


Calcul Différentiel I

STEP, MINES ParisTech*

5 mars 2021 (#72befad)

Table des matières

Notations	4
Préambule	4
Ensembles et fonctions	4
Applications linéaires et calcul matriciel	4
Multiplication scalaire-vecteur	4
Matrices	5
Mise à plat des matrices	5
Applications linéaires	6
Composition d'applications linéaires	7
Adjoint d'un opérateur	7
Vecteurs colonnes et vecteur lignes	8
Notation de Landau	9
Petit o de Landau	9
Différentielle	10
Dérivée	10
Valeurs scalaires ou vectorielles	10
Développement limité au premier ordre	10
Fonctions linéaires d'une variable scalaire	11
Différentielle de Fréchet	11
Variation d'une fonction	12
Variation et différentielle	12
Différentielle d'une fonction affine	12
Différentielle et dérivée	12
Gradient	13
Point critique	13

*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.

Point critique et extrema	13
Différentiation composante par composante	14
Valeurs et variables matricielles	14
Domaine de définition non ouvert	15
Différencier une expression	16
Règle de différentiation en chaîne	17
Notations	17
Règle de la somme	18
Règle du produit	18
Linéarité de la différentielle	19
Jacobienne, dérivées partielles et directionnelles	19
Matrice jacobienne	19
Dérivées partielles	19
Différentielles partielles	20
Arguments nommés	20
Différentielle et dérivées partielles	21
Différentielle et différentielles partielles	21
Matrice jacobienne et dérivées partielles	22
Gradient et dérivées partielles	22
Matrice jacobienne, gradient et dérivées partielles	23
Fonction continûment différentiable	23
Dérivées partielles continues	23
Matrice jacobienne continue	24
Variation des fonctions	24
Différentielle et intégrale	24
Théorème fondamental du calcul	25
Variation d'une fonction	25
Démonstration	25
Inégalité des accroissements finis I	26
Inégalité des accroissements finis II	27
Différentielles d'ordre supérieur	27
Différentielle d'ordre 2	27
Remarques	27
Variation de la différentielle I	28
Interprétation du $o(\ k\)$	28
Variation d'ordre 2	29
Variation et différentielle d'ordre deux	30
Symétrie de la différentielle d'ordre 2	30
Variation de la différentielle II	31
Dérivées partielles d'ordre 2	32
Symétrie des dérivées partielles d'ordre 2	32
Hessienne	32
Différentielle d'ordre k	33

Remarque	33
Stratification	33
Symétrie des différentielles d'ordre supérieur	34
Dérivées partielles d'ordre supérieur et multi-indices	34
Puissance symbolique	34
Développement limité d'ordre supérieur	35
Développement de Taylor avec reste intégral I	36
Développement de Taylor avec reste intégral II	36
Annexe – Intégrale de Newton	37
Intégrale de Newton	37
Linéarité	38
Majoration	38
Intégration par parties	39
Exercices	39
Dérivée sur un intervalle fermé	39
Différentiation en chaîne	39
Fonction quadratique	40
Vecteur gaussien	40
Robot manipulateur	40
Différentiation matricielle	41
Dérivée partielles, directionnelles et différentielle	41
Dérivée directionnelle d'Hadamard	41
Inégalité de la valeur moyenne	42
Egalité des accroissements finis ?	43
Convexité	43
Solutions	43
Dérivée sur un intervalle fermé	43
Différentiation en chaîne	44
Fonction quadratique	44
Vecteur gaussien	45
Robot manipulateur	47
Différentiation matricielle	47
Dérivée partielles, directionnelles et différentielle	48
Dérivée directionnelle d'Hadamard	49
Inégalité de la valeur moyenne	53
Egalité des accroissements finis ?	53
Convexité	53
Références	55

Notations

Préambule

Les fragments de code de ce document utilisent le langage Python 3. La bibliothèque NumPy est exploitée:

```
>>> from numpy import *
```

Ensembles et fonctions

La notation classique $f : A \rightarrow B$ pour désigner une fonction f d'un ensemble A dans un ensemble B suggère d'utiliser $A \rightarrow B$ pour désigner l'ensemble des fonctions de A dans B . Avec cette convention, $f : A \rightarrow B$ signifie la même chose que $f \in A \rightarrow B$.

La convention que nous adoptons a vocation à simplifier la manipulation de fonctions dont les valeurs sont des fonctions, un schéma très fréquent en calcul différentiel. Si $f : A \rightarrow B$ et $g : B \rightarrow C$, la composée des fonctions f et g , notée $g \circ f$, appartient à $A \rightarrow C$ et est définie par

$$(g \circ f)(x) = g(f(x)).$$

Si l'on applique bien f à x , puis g au résultat, il est néanmoins naturel d'inverser l'ordre d'apparition des fonctions dans la notation $g \circ f$; il faut en effet s'adapter à la notation classique (infixe ou polonaise) qui désigne par $f(x)$ l'image de x par f . Pour cette même raison, il pourra être utile de d'utiliser $B \leftarrow A$ comme une variante de la notation $A \rightarrow B$. On pourra alors utiliser la règle

$$g : C \leftarrow B, f : B \leftarrow A \implies g \circ f : C \leftarrow A$$

où les notations des ensembles et fonctions g , f , A , B et C restent dans le même ordre d'apparition et les deux occurrences de l'ensemble intermédiaire B se touchent.

Applications linéaires et calcul matriciel

Multiplication scalaire-vecteur

Pour tout scalaire $\lambda \in \mathbb{R}$ et vecteur $x \in \mathbb{R}^n$, on notera λx ou parfois $x\lambda$ la multiplication du vecteur x par le scalaire λ . Lorsque λ est non nul, on notera également x/λ le vecteur $(1/\lambda)x$.

Un vecteur de \mathbb{R}^n est représenté dans NumPy par un tableau à une dimension :

```
>>> x = array([1, 2, 3])
>>> x.ndim
1
>>> shape(x)
(3,)
>>> size(x)
3
```

La multiplication d'un scalaire et d'un vecteur est désignée par le symbole `*`:

```
>>> 2 * x
array([2, 4, 6])
```

Matrices

Nous noterons $\mathbb{R}^{m \times n}$ l'ensemble des matrices à m lignes et n colonnes à coefficients réels. Une matrice telle que

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in \mathbb{R}^{2 \times 3}$$

sera représentée avec NumPy par un tableau bi-dimensionnel:

```
>>> A = array([[1, 2, 3], [4, 5, 6]])
>>> A
array([[1, 2, 3],
       [4, 5, 6]])
>>> A.ndim
2
>>> shape(A)
(2, 3)
>>> size(A)
6
```

Mise à plat des matrices

Dans la notation $\mathbb{R}^{m \times n}$, \times est un symbole de séparation, purement syntactique : $\mathbb{R}^{2 \times 3}$ désigne ainsi l'ensemble des matrices à 2 lignes et 3 colonnes à coefficients réels et diffère de \mathbb{R}^6 qui désigne l'ensemble des 6-uplets à coefficients réels.

Ces deux ensembles sont toutefois similaires: pour toute matrice $A \in \mathbb{R}^{m \times n}$, on peut construire un mn -uplet en listant tous les coefficients de la matrices en parcourant l'ensemble des lignes de la matrice de haut en bas et chaque ligne de gauche à droite; cette façon de faire définit un vecteur de \mathbb{R}^{mn} . L'opération ainsi définie sera notée $\pi_{m \times n}$ (voire π s'il n'y a pas d'ambiguïté). Par exemple :

$$\pi_{2 \times 3} : \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in \mathbb{R}^{2 \times 3} \mapsto (1, 2, 3, 4, 5, 6) \in \mathbb{R}^6$$

Cette opération est bijective; elle-même ainsi que son inverse sont linéaires. $\mathbb{R}^{m \times n}$ et \mathbb{R}^{mn} sont donc isomorphes (en tant qu'espace vectoriels), ce que l'on notera :

$$\mathbb{R}^{m \times n} \cong \mathbb{R}^{mn}$$

Le passage de la forme matrice à la forme vecteur se fait de la façon suivante avec NumPy:

```
>>> A
array([[1, 2, 3],
       [4, 5, 6]])
>>> a = reshape(A, (6,))
>>> a
array([1, 2, 3, 4, 5, 6])
>>> reshape(a, (2, 3))
array([[1, 2, 3],
       [4, 5, 6]])
```

Applications linéaires

Notons

$$\mathbb{R}^n \xrightarrow{\ell} \mathbb{R}^m \text{ ou } \mathbb{R}^m \xleftarrow{\ell} \mathbb{R}^n$$

l'ensemble des applications linéaires de \mathbb{R}^n dans \mathbb{R}^m . La raison d'être des matrices $\mathbb{R}^{m \times n}$ est de représenter ces applications linéaires.

Si A désigne une application linéaire de \mathbb{R}^n dans \mathbb{R}^m , on peut la décomposer en m composantes A_i , des applications de \mathbb{R}^n dans \mathbb{R} telles que pour tout x dans \mathbb{R}^n , on ait $A(x) = (A_1(x), A_2(x), \dots, A_m(x))$, ce que l'on note simplement

$$A = (A_1, A_2, \dots, A_m).$$

Si l'on désigne maintenant par e_j le j -ème vecteur de la base canonique de \mathbb{R}^n

$$e_1 = (1, 0, 0, \dots, 0), e_2 = (0, 1, 0, \dots, 0), \dots, e_n = (0, 0, 0, \dots, 1),$$

il est possible d'associer à l'application linéaire $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ la matrice

$$[A] := [A_i(e_j)]_{ij} = \begin{bmatrix} A_1(e_1) & A_1(e_2) & \cdots & A_1(e_n) \\ \vdots & \vdots & \vdots & \vdots \\ A_m(e_1) & A_m(e_2) & \cdots & A_m(e_n) \end{bmatrix} \in \mathbb{R}^{m \times n}.$$

Réciproquement, étant donné une matrice

$$[a_{ij}]_{ij} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{R}^{m \times n},$$

il est possible de définir une application linéaire $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ par la relation

$$(Ax)_i := \sum_j a_{ij} x_j$$

et cette opération est l'inverse de la précédente.

Cette correspondance établit un isomorphisme d'espaces vectoriels entre les applications linéaires de \mathbb{R}^n dans \mathbb{R}^m et les matrices de taille $m \times n$ à coefficients réels :

$$\mathbb{R}^m \xleftarrow{\ell} \mathbb{R}^n \cong \mathbb{R}^{m \times n}$$

Composition d'applications linéaires

Si A et B désignent des applications linéaires de \mathbb{R}^p dans \mathbb{R}^n et de \mathbb{R}^n dans \mathbb{R}^m respectivement, la fonction composée $C = B \circ A$ est une application linéaire qui vérifie

$$C_{ij} = \sum_k B_{ik} A_{kj}.$$

Autrement dit, la composition de fonction linéaires se traduit par la multiplication des matrices associées.

Dans la suite on évitera en général l'utilisation du symbole \circ pour désigner la composition d'applications linéaires, en lui préférant le symbole \cdot ("point").

La méthode `dot` des tableaux NumPy permet de calculer ce produit matriciel :

```
>>> A = array([[1, 2, 3], [4, 5, 6]])
>>> B = array([[1, 0, 0], [0, 1, 0], [0, 0, 1]])
>>> A.dot(B)
array([[1, 2, 3],
       [4, 5, 6]])
```

Définition – Adjoint d'un opérateur

Lorsque $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ est un opérateur linéaire, on peut définir de façon unique l'opérateur *adjoint* $A^* : \mathbb{R}^m \rightarrow \mathbb{R}^n$ comme l'unique opérateur tel que pour tout $x \in \mathbb{R}^n$ et tout $y \in \mathbb{R}^m$, on ait

$$\langle y, A \cdot x \rangle = \langle A^* \cdot y, x \rangle.$$

La matrice représentant A^* est la transposée de la matrice représentant A :

```
>>> A
array([[1, 2, 3],
       [4, 5, 6]])
>>> transpose(A)
array([[1, 4],
       [2, 5],
       [3, 6]])
```

Vecteurs colonnes et vecteur lignes

Dans le cadre du calcul matriciel, on associe souvent à un vecteur $x = (x_1, \dots, x_n)$ de \mathbb{R}^n le vecteur colonne

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{n \times 1}.$$

Dans cette terminologie, un vecteur colonne n'est pas, malgré son nom, un vecteur de \mathbb{R}^n , mais bien une matrice de taille $n \times 1$. Formellement, on a associé à x une matrice $X \in \mathbb{R}^{n \times 1}$, telle que $X_{i1} = x_i$. Le produit entre une matrice et un vecteur colonne de taille compatible n'est rien d'autre qu'un produit matriciel classique.

Le vecteur x étant associé à une matrice, on peut se demander quel opérateur linéaire est associé à cette matrice. La réponse est simple: il s'agit de l'application

$$\lambda \in \mathbb{R} \mapsto \lambda x \in \mathbb{R}^n.$$

Identifier un vecteur et son opérateur linéaire de $\mathbb{R} \rightarrow \mathbb{R}^n$ permet par exemple de disposer “gratuitement” de la définition x^* (l'adjoint de l'opérateur associé à x) : il s'agit de l'application linéaire de \mathbb{R}^n dans \mathbb{R} dont la matrice est la transposée du vecteur colonne associé à x , autrement dit, la représentation de x comme vecteur ligne.

L'intérêt de la représentation des vecteurs comme vecteurs colonnes : si A est une application linéaire de \mathbb{R}^n dans \mathbb{R}^m et x un vecteur de \mathbb{R}^n , le vecteur image $y = A \cdot x \in \mathbb{R}^m$ de x par A est représenté par le vecteur colonne qui est le produit entre la représentation de A comme matrice et la représentation de x comme vecteur colonne.

