Calcul Différentiel I

Table des matières

Narratif & Notes & TODOsss

Différentielle & Dérivée Directionnelle

La progression choisie est la suivante:

- la différentielle d'une fonction en un point est introduite directement, par analogie avec le concept et les propriétés de la dérivée, une fois présentés sous la bonne forme (meilleure approximation linéaire de la variation, forme avec o plutôt que taux d'accroissement).
- on exploite un peu cette définition, on finit de faire le lien avec la dérivée, on donne les règles de combi linéaires, du produit, de la différentation en chaîne.
- sous hypothèse de différentiabilité, on donne le liens avec la dérivée partielle et les dérivée directionnnelles.
- après coup, on examine la tentation (légitime) que l'on pourrait avoir de définir la différentielle en passant par les dérivée partielles: cette approche si elle était couronnée de succès, permettrait de définir la différentielle en se ramenant à ce que l'on connaît déjà, à savoir la notion de dérivée. Et on se rend compte assez facilement que:
  - l'existence des dérivées partielles ne suffit pas à assurer l'existence de la différentielle: un exemple très simple permet de montrer que cela n'assure même pas la continuité de la fonction au point d'intérêt. Plus grave si l'on veut: la règle de dérivation en chaîne ne marche pas non plus; en particulier, on ne peut pas calculer la dérivée partielle d'une fonction composée par ce biais. Note: suppose que l'on ait dérivé la chain rule très rapidement après la définition de la différentielle et c'est légitime: c'est un grand succès du concept.
  - examiner le contre-exemple standard (1 valeur sur les axes, une autre dans le reste du domaine), diagnostiquer ce qui ne va pas (à savoir, on est aveugle au comportement de la fonction en dehors de directions privilégiées), propose une solution en travaillant sur la dérivée directionnelle. Montrer par un nouveau contre-exemple, moins évident, que ça ne va toujours pas (ni continuité ni "chain rule"). L'exemple en question travaille toujours avec deux valeurs distinctes, mais sur une parabole.
  - le nouvelle exemple pour le coup met sur la piste d'une "bonne" solution alternative, la dérivée directionnelle au sens d'Hadamard. On peut la définir au moyens des chemins, simplifier sa caractérisation. Au final, elle vérifie bien la règle de dérivation en chaîne par exemple, plus ou moins par construction, mais cela n'est pas surprenant car elle est équivalent à la notion de différentielle ! A ce stade, pas évident que la démarche adoptée soit plus simple, on peut se convaincre que le concept de différentielle est finalement pas si mal que ça … d'autant plus qu'en dimension infinie, les deux notions divergent à nouveau et la différentielle de Fréchet regagne des points.

Une partie de ça à faire dans le cours, une partie en exo, quelle frontière je ne sais pas encore exactement.

- en parralèle, on montre que tout de même, on a le droit de travailler sur les dérivée partielles si l'on sait établir que le résultat est continu, car cela garantit l'existence de la différentielle (et sa continuité).

Vecteurs / Matrices / Tenseurs

Sujet assez compliqué. Trois motivations sur ce sujet:

- Le "tout-matrice" est assez ridicule quand on y pense; l'idée qu'il faille promouvoir des vecteurs de $\mathbb{R}^n$ en matrice pour faire des calculs complique souvent les choses par rapport aux conventions tensorielles (où un vecteur est un tableau de dimension 1). C'est aussi assez incohérent avec les convention de NumPy qui pour le coup sont tensorielles par nature (contrairement à Matlab).
  Mais voilà, c'est la vision enseignée en prépa, difficile de tout déconstruire, d'autant que la démarche tensorielle vient avec ses propres problèmes de notation, conventions non partagées, etc.
  Donc on a vocation à rester compatible avec ce tout-matriciel; et à l'étendre mais de façon compatible quand nécessaire. Ainsi, "$\cdot$" interprété comme "application d'une fonction linéaire", même quand la-dite fonction linéaire est à valeurs fonctionnelle (comme dans les diff d'ordre supérieur)
- Il y a des problèmes qui supposent naturellement de considérer des fonctions avec des arguments matriciels. Par exemple, on comprend assez bien qu'on peut avoir besoin de différencier $\det A$ ou A^{-1}. Même si le problème final n'a que des paramètres vectoriels, on a envie de faire des "chain rules" avec des arguments matriciels.
- Exemple pas trivial mais typique: calculer $d^2 f \circ g$. A l'ordre 1 on a $df \circ g(x) = df(g(x)) \cdot df(x)$, ce qui est (interprétable comme) un produit de matrices.

