc} X_{11} & \dots & X_{1n} \\ \vdots & \vdots & \vdots \\ X_{m1} & \dots & X_{mn} \end{array}\right]\right)$$

avec

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

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 $\det : A \in \mathbb{R}^{n \times n} \to \det A \in \mathbb{R}$ est différentiable en l'identité ($A = I$) et calculer cette différentielle. (Solution p. 20.)

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 $\operatorname{tr} AB$ et $\operatorname{tr} BA$. (Solution p. 21.)

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. 21.)

Solutions

Exercices essentiels

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

$$\nabla f(x_1, x_2) = \left[\begin{array}{c} \partial_1(x_1 x_2^2) \\ \partial_2(x_1 x_2^2) \end{array} \right] = \left[\begin{array}{c} x_2^2 \\ 2x_1 x_2 \end{array} \right].$$

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

$$H_f(x) = J_{\nabla f}(x_1, x_2) = \left[\begin{array}{cc} \partial_1(x_2^2) & \partial_2(x_2^2) \\ \partial_1(2x_1 x_2) & \partial_2(2x_1 x_2) \end{array} \right] = \left[\begin{array}{cc} 0 & 2x_2 \\ 2x_2 & x_1 x_2 \end{array} \right].$$

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

$$\nabla L(x, \lambda) = \left[\begin{array}{c} \nabla_x (f(x) + \lambda(g(x) - c)) \\ \partial_\lambda (f(x) + \lambda(g(x) - c)) \end{array} \right] = \left[\begin{array}{c} \nabla f(x) + \lambda \nabla g(x) \\ g(x) - c \end{array} \right],$$

par conséquent

$$H_L(x, \lambda) = J_{\nabla L}(x, \lambda) = \left[\begin{array}{cc} H_f(x) + \lambda H_g(x) & \nabla g(x) \\ \nabla g(x)^\top & 0 \end{array} \right].$$

Convexité

Question 0 Le développement limité à l'ordre 2 de f en x (p. 12) fournit

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

et donc

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

Par conséquent,

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

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

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

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

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

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

L'intégrale ci-dessus étant égale à

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

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

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

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

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

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

$$\psi_\lambda(t) := \left| \begin{array}{cc} t(1 - 1/\lambda) & \text{si } t \leq \lambda \\ (t-1) & \text{sinon.} \end{array} \right.$$

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 $i \in \{1, \ldots, n\}$

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

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

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

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

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

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

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

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

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

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 équations 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/`.