Concrètement, NumPy ne nécessite pas qu'un vecteur soit d'abord transformé en matrice pour réaliser un produit matrice-vecteur. La méthode `dot` des tableaux peut être utilisée ici aussi pour réaliser cette opération:

```
>>> A = array([[1, 2, 3], [4, 5, 6]])
>>> x = array([7, 8, 9])
>>> A.dot(x)
array([ 50, 122])
```


Le produit matriciel étant associatif, tant que l'on manipule des matrices et des vecteurs, il n'y a pas lieu de préciser si $A \cdot B \cdot C$ désigne $(A \cdot B) \cdot C$ (association à gauche) ou $A \cdot (B \cdot C)$ (association à droite). Comme le produit matrice-vecteur est un produit matriciel classique, quand x est un vecteur, $A \cdot B \cdot x$ désigne indifféremment $(A \cdot B) \cdot x$ ou $A \cdot (B \cdot x)$.

Notation de Landau

Petit o de Landau

La notation $o(\|h\|^k)$, où $h \in \mathbb{R}^n$ et $k \in \mathbb{N}$, désigne une expression de la forme

$$o(\|h\|^k) := \varepsilon(h)\|h\|^k$$

où ε est une fonction définie dans un voisinage de 0 et telle que

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

En particulier, dans le cas $k = 0$, la notation $o(1)$ désigne un terme de la forme $\varepsilon(h)$ où ε est une fonction du type défini ci-dessus. Le cas général peut toujours être réduit à ce cas particulier puisque l'on a $o(\|h\|^k) = o(1)\|h\|^k$.

En dehors de tout contexte, cette notation est très ambiguë puisque l'on ne précise même pas à quel ensemble appartiennent les valeurs de ε . Les choses se précisent lorsqu'elle est utilisée dans une équation donnée, comme

$$\phi(h) = o(\|h\|^k)$$

où la fonction ϕ est connue. Cette relation signifie alors: la fonction ϕ est définie dans un voisinage de 0 et vérifie:

$$\lim_{h \rightarrow 0} \frac{\phi(h)}{\|h\|^k} = 0.$$

La fonction ε est alors définie de façon unique sur ce voisinage de 0 par la relation

$$\varepsilon(h) = \frac{\phi(h)}{\|h\|^k} \text{ si } h \neq 0 \text{ et } \varepsilon(0) = 0.$$

Exemple – Continuité Si f est une fonction définie d'un sous-ensemble de \mathbb{R}^n et que $x \in \mathbb{R}^n$, la notation

$$f(x+h) = f(x) + o(1)$$

signifie donc que f définie dans un voisinage de x et que

$$\lim_{h \rightarrow 0} f(x+h) = f(x),$$

autrement dit que x appartient à l'intérieur du domaine de définition de f et que f est continue en ce point.

Différentielle

Dérivée

Soient U un ouvert de \mathbb{R} et $f : U \rightarrow \mathbb{R}^m$. La fonction f est *dérivable* en $x \in U$ s'il existe une limite $\ell \in \mathbb{R}^n$ au *taux d'accroissement* de f en x :

$$\ell = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

Cette limite quand elle existe est unique; elle est appelée *dérivée de f en x* et notée $f'(x)$:

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}.$$

Remarque – Valeurs scalaires ou vectorielles

La formation d'un taux d'accroissement dans cette définition nécessite que l'on divise par h et par conséquent que h soit scalaire; la dérivée ne peut donc être définie que si f est une fonction d'une unique variable (scalaire). En revanche, la fonction peut être à valeurs scalaires ou vectorielles sans qu'il soit nécessaire de changer cette définition. Précisément, une fonction vectorielle $f = (f_1, \dots, f_m)$ sera dite dérivable en x si et seulement si toutes ses composantes – qui sont des fonctions scalaires – sont dérivables; on a alors

$$(f'(x))_i = f'_i(x).$$

Autrement dit, on peut dériver composante par composante. Cette approche se généralise de façon directe au cas des fonctions à valeurs matricielles.

Théorème – Développement limité au premier ordre

Soient U un ouvert de \mathbb{R} et $f : U \rightarrow \mathbb{R}^m$. La fonction f est *dérivable* en $x \in U$ si et seulement si il existe un vecteur $\ell \in \mathbb{R}^m$ tel que

$$f(x+h) = f(x) + \ell h + o(|h|).$$

Le vecteur ℓ est alors égal à $f'(x)$.

Démonstration Supposons que le taux d'accroissement de f ait une limite ℓ en x et considérons la fonction ε , à valeurs dans \mathbb{R}^m , définie quand $x+h \in U$ par $\varepsilon(0) = 0$ et si $h \neq 0$ par

$$\varepsilon(h) = \frac{f(x+h) - f(x)}{|h|} - \ell \frac{h}{|h|}.$$

Puisque U est ouvert, la fonction ε est définie dans un voisinage de $h = 0$; par construction, pour tout h on a $f(x + h) = f(x) + \ell h + \varepsilon(h)|h|$. Finalement, f étant dérivable en x de dérivée ℓ , comme pour $h \neq 0$,

$$\varepsilon(h) = \left(\frac{f(x + h) - f(x)}{h} - \ell \right) \frac{h}{|h|}$$

on a bien $\lim_{h \rightarrow 0} \varepsilon(h) = 0$. Par conséquent, avec la notation de Landau,

$$f(x + h) = f(x) + \ell h + o(|h|).$$

Réciproquement, si l'égalité $f(x + h) = f(x) + \ell h + \varepsilon(h)|h|$ est satisfaite avec un $\varepsilon(h)$ qui soit un $o(1)$, comme

$$\frac{f(x + h) - f(x)}{h} = \ell + \varepsilon(h) \frac{h}{|h|}$$

le taux d'accroissement de f en x tend bien vers ℓ quand h tend vers 0. ■

Fonctions linéaires d'une variable scalaire

Le terme ℓh dans le développement limité au premier ordre de f en x est une fonction linéaire de h . Cette remarque n'est pas anodine car toutes les applications linéaires de \mathbb{R} dans \mathbb{R}^m sont de cette forme. En effet, pour une telle fonction L et pour tout $h \in \mathbb{R}$,

$$L \cdot h = L \cdot (h \times 1) = h(L \cdot 1) = (L \cdot 1)h,$$

le vecteur $\ell = L \cdot 1$ convient donc. On peut donc caractériser la dérivabilité de f en x par l'existence d'une fonction linéaire de \mathbb{R}^m dans \mathbb{R} telle que

$$f(x) = f(x + h) + L \cdot h + o(|h|).$$

Cette caractérisation de la dérivée est directement généralisable au cas de fonctions à n variables.

Définition – Différentielle de Fréchet

Soient U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}^m$. La fonction f est *différentiable* en $x \in U$ s'il existe une application linéaire $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$ telle que $f(x + h) = f(x) + L \cdot h + o(\|h\|)$. Si c'est le cas, l'application L est unique ; nous la notons alors $df(x)$ et l'appelons *différentielle de f en x* . Elle est donc caractérisée par :

$$f(x + h) = f(x) + df(x) \cdot h + o(\|h\|).$$

La fonction f est *différentiable* (ou *différentiable sur U*) si elle est différentiable en tout point de U .

Définition – Variation d’une fonction

On appelle *variation de f en x* , pour la variation h de l’argument, la grandeur

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

Remarque – Variation et différentielle

La différentielle de f en x , quand elle existe, constitue la “meilleure” approximation linéaire de la variation de f en x , car c’est la seule telle que

$$\Delta f(x, h) = df(x) \cdot h + o(\|h\|).$$

En particulier, quand la fonction f est affine, la fonction linéaire associée est sa différentielle.

Différentielle d’une fonction affine

Toute fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ de la forme $f(x) = A \cdot x + b$ où $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ est linéaire et $b \in \mathbb{R}^m$, est différentiable en tout point x de \mathbb{R}^n et $df(x) = A$.

Démonstration Il suffit de constater que

$$\Delta f(x, h) = f(x + h) - f(x) = A \cdot (x + h) + b - A \cdot x - b = A \cdot h$$

et que par conséquent $\Delta f(x, h) = A \cdot h + o(\|h\|)$. ■

Résumons les liens entre dérivée et différentielle à ce stade :

Théorème – Différentielle et dérivée

Soient U un ouvert de \mathbb{R} , $f : U \rightarrow \mathbb{R}^m$ et $x \in U$. La fonction f est différentiable en x si et seulement si elle est dérivable en x . Dérivée et différentielle de f en x se déduisent alors l’une de l’autre par les relations

$$f'(x) = df(x) \cdot 1 \quad \text{et} \quad df(x) = (h \in \mathbb{R} \mapsto f'(x)h).$$

Démonstration Une conséquence de la caractérisation de la dérivabilité des fonctions par l’existence de développement limité au premier ordre (p. 10) et de la caractérisation des fonctions linéaires d’une variable scalaire (p. 11). ■

Définition – Gradient

Soient U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$ différentiable en $x \in U$. Le *gradient* de f en x noté $\nabla f(x)$ est l'unique vecteur de \mathbb{R}^n tel que pour tout $h \in \mathbb{R}^n$,

$$df(x) \cdot h = \langle \nabla f(x), h \rangle.$$

Démonstration (existence et unicité) La différentielle de f en x est une forme linéaire sur \mathbb{R}^n , c'est-à-dire une application linéaire de \mathbb{R}^n dans \mathbb{R} . Or pour toute application A de ce type, si un vecteur $a \in \mathbb{R}^n$ est tel que $A \cdot h = \langle a, h \rangle$ pour tout $h \in \mathbb{R}^n$, alors sélectionner successivement $h = e_i$ pour $i = 1, \dots, n$ fournit nécessairement $a = (A(e_1), \dots, A(e_n))$; il existe donc au plus un vecteur a satisfaisant ces égalités. Réciproquement, pour ce vecteur a , on a bien

$$A \cdot h = A \cdot (h_1 e_1 + \dots + h_n e_n) = \sum_i h_i A(e_i) = \sum_i a_i h_i = \langle a, h \rangle.$$

Dans notre contexte où $A = df(x)$, le gradient est donc défini de façon unique par $\nabla f(x) = (df(x) \cdot e_1, \dots, df(x) \cdot e_n)$. ■

Point critique

Soient U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$, une fonction différentiable. Le point x est un *point critique* de f si $df(x) = 0$.

Point critique et extrema

Soient U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$ une fonction différentiable. Si f admet un minimum ou un maximum local en $x \in U$, alors x est un point critique de f .

Démonstration Supposons que f admette un minimum local en x (le cas du maximum peut s'en déduire en considérant la fonction $-f$). Soit $r > 0$ tel que pour tout $y \in \mathbb{R}^n$ satisfaisant $\|y - x\| \leq r$ on ait $y \in U$ et $f(x) \leq f(y)$. Soit $h \in \mathbb{R}^n$ et t un réel non nul suffisamment petit pour que $\|th\| \leq r$. Comme f est différentiable en x , il existe une fonction ε qui soit un petit o de 1 telle que

$$f(x + th) - f(x) = df(x) \cdot (th) + \varepsilon(th)\|th\|.$$

Soit par linéarité de la différentielle,

$$df(x) \cdot h = \frac{f(x + th) - f(x)}{t} - \varepsilon(th) \frac{|t|}{t} \|h\|.$$

En faisant tendre t vers 0 dans le membre de droite de cette équation (la limite existe puisque le membre de droite est indépendant de t), on obtient un nombre positif ou nul. Le même raisonnement pouvant être appliqué pour calculer $df(x) \cdot (-h) = -df(x) \cdot h$, on en déduit que $df(x) \cdot h = 0$. ■

Différentiation composante par composante

Soient U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}^m$. La fonction $f = (f_1, \dots, f_m)$ est différentiable en $x \in U$ si et seulement si chacune de ses composantes f_i est différentiable en x . On a alors pour tout $h \in \mathbb{R}^n$

$$(df(x) \cdot h)_i = df_i(x) \cdot h.$$

Démonstration Supposons f différentiable en x ; soit ε un $o(1)$ tel que

$$f(x+h) = f(x) + df(x) \cdot h + \varepsilon(h)\|h\|.$$

En prenant la i -ème composante de cette equation, on obtient

$$f_i(x+h) = f_i(x) + (df(x) \cdot h)_i + \varepsilon_i(h)\|h\|.$$

On constate alors que l'application $h \mapsto (df(x) \cdot h)_i$ est linéaire (l'application “prendre la i -ème composante d'un vecteur de \mathbb{R}^m ” étant linéaire) et que ε_i est un $o(1)$. La i -ème composante i de f est donc différentiable et $df_i(x) \cdot h = (df(x) \cdot h)_i$.

Réciproquement, si toutes les composantes de f sont différentiables en x , c'est-à-dire si il existe pour chaque i une fonction ε_i qui soit un $o(1)$ et telle que

$$f_i(x+h) = f_i(x) + df_i(x) \cdot h + \varepsilon_i(h)\|h\|,$$

on a

$$f(x+h) = f(x) + (df_1(x) \cdot h, \dots, df_m(x) \cdot h) + \varepsilon(h)\|h\|,$$

et $\varepsilon := (\varepsilon_1, \dots, \varepsilon_m)$ est bien un $o(1)$. Comme la fonction $h \mapsto (df_1(x) \cdot h, \dots, df_m(x) \cdot h)$ est linéaire en h , on en déduit que f est différentiable en x . ■

Remarque – Valeurs et variables matricielles

On peut de façon très simple définir la différentielle d'une fonction F d'une ou de plusieurs variables scalaires et dont la valeur est matricielle, c'est-à-dire telle que

$$F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^{m \times p}.$$

Il suffit en effet d'exiger que chaque composante F_{ij} soit différentiable, puis de définir $dF(x)$ composante par composante, de façon similaire au cas vectoriel:

$$(dF(x) \cdot h)_{ij} = dF_{ij}(x) \cdot h.$$

Mais le traitement des fonctions dont les arguments sont matriciels – par exemple l'application trace $\text{tr} : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ – demande une autre approche. Dans le cas

d'une unique variable matricielle¹, donc d'une fonction de la forme

$$f : U \subset \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^p,$$

on utilisera la fonction auxiliaire f^* dont l'argument est un vecteur de $\mathbb{R}^{m \times n}$, argument qu'on "remettra sous forme matricielle" (cf. section "Mise à plat des matrices (p. 5)") avant de le fournir comme argument à f . C'est-à-dire que la fonction auxiliaire f^* est définie à partir de $\pi := \pi_{m \times n}$ comme

$$f^* = f \circ \pi^{-1}$$

On dira alors que f est différentiable en X si et seulement si f^* est différentiable en $x = \pi(X)$ et si c'est le cas, on définira la fonction $df(X) : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^p$ par

$$df(X) = df^*(x) \circ \pi.$$

Ces deux façons d'étendre la notion de la différentielle – aux fonction à valeurs matricielles et à arguments matriciels – peuvent être combinées.