Positions aujourd'hui:

- rester dans un premier temps compatible avec le conventions du tout-matriciel, se contenter de noter l'écart avec les conventions NumPy, conserver une définition de $\cdot$ qui soit plus générale.
- minimiser les présentations du tensoriel: on peut se contenter de montrer que $d^2 f$ est représentable comme un tableau à trois dimensions et de faire les calculs avec les indices pour évaluer $d^f(x) \cdot h_1 \cdot h_2$ par exemple; le cas différentielle d'ordre k n'est guère plus complexe.
- regarder s'il y a des exemples éclairants à faire en TD sur de la différentielle à argument matriciels et "bootstrapper" la définition de la différentielle à ce moment-là, en "mettant à plat" la matrice par exemple ? Ou exploiter la définition d'Hadamard pour éviter d'avoir à faire ça ?

Normes

Ne rien mettre dans ce chapitre proprement dit, mais lister ce dont on a besoin très concrètement pour inclure ces éléments dans le chapitre de topologie.

J'ai assez envie de noter par défaut $|x|$ les normes dans $\mathbb{R}^n$ et $\|L\|$ la norme d'opérateur et d'annoter ces normes par des symboles (comme $|x|_2$, $\|A\|_{22}$) dans les contextes ou il faudrait être plus précis.

Dans ce chapitre j'imagine que l'on peut (presque ?) toujours se limiter aux normes euclidiennes et à la norme d'opérateur induite, sauf peut-être si l'on en vient à montrer des choses comme le caractére intrinsèque de la définition de différentielle ? Non, même là ça va marcher.

Donc concrêtement, définition de ces deux normes, pptés habituelles (notamment $|Lx| \leq \|L\||x|$). Le coup de la norme d'opérateur associé à la représentation matricielle (via SVD), utile ou pas ? Si oui – et on peut en douter – alors il faut aussi parler de matrices dans le chapitre sur la topo. Ouch, non, éviter. En fait, il faudra sans peut-être "retarder" les rappels sur les opérateurs linéaires à ce chapitre, car c'est un gros focus du chapitre (idée d'approximation linéaire est centrale ici, avant ça serait abstrait).

Auquel cas on parle de norme en topo, on montrer les exemple classiques en dim finie et on parle d'équivalence des normes, mais on attend ce chapitre pour parler d'opérateurs et de norme. Donc un volet à rajouter ici ?

Preambule

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

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

Notations

Ensembles et Fonctions

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

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 $f : A \to B$ et $g : B \to C$, la composée des functions f et de g, notée $g \circ f$, appartient à $A \to C$ 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 functions 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 $A \to B$. On pourra alors utiliser la règle

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

ou 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

$$\left[\begin{array}{ccc} 1 & 2 & 3 \\ 4 & 5 & 6 \end{array} \right] \in \mathbb{R}^{2 \times 3}$$

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 $\mathbb{R}^{mn}$. Par exemple:

$$\left[\begin{array}{ccc} 1 & 2 & 3 \\ 4 & 5 & 6 \end{array}\right] \in \mathbb{R}^{2\times 3} \;\mapsto\; (1,2,3,4,5,6) \in \mathbb{R}^6$$

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 $A : \mathbb{R}^n \to \mathbb{R}^m$ la matrice

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

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 $A : \mathbb{R}^n \to \mathbb{R}^m$ 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'application 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 symbol $\cdot$. Le même symbole sera utilisé pour désigner le produit entre deux matrices (on évitera dans la mesure du possible de désigner le produit de deux matrices par simple juxtaposition des symboles).

Avec NumPy, la méthode `dot` des tableaux permet de réaliser cette opération:

```
>>> 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]])
```

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

$$\left[\begin{array}{c} x_1 \\ \vdots \\ x_n \end{array}\right] \in \mathbb{R}^{n \times 1}.$$

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 habituel. L'intérêt de cette opération: 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 $y = Ax \in \mathbb{R}^m$ 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

Objectif

Présenté volontairement dans le cadre le plus étroit possible qui satisfasse nos besoins (notamment, comparaison par rapport $\|h\|^k$) suffit, ce qui évite un grand nombre de subtilités. Pas jugé d'un grand intérêt en tant que tel, nous ne développons absolument pas le "calcul des o"; il s'agit juste d'avoir une notation pratique pour noter des résultats, dans le cadre bien précis du calcul différentiel et des propriétés des restes dans les développements limités. Toutes les démonstrations commencent par la traduction des o en fonctions; on est donc presque dans la situation ou l'on pourrait se passer de la notation; on aurait en contrepartie des résultats un peu plus lourd à énoncer, les conséquences seraient limitées à ça.

TODO: remarque sur rôle du $o(1)$ et comment on pourrait tout ramener à ça … retenir au moins que $o(\|h\|) = o(1)\|h\|$? La notation $o(1)$ est pratique pour désigner ε directement, sans avoir à rappeler les hypothèses en détail.