Prenons un exemple du second cas; la fonction

$$\text{tr} : A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in \mathbb{R}^{2 \times 2} \mapsto a_{11} + a_{22} \in \mathbb{R}.$$

a pour fonction auxiliaire

$$\text{tr}^* : a = (a_{11}, a_{12}, a_{21}, a_{22}) \in \mathbb{R}^4 \mapsto a_{11} + a_{22} \in \mathbb{R}.$$

Pour tout $h = (h_1, h_2, h_3, h_4) \in \mathbb{R}^4$, on a

$$\text{tr}^*(a + h) = \text{tr}^*(a) + h_{11} + h_{22}.$$

L'application $h \mapsto h_{11} + h_{22}$ étant linéaire, tr^* est différentiable en a , de différentielle $d\text{tr}^*(a) \cdot h = h_{11} + h_{22}$. L'application tr est donc différentiable en A et

$$d\text{tr}(A) \cdot H = h_{11} + h_{22} \quad \text{avec} \quad H = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix},$$

c'est-à-dire $d\text{tr}(A) = \text{tr}$.

Remarque – Domaine de définition non ouvert

La définition de la différentielle de f suppose que le domaine de définition de f soit un ensemble ouvert. Cette restriction permet de garantir qu'en tout point x

1. Si la fonction f a plusieurs arguments matriciels

$$A_1 \in \mathbb{R}^{m_1 \times n_1}, \dots, A_k \in \mathbb{R}^{m_k \times n_k},$$

on pourra définir une fonction auxiliaire f^* qui reconstruit ces matrices à partir des éléments d'un unique vecteur de \mathbb{R}^n où $n = m_1 n_1 + \dots + m_k n_k$, en procédant par exemple de gauche à droite, de haut en bas, et de la première à la dernière matrice.

considéré du domaine de définition, on puisse examiner la variation de f en x dans “toutes les directions” pour voir s’il existe une approximation linéaire.

Il y a néanmoins des façons de s’adapter quand le domaine de définition de f n’est pas ouvert :

- Si x est un point de l’intérieur de ce domaine, on peut alors considérer la restriction de f à un voisinage ouvert de x et étudier la différentiabilité de cette restriction. Le résultat (existence de la différentielle et valeur le cas échéant) est indépendant du voisinage ouvert choisi.
- Si x est un point de la frontière de ce domaine, on peut à l’inverse chercher s’il existe une extension de f à un voisinage ouvert de x qui soit différentiable en x . En général cette approche ne garantit pas une définition unique de la différentielle de f en x , mais est suffisante dans des cas importants. Par exemple, elle permet d’étudier la différentiabilité (ou dérivabilité) de fonctions d’une variable scalaire sur des intervalles fermés de \mathbb{R} (p. ??).

Différencier une expression

L’expression $df(x) \cdot h$ dépend de trois éléments : la fonction f , le point de référence x et la variation de l’argument h . Cette notation est sans ambiguïté mais peut parfois être lourde à manipuler. Dans le calcul des dérivées, nous avons l’habitude, pour signifier que la dérivée de la fonction $x \mapsto x^2$ en tout point x de \mathbb{R} est $2x$, d’écrire simplement

$$(x^2)' = 2x.$$

Le membre de gauche désigne la dérivée de la fonction $x \mapsto x^2$, évaluée en x . Avec notre notation pour la différentielle, à ce stade il nous faudrait écrire:

$$d(x \in \mathbb{R} \rightarrow x^2)(x) \cdot h = 2xh.$$

Si l’on accepte de regrouper la fonction à différencier et le point où elle est calculée en un terme unique dans cette notation, qui est une expression de x , on peut alors écrire:

$$dx^2 \cdot h = 2xh,$$

ce qui est un progrès, même si la notation n’est pas totalement dénuée d’ambiguïté². On remarque alors qu’en exploitant cette convention, le terme dx vient à désigner $d(x \mapsto x)(x)$; comme $(x)' = 1$, on a donc $dx \cdot h = 1 \times h = h$. Par conséquent, on peut réécrire l’équation ci-dessus sous la forme mémorable

$$dx^2 = 2x dx.$$

2. par exemple: est-ce que $df(x^2)$ désigne désormais la différentielle de la fonction f évaluée en x^2 ou la différentielle de la fonction $x \mapsto f(x^2)$ évaluée en x ? Les deux grandeurs ne sont pas égales ... Il faut donc savoir si l’on différencie une fonction en un point ou bien une expression par rapport à une variable. On pourra rajouter des parenthèses pour lever l’ambiguïté si nécessaire, avec $d(f)(x^2)$ dans le premier cas et $d(f(x^2))$ dans le second. Par défaut, nous supposons dans la suite que $df(x^2)$ désigne la notation “stricte” $d(f)(x^2)$.

Sous les hypothèses ad hoc, la différentielle de f et g en x est la composée des différentielles de f en x et de g en $y = f(x)$.

Règle de différentiation en chaîne

Soit $f : U \subset \mathbb{R}^p \rightarrow \mathbb{R}^n$ et $g : V \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ deux fonctions définies sur des ouverts U et V et telles que $f(U) \subset V$. Si f est différentiable en $x \in U$ et g est différentiable en $f(x) \in V$, alors la composée $g \circ f$ est différentiable en x et

$$d(g \circ f)(x) = dg(y) \cdot df(x) \quad \text{où } y = f(x).$$

Notations

La formule précédente peut s'écrire de façon plus compacte sans la variable intermédiaire y :

$$d(g \circ f)(x) = dg(f(x)) \cdot df(x).$$

Le terme $dg(f(x))$ y désigne la différentielle de g en $f(x)$ et non la différentielle de l'expression $g(f(x))$ (qui est le terme que l'on souhaite calculer).

Comment souvent, annoter les composants d'une formule avec les ensembles auxquels ils appartiennent permet de s'assurer qu'elle n'est pas de toute évidence incorrecte. Ici par exemple :

$$d(g \circ f)(x) = \overset{\mathbb{R}^m \leftarrow \mathbb{R}^p}{dg(y)} \cdot \overset{\mathbb{R}^m \leftarrow \mathbb{R}^n}{\overset{\mathbb{R}^n \leftarrow \mathbb{R}^p}{df(x)}} \quad \text{où } y = f(x).$$

Démonstration L'objectif de la preuve est de montrer que

$$g(f(x+h)) - g(f(x)) = (dg(f(x)) \cdot df(x)) \cdot h + o(\|h\|).$$

La fonction g étant différentiable en $f(x)$, il existe une fonction ε_1 qui soit un $o(1)$ et telle que

$$g(f(x) + k) - g(f(x)) = dg(f(x)) \cdot k + \varepsilon_1(k)\|k\|.$$

Choisissons $k = f(x+h) - f(x)$ dans cette équation, de telle sorte que

$$g(f(x) + k) = g(f(x) + (f(x+h) - f(x))) = g(f(x+h)).$$

Nous obtenons donc

$$g(f(x+h)) - g(f(x)) = dg(f(x)) \cdot (f(x+h) - f(x)) + \varepsilon_1(k)\|k\|.$$

Notons que la fonction $\varepsilon_2(h) := \varepsilon_1(f(x+h) - f(x))$ est définie dans un voisinage de l'origine et que par continuité de f en x , $f(x+h) - f(x)$ tend vers 0 quand

h tend vers 0, et par conséquent $\varepsilon_2(h) \rightarrow \varepsilon_2(0) = 0$ quand $h \rightarrow 0$; la fonction ε_2 est donc un $o(1)$. Avec cette notation, on a

$$g(f(x+h)) - g(f(x)) = dg(f(x)) \cdot (f(x+h) - f(x)) + \varepsilon_2(h) \|f(x+h) - f(x)\|.$$

Comme f est également différentiable en x , il existe une fonction ε_3 qui soit un $o(1)$ et telle que

$$f(x+h) - f(x) = df(x) \cdot h + \varepsilon_3(h) \|h\|.$$

En substituant cette relation dans la précédente, nous obtenons

$$g(f(x+h)) - g(f(x)) = dg(f(x)) \cdot (df(x) \cdot h) + \varepsilon(h) \|h\|$$

où $\varepsilon(0) = 0$ et dans le cas contraire,

$$\varepsilon(h) = dg(f(x)) \cdot \varepsilon_3(h) + \varepsilon_2(h) \left\| df(x) \cdot \frac{h}{\|h\|} + \varepsilon_3(h) \right\|.$$

Il suffit pour conclure de prouver que $\varepsilon(h) \rightarrow 0$ quand $h \rightarrow 0$. Or,

$$\begin{aligned} \|\varepsilon(h)\| &\leq \|dg(f(x)) \cdot \varepsilon_3(h)\| + \|\varepsilon_2(h)\| \times \|df(x) \cdot (h/\|h\|)\| + \|\varepsilon_2(h)\| \times \|\varepsilon_3(h)\| \\ &\leq \|dg(f(x))\| \times \|\varepsilon_3(h)\| + \|\varepsilon_2(h)\| \times \|df(x)\| + \|\varepsilon_2(h)\| \times \|\varepsilon_3(h)\|, \end{aligned}$$

le résultat est donc acquis. ■

Théorème – Règle de la somme

La somme $(x, y) \in \mathbb{R}^2 \mapsto x + y \in \mathbb{R}$ est différentiable en tout point et

$$d(x+y) = dx + dy.$$

Démonstration Pour tout $(x, y) \in \mathbb{R}^2$ et tout $(h_1, h_2) \in \mathbb{R}^2$, on a

$$(x+h_1) + (y+h_2) = (x+y) + (h_1+h_2).$$

L'application somme est donc différentiable et sa différentielle est l'application $(h_1, h_2) \rightarrow h_1 + h_2$, c'est-à-dire $dx + dy$. ■

Théorème – Règle du produit

L'application produit $(x, y) \in \mathbb{R}^2 \mapsto xy \in \mathbb{R}$ est différentiable en tout point et

$$dxy = xdy + ydx$$

Démonstration Soit $(x, y) \in \mathbb{R}^2$. Pour tout $h = (h_1, h_2) \in \mathbb{R}^2$, on a

$$(x + h_1)(y + h_2) = xy + xh_2 + yh_1 + h_1h_2.$$

Comme $|h_1h_2| \leq \|h\|^2$, le produit h_1h_2 est un $o(h)$. Par conséquent, l'application produit est différentiable en tout point (x, y) de \mathbb{R}^2 et sa différentielle est l'application $(h_1, h_2) \rightarrow xh_2 + yh_1$, c'est-à-dire $xdy + ydx$. ■

Corollaire – Linéarité de la différentielle

Soit U un ouvert de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}^m$ et $g : U \rightarrow \mathbb{R}^m$, différentiables en $x \in U$. Pour tous réels λ et μ , l'application $\lambda f + \mu g$ est différentiable en x et

$$d(\lambda f + \mu g)(x) = \lambda df(x) + \mu dg(x).$$

Démonstration Compte tenu du résultat concernant la différentiation composante par composante (p. 14), il suffit d'établir le résultat pour f et g à valeurs réelles. Or, l'application $x \in \mathbb{R}^n \mapsto (\lambda, f(x))$ est différentiable en x car ses composantes sont différentiables ; sa différentielle – calculée composante par composante – est l'application $h \mapsto (0, df(x) \cdot h)$. L'application λf étant le produit de λ et f , par la règle de différentiation en chaîne (p. 17), elle est différentiable en x et

$$d(\lambda f)(x) = \lambda df(x) + f(x) \times (h \rightarrow 0) = \lambda df(x).$$

De même, μg est différentiable en x et $d(\mu g)(x) = \mu dg(x)$. Par la règle de la somme, la combinaison linéaire $\lambda f + \mu g$ est donc différentiable en x et $d(\lambda f + \mu g)(x) = \lambda df(x) + \mu dg(x)$. ■

Jacobienne, dérivées partielles et directionnelles

Définition – Matrice jacobienne

Soient U un ouvert de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}^m$ et x un point de U . Quand f est différentiable en x , on appelle *matrice jacobienne* de f en x et l'on note $J_f(x)$ la matrice $\mathbb{R}^{m \times n}$ associée à la différentielle $df(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ de f en x .

Définition – Dérivées partielles

Soient U un ouvert de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}^m$ et $x \in U$. Lorsque la i -ème fonction partielle de f en x

$$y_i \mapsto f(x_1, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_n)$$

est dérivable en $y_i = x_i$, on appelle i -ème *dérivée partielle* de f en x et on note $\partial_i f(x) \in \mathbb{R}^m$ sa dérivée. Alternativement,

$$\begin{aligned}\partial_i f(x) &= \lim_{t \rightarrow 0} \frac{f(x + te_i) - f(x)}{t} \\ &= \lim_{t \rightarrow 0} \frac{f(x_1, \dots, x_i + t, \dots, x_n) - f(x_1, \dots, x_n)}{t}\end{aligned}$$

quand le second membre existe.

Remarque – Différentielles partielles

La dérivée partielle permet d'étudier séparément l'influence de chaque variable scalaire de f sur sa variation. Mais dans certaines situations il est plus naturel de regrouper les variables dont dépend f en plusieurs variables vectorielles. Ainsi, pour étudier l'application produit scalaire dans \mathbb{R}^n

$$\langle \cdot, \cdot \rangle : (x_1, \dots, x_n, y_1, \dots, y_n) \in \mathbb{R}^{2n} \mapsto \sum_{k=1}^n x_k y_k \in \mathbb{R},$$

il est raisonnable de partitionner la variable en $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ d'une part et $y = (y_1, \dots, y_n) \in \mathbb{R}^n$ d'autre part. Pour gérer ce type de situation, la *différentielle partielle* par rapport au i -ème argument d'une fonction

$$f : U \subset \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_k} \rightarrow \mathbb{R}^m$$

est définie comme la différentielle de la i -ème fonction partielle de f en $x = (x_1, \dots, x_k)$

$$y_i \in \mathbb{R}^{n_i} \mapsto f(x_1, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_k)$$

quand celle-ci existe. Elle est notée $\partial_i f(x)$, comme la dérivée partielle, ce qui n'est pas trop ambigu tant que l'on explicite comment l'argument de f est décomposé.

Arguments nommés

Les conventions attribuant un nom aux arguments d'une fonction permettent parfois de rendre les dérivées et différentielles partielles plus intelligibles. Si le i -ème argument d'une fonction f est désigné par un symbole x , on pourra noter $\partial_x f$ (ou $\partial f / \partial x$) sa dérivée partielle (ou différentielle partielle selon le cas) au lieu de $\partial_i f$.

Si l'on considère par exemple la fonction m définie par

$$m : (x, y, z, t) \in \mathbb{R}^4 \rightarrow x^2 + y^2 + z^2 - c^2 t^2,$$

comme dans la théorie de la relativité, les dérivées partielles par rapport aux variables d'espace x, y, z sont données par $\partial_x m(x, y, z, t) = 2x$, $\partial_y m(x, y, z, t) =$

$2y$, $\partial_z m(x, y, z, t) = 2z$ et par rapport à la variable de temps t par $\partial_t m(x, y, z, t) = -2c^2 t$. Si l'on préfère regrouper les trois premiers arguments en un vecteur d'espace $\mathbf{x} = (x, y, z)$, on a alors

$$m : (\mathbf{x}, t) \in \mathbb{R}^3 \times \mathbb{R} \rightarrow \|\mathbf{x}\|^2 - c^2 t^2,$$

et la différentielle partielle

$$\partial_{\mathbf{x}} m(\mathbf{x}, t) = 2(xdx + ydy + zdz).$$

Proposition – Différentielle et dérivées partielles

Soient U un ouvert de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}^m$ et x un point de U . Lorsque f est différentiable en x , toutes ses dérivées partielles existent et vérifient

$$\partial_i f(x) = df(x) \cdot e_i,$$

ou de façon équivalente, pour tout $h \in \mathbb{R}^n$

$$df(x) \cdot h = \sum_{i=1}^n \partial_i f(x) h_i.$$

Différentielle et différentielles partielles

Sous les même hypothèses, si l'on considère désormais f comme une fonction $f : U \subset \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k} \rightarrow \mathbb{R}^m$ (avec $n_1 + \cdots + n_k = n$), toutes les différentielles partielles de f en x existent et pour tout $h = (h_1, \dots, h_k) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$,

$$df(x) \cdot h = \sum_{i=1}^n \partial_i f(x) \cdot h_i.$$

Démonstration La différentiabilité de f en x établit l'existence d'une fonction ε qui soit un $o(1)$ et telle que

$$f(x + h) = f(x) + df(x) \cdot h + \varepsilon(h) \|h\|.$$

Soit t un réel non nul ; substituer $h := te_i$ dans cette relation fournit

$$f(x + te_i) = f(x) + df(x) \cdot (te_i) + \varepsilon(te_i) \|te_i\|.$$

En exploitant la linéarité de la différentielle, on obtient donc

$$df(x) \cdot e_i = \frac{f(x + te_i) - f(x)}{t} + \varepsilon(te_i) \frac{|t|}{t}.$$

Par conséquent, en passant à la limite quand $t \rightarrow 0$, on obtient

$$df(x) \cdot e_i = \lim_{t \rightarrow 0} \frac{f(x + te_i) - f(x)}{t} =: \partial_i f(x)$$

Pour obtenir la seconde forme de cette relation, il suffit de décomposer un vecteur $h = (h_1, \dots, h_n)$ sous la forme

$$h = h_1 e_1 + \dots + h_n e_n$$

et d'exploiter la linéarité de la différentielle ; on obtient

$$df(x) \cdot h = df(x) \cdot (h_1 e_1 + \dots + h_n e_n) = \sum_i (df(x) \cdot e_i) h_i = \sum_i \partial_i f(x) h_i,$$

comme attendu. ■

Corollaire – Matrice jacobienne et dérivées partielles

Soient U un ouvert de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}^m$ et x un point de U . Si f est différentiable en x , on a

$$[J_f(x)]_{ij} = \partial_j f_i(x),$$

c'est-à-dire

$$J_f(x) = \begin{bmatrix} \partial_1 f_1(x) & \partial_2 f_1(x) & \cdots & \partial_n f_1(x) \\ \partial_1 f_2(x) & \partial_2 f_2(x) & \cdots & \partial_n f_2(x) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_1 f_m(x) & \partial_2 f_m(x) & \cdots & \partial_n f_m(x) \end{bmatrix}$$

Démonstration Par définition, la matrice jacobienne de f en x se déduit de la différentielle par $[J_f(x)]_{ij} = (df(x) \cdot e_j)_i$. Comme $\partial_j f(x) = df(x) \cdot e_j$, on a $[J_f(x)]_{ij} = (\partial_j f(x))_i$. Les fonctions vectorielles se dérivant composante par composante, on en déduit que $[J_f(x)]_{ij} = \partial_j f_i(x)$. ■

Gradient et dérivées partielles

Soient U un ouvert de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}$ et x un point de U . Si f est différentiable en x , on a

$$\nabla f(x) = (\partial_1 f(x), \dots, \partial_n f(x)).$$

Démonstration Pour tout $h \in \mathbb{R}^n$, on a

$$\langle \nabla f(x), h \rangle = df(x) \cdot h = \sum_i \partial_i f(x) \cdot h_i = \langle (\partial_1 f(x), \dots, \partial_n f(x)), h \rangle,$$

ce qui établit le résultat. ■

Remarque – Matrice jacobienne, gradient et dérivées partielles

On remarquera qu'avec les résultats ci-dessus, il est techniquement possible de définir le gradient $\nabla f(x)$ ou la matrice jacobienne $J_f(x)$ en supposant uniquement que les dérivées partielles de f en x existent, ce qui peut arriver alors que f n'est pas différentiable en x . Mais cette extension est à prendre avec précaution. En effet, dans ce cadre étendu, on ne peut plus transposer aux gradients et matrices jacobiennes tous les résultats valides pour les différentielles. Par exemple, si $J_g(f(x))$ et $J_f(x)$ existent (au sens où toutes les dérivées partielles concernées existent), on peut former le produit matriciel $J_g(f(x))J_f(x)$, mais sans aucune garantie que $J_{g \circ f}(x)$ existe et/ou soit égal à ce produit, car la règle de différentiation en chaîne (p. 17) requiert l'existence des différentielles.

Proposition – Fonction continûment différentiable

Soit U un ouvert de \mathbb{R}^n . Une fonction $f : U \rightarrow \mathbb{R}^m$ est *continûment différentiable* si elle est différentiable et si l'application différentielle

$$df : x \in U \subset \mathbb{R}^n \mapsto df(x) \in (\mathbb{R}^n \xrightarrow{\ell} \mathbb{R}^m)$$

est continue.

Cette définition constitue un moyen de vérifier l'existence de la différentielle (et sa continuité) en passant par les dérivées partielles :

Proposition – Dérivées partielles continues

Soit U un ouvert de \mathbb{R}^n . Une fonction $f : U \rightarrow \mathbb{R}^m$ est continûment différentiable si et seulement si toutes ses dérivées partielles existent et sont continues.

Démonstration Si f est différentiable de différentielle continue, ses dérivées partielles sont définies et $\partial_i f(x) = df(x) \cdot e_i$. Le second membre de cette relation est une fonction continue de x , donc les dérivées partielles sont continues.

Réciproquement, soit $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction dont les dérivées partielles existent et sont continues (la preuve dans le cas d'une fonction à valeurs vectorielles se déduit du résultat dans le cas scalaire). Soit $a \in U$ et $r > 0$ telle que la boule fermée centrée en a et de rayon r soit dans U ; soit $h \in \mathbb{R}^n$ tel que $\|h\| \leq r$. La variation de f entre a et $a + h$ satisfait

$$f(a + h) - f(a) = \sum_{i=1}^n f(a + (h_1, \dots, h_i, 0, \dots)) - f(a + (h_1, \dots, h_{i-1}, 0, \dots)).$$

Or, par le théorème fondamental du calcul (p. 25), comme pour tout i la fonction

$$t \in [0, 1] \mapsto f(a + (h_1, \dots, th_i, 0, \dots))$$

est dérivable de dérivée $\partial_i f(a + (h_1, \dots, th_i, 0, \dots))h_i$, on a

$$f(a + (h_1, \dots, h_i, 0, \dots)) - f(a + (h_1, \dots, h_{i-1}, 0, \dots)) = h_i \int_0^1 \partial_i f(a + (h_1, \dots, th_i, 0, \dots)) dt.$$

Par ailleurs, comme

$$\partial_i f(a)h_i = h_i \int_0^1 \partial_i f(a) dt,$$

on a

$$f(a + h) - f(a) - \sum_i \partial_i f(a)h_i = \sum_{i=1}^n h_i \int_0^1 [\partial_i f(a + (h_1, \dots, th_i, 0, \dots)) - \partial_i f(a)] dt.$$

Par continuité des dérivées partielles en a , si r est choisi de telle sorte que $|\partial_i f(b) - \partial_i f(a)| \leq \varepsilon/n$ quand $|b - a| \leq r$, alors l'inégalité triangulaire et la majoration des intégrales (p. ??) ci-dessus conduisent à

$$\left| f(a + h) - f(a) - \sum_i \partial_i f(a)h_i \right| \leq \sum_{i=1}^n |h_i| \varepsilon/n \leq \varepsilon \|h\|.$$

La fonction f est donc différentiable en a , de différentielle $h \mapsto \sum_i \partial_i f(a)h_i$. La matrice (jacobienne) représentant $df(x)$ ayant pour coefficients les dérivées partielles de f en x , elle est une fonction continue de x , comme $df(x)$. ■

Remarque – Matrice jacobienne continue

S'il l'on adopte la définition étendue de jacobien de la remarque "Matrice jacobienne, gradient et dérivées partielles (p. 23)", la définition de continûment différentiable peut être reformulée comme "la matrice jacobienne existe et est continue".

Variation des fonctions

Différentielle et intégrale

Pour comparer $f(a + h)$ et $f(a)$, lorsque la fonction f est continue en a , nous disposons de l'égalité $f(a + h) = f(a) + o(1)$, mais cette relation est asymptotique. Pour maîtriser l'écart entre $f(a + h)$ et $f(a)$ au moyen de cette formule, nous devons être en mesure de faire tendre h vers 0. Si la grandeur h est fixée, cette relation est inexploitable.

Toutefois, dans cette situation, si f est différentiable sur tout le segment $[a, a+h]$, il est possible de relier $f(a+h)$ à $f(a)$ en intégrant les variations infinitésimales de f le long de $[a, a+h]$. La seule notion d'intégrale dont nous avons besoin, minimaliste et construite exclusivement au service du calcul différentiel, est l'intégrale de Newton, présentée en annexe (p. 37) ; dans de ce chapitre, c'est toujours cette intégrale dont nous ferons usage implicitement.

Théorème – Théorème fondamental du calcul

Si $f : [a, b] \rightarrow \mathbb{R}$ est dérivable, alors f' est intégrable et

$$f(b) - f(a) = \int_a^b f'(x) dx.$$

Démonstration Cf. l'annexe consacrée à l'intégrale de Newton (p. 37). ■

Théorème – Variation d'une fonction

Soient U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}^m$, soient $a \in U$ et $h \in \mathbb{R}^n$ tels que le segment

$$[a, a+h] = \{a+th \mid t \in [0, 1]\}$$

soit inclus dans U . Si f est différentiable en tout point de $[a, a+h]$,

$$f(a+h) = f(a) + \int_0^1 df(a+th) \cdot h dt.$$

Démonstration

Considérons la fonction $\phi : [0, 1] \rightarrow \mathbb{R}^m$ définie par

$$\phi(t) = f(a+th)$$

La fonction ϕ est dérivable sur $[0, 1]$ comme composée des fonctions différentiables f et $t \mapsto a+th$; sa dérivée est donnée par

$$\begin{aligned} \phi'(t) &= d\phi(t) \cdot 1 \\ &= (df(a+th) \cdot d(t \mapsto a+th)) \cdot 1 \\ &= df(a+th) \cdot (d(t \mapsto a+th) \cdot 1) \\ &= df(a+th) \cdot (t \mapsto a+th)' \\ &= df(a+th) \cdot h \end{aligned}$$

Par le théorème fondamental du calcul, on a donc

$$f(a+h) - f(a) = \phi(1) - \phi(0) = \int_0^1 \phi'(t) dt = \int_0^1 df(a+th) \cdot h dt.$$

Théorème – Inégalité des accroissements finis I

Soit $f : [a, a + h] \rightarrow \mathbb{R}^m$ où $a \in \mathbb{R}$, $h \in [0, +\infty[$. Si f est dérivable sur $[a, a + h]$ et M est un majorant de $\|f'\|$, c'est-à-dire si

$$\text{pour tout } t \in [a, b], \|f'(t)\| \leq M.$$

Alors

$$\|f(a + h) - f(a)\| \leq Mh.$$

Démonstration Par construction, la fonction f' est intégrable au sens de Newton et

$$f(a + h) - f(a) = \int_a^b f'(t) dt.$$

Elle est donc également intégrable au sens de Henstock-Kurzweil (cf. chapitre “Calcul Intégral I”); en combinant la définition de l'intégrale de Henstock-Kurzweil et le lemme de Cousin, on peut trouver des approximations arbitrairement précises de l'intégrale de f' par des sommes de Riemann³ : pour tout $\varepsilon > 0$, il existe une subdivision pointée \mathcal{D} de l'intervalle $[a, b]$ telle que

$$\|f(a + h) - f(a) - S(f', \mathcal{D})\| = \left\| \int_a^{a+h} f'(t) dt - S(f', \mathcal{D}) \right\| \leq \varepsilon.$$

En exploitant l'inégalité triangulaire, on obtient donc

$$\|f(a + h) - f(a)\| \leq \|S(f', \mathcal{D})\| + \varepsilon.$$

Supposons que $\mathcal{D} = \{(t_i, [x_i, x_{i+1}]) \mid 0 \leq i \leq k-1\}$. En utilisant à nouveau l'inégalité triangulaire, on peut majorer en norme la somme de Riemann $S(f', \mathcal{D})$:

$$\|S(f', \mathcal{D})\| = \left\| \sum_{i=0}^{k-1} f'(t_i)(x_{i+1} - x_i) \right\| \leq \sum_{i=0}^{k-1} \|f'(t_i)\| |x_{i+1} - x_i|.$$

Comme $\|f'(t_i)\| \leq M$ pour tout $i \in \{0, \dots, k-1\}$,

$$\sum_{i=0}^{k-1} \|f'(t_i)\| |x_{i+1} - x_i| \leq \sum_{i=0}^{k-1} M |x_{i+1} - x_i| \leq M \sum_{i=0}^{k-1} |x_{i+1} - x_i|$$

Finalement, comme $a = x_0 \leq x_1 \leq \dots \leq x_k = a + h$,

$$\sum_{i=0}^{k-1} |x_{i+1} - x_i| = \sum_{i=0}^{k-1} (x_{i+1} - x_i) = x_k - x_0 = (a + h) - a = h$$

et donc $\|S(f', \mathcal{D})\| \leq Mh$. Par conséquent, $\|f(a + h) - f(a)\| \leq Mh + \varepsilon$ et comme le choix de $\varepsilon > 0$ est arbitraire, on en déduit le résultat cherché : $\|f(a + h) - f(a)\| \leq Mh$. ■

3. l'intégrabilité de f' signifie que quelle que soit la précision $\varepsilon > 0$ cherchée on pourra trouver une jauge telle que pour toute subdivision pointée subordonnée à cette jauge, l'écart entre la somme de Riemann et l'intégrale est au plus ε . Le lemme de Cousin affirme que pour toute jauge il existe effectivement une subdivision pointée qui y soit subordonnée.

Théorème – Inégalité des accroissements finis II

Soient U un ouvert de \mathbb{R}^n , et $f : U \rightarrow \mathbb{R}^m$ supposée différentiable en tout point d'un segment $[a, a + h]$ inclus dans U et dont la différentielle est majorée en norme par M sur $[a, a + h]$, c'est-à-dire telle que

$$\text{pour tout } x \in [a, a + h], \quad \|df(x)\| \leq M.$$

Alors

$$\|f(a + h) - f(a)\| \leq M\|h\|.$$

Démonstration Considérons la fonction $\phi : t \in [0, 1] \mapsto f(a + th)$. Nous avons déjà montré dans la démonstration de “Variation d’une fonction (p. 25)” que cette fonction est dérivable, de dérivée $\phi'(t) = df(a + th) \cdot h$. De plus,

$$\|\phi'(t)\| = \|df(a + th) \cdot h\| \leq \|df(a + th)\| \|h\| \leq M\|h\|.$$

Par l’inégalité des accroissements finis dans le cas d’une variable réelle (p. 26),

$$\|f(a + h) - f(a)\| = \|\phi(1) - \phi(0)\| \leq M\|h\| \times 1 = M\|h\|.$$

■

Différentielles d’ordre supérieur

Définition – Différentielle d’ordre 2

Soit $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction différentiable dans un voisinage d’un point x de U . On dira que f est *deux fois différentiable en x* si pour tout vecteur h de \mathbb{R}^n , la fonction $x \mapsto df(x) \cdot h$ 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 dans \mathbb{R}^n ,

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

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

$$d^2f(x) \cdot h \cdot k = d(x \mapsto df(x) \cdot h)(x) \cdot k.$$

Remarques

- On peut vérifier que le terme $d(x \mapsto df(x) \cdot h)(x)$ dépend bien linéairement de h , ce qui justifie l’assertion que $d^2f(x)$ est linéaire et la notation “.” lorsqu’elle est appliquée à un argument h .

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

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

ce qui se décline successivement en

$$d^2 f(x) \cdot h \in (\mathbb{R}^n \xrightarrow{\ell} \mathbb{R}^m), \text{ et } (d^2 f(x) \cdot h) \cdot k \in \mathbb{R}^m.$$

- Pour alléger ces notations, on pourra considérer que dans les notations d'espace fonctionnels, le symbole “ \rightarrow ” associe à droite, par exemple:

$$A \rightarrow B \rightarrow C := A \rightarrow (B \rightarrow C),$$

$$A \rightarrow B \rightarrow C \rightarrow D := A \rightarrow (B \rightarrow (C \rightarrow D)).$$

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, par exemple:

$$L \cdot h \cdot k := (L \cdot h) \cdot k,$$

$$L \cdot h \cdot k \cdot l := ((L \cdot h) \cdot k) \cdot l.$$

Proposition – Variation de la différentielle I

Si $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ est une fonction deux fois différentiable en $x \in U$,

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

Remarque – Interprétation du $o(\|k\|)$

L'équation ci-dessus s'applique à des fonctions linéaires de \mathbb{R}^n dans \mathbb{R}^m . Elle doit donc être interprétée comme l'existence d'une fonction E , définie dans un voisinage de 0 dans \mathbb{R}^n , vérifiant

$$E(k) \in \mathbb{R}^n \xrightarrow{\ell} \mathbb{R}^m \text{ et } \lim_{h \rightarrow 0} E(k) = E(0) = 0,$$

telle que

$$df(x+k) = df(x) + (h \mapsto d^2 f(x) \cdot h \cdot k) + E(k)\|k\|.$$

Démonstration Par définition de la différentielle d'ordre 2 en x (p. 27), pour tout vecteur h de \mathbb{R}^n fixé, on a, pour tout vecteur k de \mathbb{R}^n ,

$$df(x+k) \cdot h = df(x) \cdot h + d^2f(x) \cdot h \cdot k + o(\|k\|),$$

c'est-à-dire qu'il existe pour tout h une fonction ε_h , définie dans un voisinage de $0 \in \mathbb{R}^n$, nulle et continue en 0, telle que

$$df(x+k) \cdot h = df(x) \cdot h + d^2f(x) \cdot h \cdot k + \varepsilon_h(k)\|k\|,$$

Pour tout vecteur k non nul, on a

$$\varepsilon_h(k) = \frac{1}{\|k\|} (df(x+k) \cdot h - df(x) \cdot h - d^2f(x) \cdot h \cdot k),$$

le terme $\varepsilon_h(k)$ est donc linéaire en h ; notons $E(k)$ l'application linéaire de \mathbb{R}^n dans \mathbb{R}^m qui est nulle quand $k = 0$ et définie dans le cas contraire par $E(k) \cdot h = \varepsilon_h(k)$. On a donc pour tout h

$$df(x+k) \cdot h = df(x) \cdot h + d^2f(x) \cdot h \cdot k + (E(k) \cdot h)\|k\|,$$

soit

$$df(x+k) = df(x) + (h \mapsto d^2f(x) \cdot h \cdot k) + E(k)\|k\|,$$

Par ailleurs, pour tout couple de vecteurs h et k de \mathbb{R}^n , on a

$$\begin{aligned} \|E(k) \cdot h\| &= \left\| E(k) \cdot \left(\sum_i h_i e_i \right) \right\| \\ &\leq \sum_i \|E(k) \cdot e_i\| |h_i| \\ &\leq \left(\sum_i \|E(k) \cdot e_i\| \right) \|h\| = \left(\sum_i \|\varepsilon_{e_i}(k)\| \right) \|h\| \end{aligned}$$

donc la norme d'opérateur de $E(k)$ vérifie

$$\|E(k)\| \leq \sum_i \|\varepsilon_{e_i}(k)\| \rightarrow 0 \text{ quand } k \rightarrow 0,$$

ce qui prouve le résultat cherché. ■

Variation d'ordre 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 2* de f en x , associée aux variations h et k de l'argument, la grandeur

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

Théorème – Variation et différentielle d'ordre deux

Pour tout $\varepsilon > 0$, il existe un $\eta > 0$ tel que si $\|h\| \leq \eta$ et $\|k\| \leq \eta$, alors

$$\|\Delta^2 f(x, h, k) - d^2 f(x) \cdot h \cdot k\| \leq \varepsilon(\|h\| + \|k\|)^2.$$

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

$$\begin{aligned} e &= (f(x+h+k) - f(x+k)) - (f(x+h) - f(x)) - d^2 f(x) \cdot h \cdot k \\ &= (f(x+h+k) - f(x+h) - d^2 f(x) \cdot h \cdot k) \\ &\quad - (f(x+k) - f(x) - d^2 f(x) \cdot 0 \cdot k) \end{aligned}$$

Par conséquent, si l'on définit g par

$$g(u) = f(x+u+k) - f(x+u) - d^2 f(x) \cdot u \cdot k,$$

la différence vaut $e = g(h) - g(0)$. Cette différence peut être majorée par l'inégalité des accroissements finis (p. 27) : g est différentiable sur le segment $[0, h]$ et

$$dg(u) = df(x+u+k) - df(x+u) - (h \mapsto d^2 f(x) \cdot h \cdot k).$$

Comme

$$\begin{aligned} dg(u) &= (df(x+u+k) - df(x) - (h \mapsto d^2 f(x) \cdot h \cdot (u+k))) \\ &\quad - (df(x+u) - df(x) - (h \mapsto d^2 f(x) \cdot h \cdot u)), \end{aligned}$$

par le théorème contrôlant la variation de la différentielle (p. 28), pour $\varepsilon > 0$ quelconque, comme $\|u+k\| \leq \|h\| + \|k\|$ et $\|u\| \leq \|h\|$, on peut trouver un $\eta > 0$ tel que si $\|h\| < \eta$ et $\|k\| < \eta$, alors

$$\|dg(u)\| \leq \frac{\varepsilon}{2}(\|h\| + \|k\|) + \frac{\varepsilon}{2}\|h\|.$$

Par conséquent, le théorème des accroissements finis (p. 27) fournit

$$\|e\| = \|dg(u) - dg(0)\| \leq \left(\frac{\varepsilon}{2}(\|h\| + \|k\|) + \frac{\varepsilon}{2}\|h\| \right) \|h\| \leq \varepsilon(\|h\| + \|k\|)^2.$$

■

Théorème – Symétrie de la différentielle d'ordre 2

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

$$d^2 f(x) \cdot h \cdot k = d^2 f(x) \cdot k \cdot h.$$

Démonstration Notons au préalable que

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

La variation d'ordre 2 de f en x est donc symétrique par rapport à ses arguments h et k . On peut alors exploiter la relation entre variation d'ordre 2 et différentielle d'ordre 2 (p. 30) en notant que

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

On obtient pour tout $\varepsilon > 0$ et quand h et k sont suffisamment petits,

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

Si h et k sont arbitraires, en substituant th à h et tk à k pour un $t > 0$ suffisamment petit pour que l'inégalité ci-dessus soit valable, comme

$$d^2 f(x) \cdot th \cdot tk - d^2 f(x) \cdot tk \cdot th = t^2 \times (d^2 f(x) \cdot h \cdot k - d^2 f(x) \cdot k \cdot h)$$

et

$$2\varepsilon(\|th\| + \|tk\|)^2 = t^2 \times 2(\|h\| + \|k\|)^2,$$

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

Théorème – Variation de la différentielle II

Si $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ est une fonction deux fois différentiable en $x \in U$,

$$df(x + k) = df(x) + d^2 f(x) \cdot k + o(\|k\|)$$

Démonstration Par le lemme sur la variation de la différentielle (p. 28), on sait que

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

La différentielle d'ordre 2 étant symétrique (p. 30),

$$d^2 f(x) \cdot h \cdot k = d^2 f(x) \cdot k \cdot h = (d^2 f(x) \cdot k) \cdot h,$$

ce qui fournit l'égalité cherchée. ■

Définition – Dérivées partielles d'ordre 2

Soient U un ouvert de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}^m$ et $x \in U$. Soient i et j deux indices dans $\{1, \dots, n\}$. Lorsque la j -ème dérivée partielle de f est définie sur U et admet en x une i -ème dérivée partielle, on l'appelle *dérivée partielle d'ordre 2* de f en x par rapport aux j -ème et i -ème variables et on la note $\partial_{ij}^2 f(x)$:

$$\partial_{ij}^2 f(x) := \partial_i(x \mapsto \partial_j f(x))(x).$$

Proposition – Symétrie des dérivées partielles d'ordre 2

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 , alors pour toute paire d'indice i et j la dérivée partielle $\partial_{ij} f(x)$ existe et

$$\partial_{ij} f(x) = \partial_{ji} f(x) = d^2 f(x) \cdot e_i \cdot e_j.$$

Démonstration Si f est deux fois différentiable, on a $\partial_j f(x) = d(f(x)) \cdot e_j$, puis $\partial_{ij}^2 f(x) = d(d(f(x)) \cdot e_j) \cdot e_i$. Par définition de la différentielle d'ordre 2 (p. 27),

$$d^2 f(x) \cdot e_j \cdot e_i = d(d(f(x)) \cdot e_j) \cdot e_i,$$

on en déduit donc que $\partial_{ij}^2 f(x) = d^2 f(x) \cdot e_j \cdot e_i$. Par symétrie de la différentielle d'ordre 2 (p. 30), $\partial_{ij}^2 f(x) = \partial_{ji}^2 f(x)$. ■

Définition – Hessienne

Soit $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction deux fois différentiable en $x \in U$. On appelle *Hessienne* de f et x et l'on note $\nabla^2 f(x)$ l'application linéaire $\mathbb{R}^n \rightarrow \mathbb{R}^n$ telle que pour tout couple de vecteurs h et k de \mathbb{R}^n

$$d^2 f(x) \cdot h \cdot k = \langle \nabla^2 f(x) \cdot h, k \rangle.$$

La *matrice hessienne* $H_f(x)$ est la matrice associée à cette application linéaire ; elle est donnée par

$$(H_f(x))_{ij} = \partial_{ij}^2 f(x).$$

Démonstration (expression de la matrice hessienne) Elle résulte directement de la définition de $\nabla^2 f(x)$ et des liens entre $d^2 f(x)$ et $\partial_{ij}^2 f(x)$ établis par la proposition “Symétrie des dérivées partielles d'ordre 2” (p. 32). ■

La notion de différentielle d'ordre 2 se généralise sans difficulté à un ordre plus élevé, 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

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

Remarque

On a

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

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. 33),

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

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. 30). 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. 33), 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

Les dérivées partielles d'ordre supérieur se définissent par récurrence, de manière similaire aux dérivées partielles d'ordre 2. Pour simplifier la notation $\partial_{i_1 \dots i_k}^k f(x)$, on exploite 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

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

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 le théorème des accroissements finis (p. 27),

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

ce qui établit que $r(h) = o(\|h\|^j)$.

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

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

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

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

Par le lemme de stratification (p. 33) et la symétrie des différentielles d'ordre supérieur (p. 34), 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 $dr(h) \cdot k = o(\|h\|^{j-1})$ 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 $f : [a, a+h] \rightarrow \mathbb{R}^m$ 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 (p. 25). Si l'on suppose la relation vérifiée à l'ordre j , et f $j+2$ fois dérivable, par intégration par parties (p. 39), on obtient

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

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 $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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. 36), 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 – Intégrale de Newton

Définition – Intégrale de Newton

Soit $f : [a, b] \rightarrow \mathbb{R}^m$. On dit que f est *intégrable au sens de Newton* si elle admet une primitive $F : [a, b] \rightarrow \mathbb{R}^m$. L'intégrale de f entre a et b est alors définie par

$$\int_a^b f(x) dx = F(b) - F(a).$$

La primitive F de f quand elle existe étant déterminée à une constante près, cette définition est non-ambiguë.

Une autre façon de voir les choses : l'intégrale de Newton est définie de telle sorte que le théorème fondamental du calcul (p. 25) soit trivialement satisfait, en toute généralité. Pour d'autres intégrales, comme l'intégrale de Riemann ou l'intégrale de Lebesgue, il sera nécessaire de faire des hypothèses supplémentaires sur la fonction f' (par exemple, f' continue) pour que ce résultat soit valable. L'intégrale de Henstock-Kurzweil, qui sera exposée dans le cours de calcul intégral, vérifie bien le théorème fondamental du calcul en toute généralité : elle étend donc l'intégrale de Newton (et celle de Riemann, ainsi que celle de Lebesgue).

L'intégrale de Newton est un outil assez primitif⁴ et difficile à exploiter ; elle vérifie tout de même quelques propriétés bien utiles.

4. sans mauvais jeu de mots . . .

Proposition – Linéarité

Soit $f : [a, b] \rightarrow \mathbb{R}^m$, $g : [a, b] \rightarrow \mathbb{R}^m$, et λ, μ deux constantes réelles. Si f et g sont intégrables au sens de Newton, $\lambda f + \mu g$ également et

$$\int_a^b \lambda f(x) + \mu g(x) dx = \lambda \int_a^b f(x) dx + \mu \int_a^b g(x) dx.$$

Démonstration Par hypothèse, f a une primitive F , g a une primitive G ,

$$\int_a^b f(x) dx = F(b) - F(a) \quad \text{et} \quad \int_a^b g(x) dx = G(b) - G(a).$$

La fonction $\lambda F + \mu G$ est une primitive de $\lambda f + \mu g$ et donc

$$\begin{aligned} \int_a^b \lambda f(x) + \mu g(x) dx &= (\lambda F(b) + \mu G(b)) - (\lambda F(a) + \mu G(a)) \\ &= \lambda(F(b) - F(a)) + \mu(G(b) - G(a)) \\ &= \lambda \int_a^b f(x) dx + \mu \int_a^b g(x) dx. \end{aligned}$$

■

Théorème – Majoration

Si $f : [a, b] \rightarrow \mathbb{R}$ est une fonction intégrable au sens de Newton telle que $|f| \leq M$,

$$\left| \int_a^b f(x) dx \right| \leq M(b - a).$$

Démonstration La fonction $g : x \in [a, b] \mapsto f(x) - M$ est intégrable au sens de Newton et négative. Si G est une primitive de g , elle est donc décroissante. Par conséquent,

$$\int_a^b (f(x) - M) dx = \int_a^b f(x) dx - M(b - a) = G(b) - G(a) \leq 0.$$

On peut de même montrer en intégrant la fonction $x \in [a, b] \mapsto f(x) + M$ que

$$\int_a^b f(x) dx + M(b - a) \geq 0,$$

ce qui fournit le résultat cherché. ■

Théorème – Intégration par parties

Soit $f : [a, b] \rightarrow \mathbb{R}$ et $g : [a, b] \rightarrow \mathbb{R}$ deux fonctions dérivables. Si la fonction fg' est intégrable au sens de Newton, la fonction $f'g$ également et

$$\int_a^b f'(x)g(x) dx = (f(b)g(b) - f(a)g(a)) - \int_a^b f(x)g'(x) dx.$$

Démonstration Comme $(fg)' = f'g + fg'$, on a $f'g = (fg)' - fg'$. Or, $(fg)'$ est intégrable au sens de Newton (fg est une de ses primitives), fg' est intégrable au sens de Newton par hypothèse, donc $f'g$ est intégrable comme combinaison linéaire de fonctions intégrables. De plus,

$$\begin{aligned} \int_a^b f'(x)g(x) dx &= \int_a^b (fg)'(x) dx - \int_a^b f(x)g'(x) dx \\ &= (f(b)g(b) - f(a)g(a)) - \int_a^b f(x)g'(x) dx. \end{aligned}$$

■

Exercices

Dérivée sur un intervalle fermé

Montrer qu'une fonction f est dérivable sur l'intervalle fermé $[a, b]$ – $f'(a)$ et $f'(b)$ désignant alors les dérivées à droite de f en a et à gauche de f en b – si et seulement si il existe un $\varepsilon > 0$ et une extension g de f sur $]a - \varepsilon, b + \varepsilon[$ tel que g soit dérivable et qu'alors, $f' = g'|_{[a, b]}$. (Solution p. 43.)

Différentiation en chaîne

La règle générale de différentiation en chaîne (p. 17) s'applique à la composée de deux fonctions différentiables $f : U \subset \mathbb{R}^p \rightarrow \mathbb{R}^n$ et $g : V \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$.

Question 1 Calculer $d(g \circ f)$ quand $p = n = 1$ (on utilisera les dérivées de f et g). (Solution p. 44.)

Question 2 Calculer $d(g \circ f)$ quand $p = m = 1$ (on utilisera les dérivées et/ou gradients de f et g). (Solution p. 44.)

Fonction quadratique

Soit $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$ un opérateur linéaire, b un vecteur de \mathbb{R}^n et $c \in \mathbb{R}$. On considère la fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ définie par

$$f(x) = \frac{1}{2} \langle x, A \cdot x \rangle + \langle b, x \rangle + c.$$

Question 1 Montrer que f est 2 fois différentiable en tout point x de \mathbb{R}^n ; calculer $\nabla f(x)$ et $\nabla^2 f(x)$. (Solution p. 44.)

Question 2 Soit $x \in \mathbb{R}^n$; on suppose que $\nabla^2 f(x)$ est inversible. Montrer que la fonction f admet un unique point critique x_0 et le calculer en fonction de x , $\nabla f(x)$ et $\nabla^2 f(x)$. (Solution p. 45.)

Vecteur gaussien

La densité de probabilité associée à un vecteur gaussien $X \in \mathbb{R}^d$ est proportionnelle à la fonction

$$f : x \in \mathbb{R}^d \mapsto \exp \left(-\frac{1}{2} \langle x, \Sigma^{-1} \cdot x \rangle \right)$$

où $\Sigma : \mathbb{R}^d \rightarrow \mathbb{R}^d$ est un opérateur linéaire autoadjoint (c'est-à-dire que $\Sigma^* = \Sigma$) tel que $\langle x, \Sigma \cdot x \rangle > 0$ quand $x \neq 0$.

Question 1 Montrer que la fonction f est différentiable et calculer son gradient. (Solution p. 45.)

Question 2 Montrer que la fonction f est deux fois différentiable et calculer sa hessienne. (Solution p. 46.)

Robot manipulateur

Les coordonnées cartésiennes x et y de l'effecteur final d'un robot dans le plan, composé de deux corps rigides de longueur ℓ_1 et ℓ_2 et d'articulation rotoïdes sont données par

$$\begin{cases} x &= \ell_1 \cos \theta_1 + \ell_2 \cos(\theta_1 + \theta_2) \\ y &= \ell_1 \sin \theta_1 + \ell_2 \sin(\theta_1 + \theta_2) \end{cases}$$

où θ_1 et θ_2 sont les coordonnées articulaires du robot.

Montrer que l'application $f : (\theta_1, \theta_2) \in \mathbb{R}^2 \mapsto (x, y) \in \mathbb{R}^2$ est différentiable et déterminer sa matrice jacobienne. (Solution p. 47.)

Différentiation matricielle

Source: (Tao 2013)

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. 47.)

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. 48.)

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. 48.)

Dérivée partielles, directionnelles et différentielle

Question 1 Construire une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ dont les dérivées partielles existent en $(0, 0)$ mais qui ne soit pas différentiable en ce point. (Solution p. 48.)

Question 2 Construire une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ dont la dérivée dans la direction $h \in \mathbb{R}^2$

$$f'(x, h) := \lim_{t \rightarrow 0} \frac{f(x + th) - f(x)}{t}$$

existe en $x = (0, 0)$ pour tout $h \in \mathbb{R}^2$, mais qui ne soit pas différentiable en ce point. (Solution p. 49.)

Dérivée directionnelle d'Hadamard

Source: (Shapiro 1990)

Soient U un ouvert de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}^m$ et $x \in U$. La fonction f est *directionnellement dérivable* si pour tout vecteur $h \in \mathbb{R}^n$, la dérivée directionnelle

$$f'(x, h) := (t \mapsto f(x + th))'(0) = \lim_{t \rightarrow 0} \frac{f(x + th) - f(x)}{t}$$

est bien définie.

On introduit une variante à cette définition: la fonction f est *directionnellement dérivable au sens de Hadamard* en x si pour tout chemin $\gamma : I \subset \mathbb{R} \rightarrow \mathbb{R}^n$, défini sur un intervalle ouvert I contenant 0, tel que $\gamma(I) \subset U$, $\gamma(0) = x$ et $\gamma'(0)$ existe, la dérivée $(f \circ \gamma)'(0)$ existe.

Question 1 Montrer que si f est directionnellement dérivable au sens de Hadamard en x , alors f est directionnellement dérivable au sens classique. (Solution p. 49.)

Question 2 Montrer que si f est directionnellement dérivable au sens de Hadamard en x , la grandeur $(f \circ \gamma)'(0)$ ne dépend de γ qu'à travers $\gamma'(0)$ et que par conséquent

$$(f \circ \gamma)'(0) = f'(x, \gamma'(0)).$$

(Solution p. 49.)

Question 3 – Dérivation en chaîne Soit $f : U \subset \mathbb{R}^p \rightarrow \mathbb{R}^n$ et $g : V \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ deux fonctions définies sur des ouverts U et V et telles que $f(U) \subset V$. Montrer que si f est directionnellement dérivable au sens de Hadamard en $x \in U$ et que g est directionnellement dérivable au sens de Hadamard en $f(x) \in V$, alors la composée $g \circ f$ est directionnellement dérivable au sens de Hadamard en x et

$$(g \circ f)'(x, h) = g'(f(x), f'(x, h)).$$

(Solution p. 50.)

Question 4 Montrer que f est directionnellement dérivable au sens de Hadamard en x si et seulement si la limite

$$\lim_{(t,k) \rightarrow (0,h)} \frac{f(x + tk) - f(x)}{t}$$

existe et que la limite est alors égale à $f'(x, h)$. (Solution p. 50.)

Question 5 Une fonction dérivable directionnellement au sens de Hadamard en x est *différentiable au sens de Hadamard* en x si de plus $f'(x, h)$ est une fonction linéaire de h . Montrer que f est différentiable en x au sens de Hadamard si et seulement si elle est différentiable en x au sens de Fréchet. (Solution p. 52.)

Inégalité de la valeur moyenne

Soit $f : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}^m$ une fonction intégrable au sens de Newton; on appelle *valeur moyenne de f* la grandeur

$$\langle f \rangle := \frac{1}{b-a} \int_a^b f(x) dx.$$

Quel est le lien entre $\langle f \rangle$ et la grandeur $\sup_{x \in [a,b]} \|f(x)\|$? (Solution p. 53.)

Egalité des accroissements finis ?

Soit $f : [0, 2\pi] \rightarrow \mathbb{R}^2$ la fonction définie par

$$f(t) = (\cos t, \sin t)$$

Peut-on trouver un $t \in [0, 2\pi]$ tel que $f(2\pi) - f(0) = f'(t) \times 2\pi$? (Solution p. 53.)

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. 53.)

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 $h \in \mathbb{R}^n$,

$$d^2 f(x)(\cdot h)^2 = \langle \nabla^2 f(x) \cdot h, h \rangle \geq 0.$$

(Solution p. 54.)

Question 2 Montrer la réciproque de ce résultat. (Solution p. 54.)

Solutions

Dérivée sur un intervalle fermé

Si une fonction g dérivable sur $]a - \varepsilon, b + \varepsilon[$ étend la fonction f définie sur $[a, b]$, il est clair que f est dérivable en tout point de $[a, b]$ et que $g'|_{[a, b]} = f'$.

Réciproquement, si f est dérivable sur $[a, b]$ (à droite en a et à gauche en b), alors la fonction $g :]a - 1, b + 1[$ définie par

$$g(x) = \begin{cases} f(a) + f'(a) \times (x - a) & \text{si } x < a \\ f(x) & \text{si } x \in [a, b] \\ f(b) + f'(b) \times (x - b) & \text{si } x > b \end{cases}$$

étend f et est dérivable par construction.

Différentiation en chaîne

Question 1 Les fonction f et g sont différentiables donc dérivables (cf. Différentielle et dérivée (p. 12)). Comme fonctions d'une variable, en raison du lien entre différentielle et dérivée (p. 12), on a $df(x) \cdot h = f'(x)h$ et $dg(x) \cdot h = g'(x)h$. Par la règle de différentiation en chaîne (p. 17), on obtient

$$d(g \circ f)(x) = dg(f(x)) \cdot df(x).$$

On en déduit

$$\begin{aligned} d(g \circ f)(x) \cdot h &= (dg(f(x)) \cdot df(x)) \cdot h \\ &= dg(f(x)) \cdot (df(x) \cdot h) \\ &= dg(f(x)) \cdot (f'(x)h) \\ &= g'(f(x))(f'(x)h) \\ &= (g'(f(x))f'(x))h. \end{aligned}$$

Question 2 La fonction f dépendant d'une variable scalaire, elle est dérivable et $df(x) \cdot h = f'(x)h$. Quant à g qui est à valeur scalaire, sa différentielle en x est reliée à son gradient par $dg(x) \cdot h = \langle \nabla g(x), h \rangle$. La règle de différentiation en chaîne (p. 17), $d(g \circ f)(x) = dg(f(x)) \cdot df(x)$ se décline donc en

$$\begin{aligned} d(g \circ f)(x) \cdot h &= dg(f(x)) \cdot (df(x) \cdot h) \\ &= dg(f(x)) \cdot (f'(x)h) \\ &= \langle \nabla g(f(x)), f'(x)h \rangle \\ &= \langle \nabla g(f(x)), f'(x) \rangle h. \end{aligned}$$

Fonction quadratique

Question 1 Pour tout $x \in \mathbb{R}^n$ et tout $h \in \mathbb{R}^n$, on a

$$\begin{aligned} f(x+h) - f(x) &= \frac{1}{2} \langle (x+h), A \cdot (x+h) \rangle + \langle b, x+h \rangle + c - \frac{1}{2} \langle x, A \cdot x \rangle - \langle b, x \rangle - c \\ &= \frac{1}{2} \langle x, A \cdot h \rangle + \frac{1}{2} \langle h, A \cdot x \rangle + \langle b, h \rangle + \frac{1}{2} \langle h, A \cdot h \rangle \\ &= \frac{1}{2} \langle A^* \cdot x, h \rangle + \frac{1}{2} \langle A \cdot x, h \rangle + \langle b, h \rangle + \frac{1}{2} \langle h, A \cdot h \rangle. \end{aligned}$$

Comme $|\langle h, A \cdot h \rangle| \leq \|h\| \times \|A\| \|h\|$, ce terme est un $o(\|h\|)$. On en conclut que

$$f(x+h) - f(x) = \left\langle \frac{1}{2}(A + A^*) \cdot x + b, h \right\rangle + o(\|h\|).$$

La fonction f est donc différentiable en x , de gradient

$$\nabla f(x) = \frac{1}{2}(A + A^*) \cdot x + b.$$

Pour tout $h \in \mathbb{R}^n$, la fonction $x \mapsto \langle \nabla f(x), h \rangle$ vérifie

$$\langle \nabla f(x+k), h \rangle - \langle \nabla f(x), h \rangle = \left\langle \frac{1}{2}(A + A^*) \cdot k, h \right\rangle.$$

Elle est donc différentiable et

$$d^2 f(x) \cdot h \cdot k = \left\langle \frac{1}{2}(A + A^*) \cdot k, h \right\rangle.$$

Par symétrie de la différentielle d'ordre 2,

$$d^2 f(x) \cdot h \cdot k = \left\langle \frac{1}{2}(A + A^*) \cdot h, k \right\rangle,$$

donc

$$\nabla^2 f(x) = \frac{1}{2}(A + A^*).$$

Question 2 Si $\nabla^2 f(x)$ est inversible (cet opérateur est constant), comme

$$\nabla f(y) = \frac{1}{2}(A + A^*) \cdot y + b = \nabla^2 f(x) \cdot y + b,$$

résoudre $\nabla f(y) = 0$ revient à rechercher les solutions de

$$\nabla^2 f(x) \cdot y + b = \nabla^2 f(x) \cdot y + (\nabla f(x) - \nabla^2 f(x) \cdot x) = 0.$$

Il existe donc un unique point critique pour f , donné par

$$y = x - (\nabla^2 f(x))^{-1} \nabla f(x).$$

Vecteur gaussien

Question 1 La fonction

$$f : x \in \mathbb{R}^d \mapsto \exp \left(-\frac{1}{2} \langle x, \Sigma^{-1} \cdot x \rangle \right)$$

apparaît comme la composée des fonctions

$$x \in \mathbb{R}^d \mapsto -\frac{1}{2} \langle x, \Sigma^{-1} \cdot x \rangle \quad \text{et} \quad \exp : \mathbb{R} \rightarrow \mathbb{R}.$$

La fonction \exp est dérivable, et donc différentiable sur tout \mathbb{R} avec $d(\exp(y)) = \exp'(y)dy = \exp(y)dy$, c'est-à-dire

$$d \exp(y) \cdot h = \exp(y) \times h.$$

Quand à la première fonction, pour tout $h \in \mathbb{R}^d$, on a

$$-\frac{1}{2} \langle x+h, \Sigma^{-1} \cdot (x+h) \rangle = \\ -\frac{1}{2} (\langle x, \Sigma^{-1} \cdot x \rangle + \langle x, \Sigma^{-1} h \rangle + \langle h, \Sigma^{-1} \cdot x \rangle + \langle h, \Sigma^{-1} \cdot h \rangle).$$

D'une part, comme Σ est autoadjoint (et inversible), Σ^{-1} également et

$$\langle x, \Sigma^{-1} \cdot h \rangle + \langle h, \Sigma^{-1} \cdot x \rangle = 2 \langle \Sigma^{-1} \cdot x, h \rangle,$$

d'autre part

$$|\langle h, \Sigma^{-1} \cdot h \rangle| \leq \|h\| \times \|\Sigma^{-1} \cdot h\| \leq \|h\| \times \|\Sigma^{-1}\| \times \|h\| = o(\|h\|).$$

La fonction est donc différentiable sur \mathbb{R}^n , avec

$$d\left(-\frac{1}{2} (\langle x, \Sigma^{-1} \cdot x \rangle)\right) \cdot h = -\langle \Sigma^{-1} \cdot x, h \rangle.$$

La fonction f est donc différentiable sur \mathbb{R}^d comme composée de fonctions différentiables et l'on a

$$df(x) \cdot h = -\exp\left(-\frac{1}{2} (\langle x, \Sigma^{-1} \cdot x \rangle)\right) \langle \Sigma^{-1} \cdot x, h \rangle = \langle -f(x) \times \Sigma^{-1} \cdot x, h \rangle,$$

le gradient de f vaut donc

$$\nabla f(x) = -f(x) \times \Sigma^{-1} \cdot x.$$

Question 2 De l'équation

$$df(x) \cdot h = \langle -f(x) \times \Sigma^{-1} \cdot x, h \rangle = -f(x) \langle \Sigma^{-1} \cdot h, x \rangle$$

on déduit que $x \mapsto df(x) \cdot h$ est différentiable comme produit de fonctions scalaires différentiables (la fonction $x \mapsto \langle \Sigma^{-1} \cdot h, x \rangle$ étant linéaire). On a de plus

$$d(x \mapsto df(x) \cdot h) \cdot k = -(df(x) \cdot k) \times \langle \Sigma^{-1} \cdot x, h \rangle - f(x) \times \langle \Sigma^{-1} \cdot h, k \rangle \\ = \langle -f(x) \times \Sigma^{-1} \cdot x, k \rangle \langle \Sigma^{-1} \cdot x, h \rangle + \langle -f(x) \times \Sigma^{-1} \cdot h, k \rangle$$

Pour des vecteurs arbitraires u et v dans \mathbb{R}^n , on a

$$\langle u, k \rangle \langle v, h \rangle = \langle k, u \rangle \langle v, h \rangle = k^* \cdot u \times v^* \cdot h = (v \cdot u^* \cdot k)^* \cdot h = \langle (v \cdot u^*) \cdot k, h \rangle,$$

par conséquent

$$d(x \mapsto df(x) \cdot h) \cdot k = -f(x) \langle (\Sigma^{-1} \cdot x \cdot x^* \cdot \Sigma^{-1} + \Sigma^{-1})h, k \rangle.$$

La Hessienne de f en x est donc donnée par

$$\nabla^2 f(x) = -f(x)(\Sigma^{-1} \cdot x \cdot x^* \cdot \Sigma^{-1} + \Sigma^{-1}).$$

Robot manipulateur

Des équations

$$\begin{cases} x &= \ell_1 \cos \theta_1 + \ell_2 \cos(\theta_1 + \theta_2) \\ y &= \ell_1 \sin \theta_1 + \ell_2 \sin(\theta_1 + \theta_2) \end{cases}$$

on déduit que les dérivées partielles de x et de y par rapport à θ_1 et θ_2 existent et vérifient

$$\begin{aligned} \partial_1 x(\theta_1, \theta_2) &= -\ell_1 \sin \theta_1 - \ell_2 \sin(\theta_1 + \theta_2), \\ \partial_2 x(\theta_1, \theta_2) &= -\ell_2 \sin(\theta_1 + \theta_2), \\ \partial_1 y(\theta_1, \theta_2) &= \ell_1 \cos \theta_1 + \ell_2 \cos(\theta_1 + \theta_2), \\ \partial_2 y(\theta_1, \theta_2) &= \ell_2 \cos(\theta_1 + \theta_2). \end{aligned}$$

Ces grandeurs étant continues, la fonction $f = (x, y)$ est continûment différentiable et donc différentiable. Si l'on note $s_1 = \sin \theta_1$, $s_{12} = \sin(\theta_1 + \theta_2)$, $c_1 = \cos \theta_1$ et $c_{12} = \cos(\theta_1 + \theta_2)$, on obtient donc

$$J_f(\theta_1, \theta_2) = \begin{bmatrix} -\ell_1 s_1 - \ell_2 s_{12} & -\ell_2 s_{12} \\ \ell_1 c_1 + \ell_2 c_{12} & \ell_2 c_{12} \end{bmatrix}.$$

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

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

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

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

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

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

Chaque coefficient de $\text{co}(A)^t$ (la transposée de la comatrice de A) est une fonction polynomiale des coefficients a_{ij} de A ; chaque coefficient de $\text{co}(A)^t$ 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 $\text{inv}(A) = A^{-1}$; comme $\text{inv}(I + H) = I + d\text{inv}(I) \cdot H + o(\|H\|)$, l'identité $(I + H)(I + H)^{-1} = I$ fournit :

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

et donc

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

Dérivée partielles, directionnelles et différentielle

Question 1 Les dérivées partielles de la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par

$$f(x, y) = \begin{cases} 0 & \text{si } x = 0 \text{ ou } y = 0, \\ 1 & \text{sinon.} \end{cases}$$

existent en $(0, 0)$ et sont nulles, puisque les fonctions partielles associées sont nulles. Mais f n'est pas continue en l'origine; elle n'y est donc a fortiori pas différentiable.

Question 2 Les dérivées directionnelles de la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par

$$f(x, y) = \begin{cases} 1 & \text{si } x > 0 \text{ et } y = x^2, \\ 0 & \text{sinon.} \end{cases}$$

existent en $(0, 0)$ et sont nulles pour tout $h \in \mathbb{R}^2$, puisque les fonctions associées $t \in \mathbb{R} \mapsto f(th)$ sont nulles pour $|t|$ suffisamment petit. Mais f n'est pas continue en l'origine ; elle n'y est donc a fortiori pas différentiable.

Dérivée directionnelle d'Hadamard

Question 1 Supposons que f soit directionnellement dérivable au sens de Hadamard en x . Pour tout $h \in \mathbb{R}^n$, par continuité de l'application $t \in \mathbb{R} \mapsto x + th$, pour $\varepsilon > 0$ assez petit, et parce que le domaine de définition de f est ouvert, l'image de la fonction

$$\gamma : t \in]-\varepsilon, \varepsilon[\mapsto x + th$$

est incluse dans le domaine de définition de f , est telle que $\gamma(0) = x$, $\gamma'(0) = h$. Par conséquent, la dérivée de $f \circ \gamma$ en 0 existe, et c'est par construction la dérivée directionnelle de f en x dans la direction h . La fonction f est donc directionnellement dérivable en x au sens classique.

Question 2 Supposons que f soit directionnellement dérivable au sens de Hadamard en x . Pour montrer que l'expression $(f \circ \gamma)'(0)$ ne dépend de γ qu'à travers $\gamma'(0)$, nous allons considérer un second chemin arbitraire $\beta : J \rightarrow \mathbb{R}^n$, où J est un intervalle ouvert de \mathbb{R} contenant 0, tel que $\beta(0) = x$, $\beta'(0) = \gamma'(0)$ et montrer que

$$(f \circ \gamma)'(0) = (f \circ \beta)'(0).$$

L'idée de la démonstration consiste à construire un troisième chemin α qui en "mélangeant" les chemins β et γ , satisfait les hypothèses de la définition de "directionnellement dérivable au sens de Hadamard", est tel que $\alpha'(0) = \beta'(0) = \gamma'(0)$ et également tel que d'une part $(f \circ \alpha)'(0) = (f \circ \beta)'(0)$ et d'autre part $(f \circ \alpha)'(0) = (f \circ \gamma)'(0)$.

Un chemin qui permette de tenir ce raisonnement est le suivant. Tout d'abord, choisissons $\varepsilon > 0$ tel que $]-\varepsilon, \varepsilon[\subset I \cap J$, puis définissons $\alpha :]-\varepsilon, \varepsilon[\rightarrow \mathbb{R}^n$ par

$$\alpha(t) = \begin{cases} x & \text{si } t = 0, \\ \beta(t) & \text{si } \varepsilon/2^{2k+1} \leq |t| < \varepsilon/2^{2k}, \text{ pour un entier } k \in \mathbb{N}, \\ \gamma(t) & \text{si } \varepsilon/2^{2k+2} \leq |t| < \varepsilon/2^{2k+1}, \text{ pour un entier } k \in \mathbb{N}. \end{cases}$$

Les hypothèses de la définition sont facilement vérifiées, ainsi que la preuve que $\alpha'(0) = \beta'(0) = \gamma'(0)$. Avec l'hypothèse de différentiabilité au sens de Hadamard, nous savons donc que la dérivée $(f \circ \alpha)'(0)$ existe. On peut la calculer comme la limite de

$$(f \circ \alpha)'(0) = \lim_{k \rightarrow +\infty} \frac{f(\alpha(t_k)) - f(x)}{t_k}$$

où t_k est une suite arbitraire de valeurs non nulles tendant vers 0. Or, si l'on choisit $t_k = \varepsilon/2^{2k+1}$, on trouve

$$\lim_{k \rightarrow +\infty} \frac{f(\alpha(t_k)) - f(x)}{t_k} = \lim_{k \rightarrow +\infty} \frac{f(\beta(t_k)) - f(x)}{t_k} = (f \circ \beta)'(0)$$

et si l'on choisit $t_k = \varepsilon/2^{2k+2}$, on trouve

$$\lim_{k \rightarrow +\infty} \frac{f(\alpha(t_k)) - f(x)}{t_k} = \lim_{k \rightarrow +\infty} \frac{f(\gamma(t_k)) - f(x)}{t_k} = (f \circ \gamma)'(0),$$

ce qui prouve le résultat d'indépendance souhaité. Pour prouver que $(f \circ \gamma)'(0) = f'(x, \gamma'(0))$, il suffit d'associer à un chemin quelconque γ le chemin "canonique" $\beta : t \mapsto x + t\gamma'(0)$ de la question 1, qui est tel que $\beta'(0) = \gamma'(0)$ d'une part et d'autre part $(f \circ \beta)'(0) = f'(x, \beta'(0))$ par construction. On en déduit que

$$(f \circ \gamma)'(0) = (f \circ \beta)'(0) = f'(x, \beta'(0)) = f'(x, \gamma'(0)).$$

Question 3 Soit $\gamma : I \subset \mathbb{R} \rightarrow \mathbb{R}^n$, un chemin défini sur un intervalle ouvert I contenant 0, tel que $\gamma(I) \subset U$, $\gamma(0) = x$ et $\gamma'(0)$ existe. Alors, sous les hypothèses du théorème de dérivée en chaîne que nous souhaitons montrer, le chemin $\beta = f \circ \gamma$ est défini sur I , vérifie $\beta(I) \subset V$, $\beta(0) = f(x)$ et par hypothèse de dérivabilité directionnelle au sens de Hadamard sur f en x , $\beta'(0) = f'(x, \gamma'(0))$. Par hypothèse de dérivabilité directionnelle au sens de Hadamard sur g en $f(x)$,

$$((g \circ f) \circ \gamma)'(0) = (g \circ \beta)'(0) = g'(f(x), \beta'(0)) = g'(f(x), f'(x, \gamma'(0))),$$

ce qui prouve la dérivabilité directionnelle au sens de Hadamard pour la composée $g \circ f$ en x . Il suffit d'associer à un vecteur h le chemin canonique $t \mapsto x + th$ pour obtenir la relation

$$(g \circ f)'(x, h) = g'(f(x), f'(x, h)).$$

Question 4 Tout d'abord, si la limite

$$\lim_{(t,k) \rightarrow (0,h)} \frac{f(x + tk) - f(x)}{t}$$

existe, elle est égale à la limite obtenue en fixant $k = h$

$$\lim_{t \rightarrow 0} \frac{f(x + th) - f(x)}{t}$$

qui est par définition $f'(x, h)$.

Supposons que cette limite existe et montrons que f a une dérivée directionnelle au sens de Hadamard. Soit γ un chemin satisfaisant les hypothèses de cette

définition. La fonction $f \circ \gamma$ est dérivable en 0 si et seulement si le taux d'accroissement associé converge en 0. Or, ce taux d'accroissement peut s'écrire sous la forme

$$\frac{f(\gamma(t)) - f(\gamma(0))}{t} = \frac{f\left(x + t \frac{\gamma(t) - \gamma(0)}{t}\right) - f(x)}{t}.$$

Le chemin γ étant dérivable en 0,

$$k(t) := \frac{\gamma(t) - \gamma(0)}{t} \rightarrow \gamma'(0) \text{ quand } t \rightarrow 0$$

donc par hypothèse, le taux d'accroissement de $f \circ \gamma$ a une limite en 0.

Réciproquement, suppose que f soit directionnellement dérivable au sens de Hadamard en 0. Pour montrer que la limite

$$\lim_{(t,k) \rightarrow (0,h)} \frac{f(x + tk) - f(x)}{t}$$

existe, il nous suffit de montrer que pour toute suite t_i de valeurs non nulles tendant vers 0 et toute suite de vecteurs k_i convergeant vers h , la limite

$$\lim_{i \rightarrow +\infty} \frac{f(x + t_i k_i) - f(x)}{t_i}$$

existe. On peut imposer la restriction que la suite $|t_i|$ soit strictement décroissante et le résultat reste valable.

Pour tout $t \in \mathbb{R}^*$ notons $j(t)$ le plus petit parmi les entiers j satisfaisant

$$|t - t_j| = \min_{i \in \mathbb{N}} |t - t_i|,$$

puis définissons $\gamma(t)$ par $\gamma(0) = x$ et si $t \neq 0$,

$$\gamma(t) = x + t k_{j(t)}.$$

S'il est défini sur un intervalle $] -\varepsilon, \varepsilon[$ assez petit, γ satisfait les hypothèses de la dérivabilité directionnelle. Le point critique à vérifier est que γ est dérivable en 0. Mais par construction

$$\frac{\gamma(t) - \gamma(0)}{t} = k_{j(t)}$$

et $j(t)$ tend vers $+\infty$ quand t tend vers 0; par conséquent la limite existe et

$$\lim_{t \rightarrow 0} \frac{\gamma(t) - \gamma(0)}{t} = h.$$

Par construction

$$\frac{f(\gamma(t_i)) - f(\gamma(0))}{t_i} = \frac{f(x + t_i k_i) - f(x)}{t_i}.$$

Comme la fonction est dérivable directionnellement au sens de Hadamard,

$$\lim_{i \rightarrow +\infty} \frac{f(x + t_i k_i) - f(x)}{t_i}$$

existe.

Question 5 Si f est différentiable au sens de Fréchet, notons ε la fonction définie dans un voisinage de 0, continue et nulle en 0, telle que

$$f(x+h) = f(x) + df(x) \cdot h + \varepsilon(h)\|h\|.$$

On a alors pour tout $t \in \mathbb{R}$ non nul et tout vecteur $k \in \mathbb{R}^n$ suffisamment petits, en posant $h = tk$,

$$\begin{aligned} \frac{f(x+tk) - f(x)}{t} &= \frac{1}{t} df(x) \cdot tk + \frac{1}{t} \varepsilon(tk)\|tk\| \\ &= df(x) \cdot k + \varepsilon(tk) \frac{|t|}{t} \|k\|. \end{aligned}$$

Le terme $df(x) \cdot k$ tend vers $df(x) \cdot h$ quand $k \rightarrow h$ et le second terme du membre de droite tend vers 0 quand t et k tendent vers 0, donc

$$\lim_{(t,k) \rightarrow (0,h)} \frac{f(x+tk) - f(x)}{t} = df(x) \cdot h.$$

Par conséquent la fonction f est directionnellement dérivable au sens de Hadamard. Le membre de droite, égal à $f'(x, h)$, est linéaire en h ; la fonction f est donc différentiable au sens de Hadamard.

Réciproquement, supposons que f est différentiable au sens de Hadamard. Pour montrer que f est différentiable au sens de Fréchet, de différentielle $f'(x, h)$, montrons que

$$\frac{\|f(x+h) - f(x) - f'(x, h)\|}{\|h\|} \rightarrow 0 \text{ quand } h \rightarrow 0,$$

ou de façon équivalente, que

$$\frac{f\left(x + \|h\| \frac{h}{\|h\|}\right) - f(x)}{\|h\|} - f'\left(x, \frac{h}{\|h\|}\right) \rightarrow 0 \text{ quand } h \rightarrow 0.$$

Il nous suffit de montrer que pour toute suite $t_i > 0$ telle que $t_i \rightarrow 0$ quand $i \rightarrow +\infty$ et $k_i \in \mathbb{R}^n$ telle que $\|k_i\| = 1$,

$$\frac{f(x + t_i k_i) - f(x)}{t_i} - f'(x, k_i) \rightarrow 0 \text{ quand } i \rightarrow +\infty.$$

Imaginons au contraire que cette expression ne tende pas vers 0. Alors on pourrait trouver un $\varepsilon > 0$ et une sous-suite de (t_i, k_i) , notée de (t'_i, k'_i) , telle que pour tout i ,

$$\left\| \frac{f(x + t'_i k'_i) - f(x)}{t'_i} - f'(x, k'_i) \right\| \geq \varepsilon.$$

Mais la suite des k'_i est de norme égale à 1; la sphère fermée de centre 1 étant compacte, il existe des sous-suites t''_i et k''_i de t'_i et k'_i et un $h \in \mathbb{R}^n$ tels que

$\|h\| = 1$ et $k_i'' \rightarrow h$. Par hypothèse de dérivabilité au sens de Hadamard, on aurait

$$\frac{f(x + t_i'' k_i'') - f(x)}{t_i''} \rightarrow f'(x, h) \quad \text{quand } i \rightarrow +\infty$$

ce qui contredit l'inégalité ci-dessus et prouve la contradiction. Par conséquent, f est bien différentiable au sens de Fréchet.

Inégalité de la valeur moyenne

Soit $F : [a, b] \rightarrow \mathbb{R}^m$ une primitive de f . Par définition de l'intégrale de Newton (p. 37),

$$\langle f \rangle = \frac{1}{b-a} \int_a^b f(x) dx = \frac{F(b) - F(a)}{b-a}.$$

Or si $\|F'\| = \|f\|$ est borné sur $[a, b]$, par le théorème des accroissements finis (p. 26),

$$\|F(b) - F(a)\| \leq \sup_{x \in [a, b]} \|f(x)\| \times (b-a),$$

et donc

$$\|\langle f \rangle\| \leq \sup_{x \in [a, b]} \|f(x)\|.$$

Il va de soi que cette inégalité reste vérifiée si $\|f\|$ est non-bornée, c'est-à-dire si $\sup_{x \in [a, b]} \|f(x)\| = +\infty$.

Egalité des accroissements finis ?

La dérivée de f est donnée par $f'(t) = (-\sin t, \cos t)$; en particulier pour tout $t \in [0, 2\pi]$, $\|f'(t)\| = 1$. Or $f(2\pi) - f(0) = 0$, donc il est impossible de trouver un t tel que $f(2\pi) - f(0) = f'(t) \times 2\pi$.

Convexité

Question 0 Le développement limité à l'ordre 2 de f en x (p. 35) 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

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

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.

Références

- Shapiro, A. 1990. “On Concepts of Directional Differentiability.” *Journal of Optimization Theory and Applications* 66 (3): 477–87.
- Tao, Terence. 2013. “Matrix Identities as Derivatives of Determinant Identities.” <https://terrytao.wordpress.com/2013/01/13/matrix-identities-as-derivatives-of-determinant-identities/>.