Petit o de Landau

La notation $o(\|h\|^k)$, où $h \in \mathbb{R}^n$ et $k \in \mathbb{N}$, désigne toute 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

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

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 V de 0 et vérifie:

$$\lim_{h\to 0} \frac{\phi(h)}{\|h\|} = 0.$$

La fonction ε est alors définie de façon unique sur V par la relation

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

Continuité

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

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

(ce qui correspond au cas ou $k = 0$ puisque $\|h\|^0 = 1$) signifie donc que f définie dans un voisinage de x et que

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

autrement dit que x appartient à l'intérieur du domaine de f et que f y est continue.

Différentielle

Dérivée

Soit $f : U \subset \mathbb{R} \to \mathbb{R}^m$ où U est ouvert. 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:

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

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

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

Remarque

Cette définition de la dérivée nécessite la formation d'un taux d'accroissement et par conséquent que h soit scalaire puisque l'on divise par h; il ne peut être utilisé que si la fonction f n'ait qu'un argument scalaire. En revanche, la fonction peut être à valeurs scalaires ou vectorielles sans qu'il soit nécessaire de changer cette définition. Plus précisement, une fonction vectorielle $f = (f_1, \cdots, f_m)$ sera 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.

Remarque

La dérivabilité peut être définie de façon équivalente en passant par la notion de développement limité à l'ordre 1.

Développement limité au premier ordre

Soit $f : U \subset \mathbb{R} \to \mathbb{R}^m$ où U est ouvert. 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 $h \neq 0$,

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

on a bien $\lim_{h \to 0} \varepsilon(h) = 0$. 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 une fonction ε conforme à la notation de Landau,

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

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

■

Fonctions linéaires d'une variable scalaire

Ce développement limité de f en x à l'ordre 1 fournit de façon explicite une approximation linéaire de la variation $\Delta f(x,h)$ de f en x:

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

Cette remarque n'est pas anodine car toutes les applications linéaires de $\mathbb{R}$ dans $\mathbb{R}^m$ sont de la forme $h \mapsto \ell h$ pour un certain vecteur ℓ. En effet, L étant linéaire, 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. Par conséquent, on peut 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 fonction à n variables.

Différentielle de Fréchet

Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ où U est ouvert. La fonction f est *différentiable* en $x \in U$ s'il existe une application linéaire $L : \mathbb{R}^n \to \mathbb{R}^m$ 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* si elle est différentiable en tout point de U.

Remarque

Si l'on considère à nouveau $\Delta f(x,h)$, la variation de f en x, associée à la variation h de l'argument

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

on réalise que la différentielle de f en x, quand elle existe, constitue une approximation de cette variation qui est linéaire en h

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

et d'une certaine façon la meilleure puisque cette relation la définit de façon unique.

Note

On pourra parler de fonction f différentiable *sur U* si le domaine de définition de la fonction n'est pas évident dans le contexte (cas particulier des expressions, à suivre).

Différentiation composante par composante

Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ où U est ouvert. La fonction $f = (f_1, \cdots, f_m)$ 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 $h \in \mathbb{R}^n$

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

Preuve

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 différentables 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

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

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

$\blacksquare$

Domaine de définition non ouvert

La définition de 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 considéré du domaine de définition, on peut 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 $\mathbb{R}$.

Résumons les liens entre dérivée et différentielle:

Différentielle et Dérivée

Soit $f : U \subset \mathbb{R} \to \mathbb{R}^m$ où U est ouvert et soit $x \in U$. La fonction f est différentiable en a si et seulement si elle est dérivable en x. Dérivée et différentielle de f en a se déduisent alors l'une de l'autre par les relations

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

Démonstration

Une conséquence de la caractérisation de la dérivabilité des fonctions par l'existence de développement limité au premier ordre et de la caractérisation des fonctions linéaires d'une variable scalaire.

■

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 ambiguité mais peut parfois être lourde à manipuler. Dans le calcul des dérivées, nos avons pris l'habitude, pour affirmer que la dérivée de la fonction $x \mapsto x^2$ et 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:

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

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'ambiguité [1]. 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

$$dx^2 = 2xdx.$$

Note

Même si la notation de la différentielle en x donne un indice sur l'étape suivante, il faut probablement retarder l'apparition de la notion d'application différentielle et construire une familiarité avec la notion de différentielle en a avant de passer à l'étape d'après. La notion d'application différentielle ne devient nécessaire que pour parler de fonction continûment différentiable et de différentielle d'ordre supérieur.

Sous les hypothèse 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 $f : U \subset \mathbb{R}^p \to \mathbb{R}^n$ et $g : V \subset \mathbb{R}^n \to \mathbb{R}^m$ 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

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

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).

1. 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'ambiguité si nécessaire, avec $d(f)(x^2)$ dans le premier cas et $d(f(x^2))$ dans le second. Par défaut, nous supposerons dans la suite que $df(x^2)$ désigne la notation "stricte" $d(f)(x^2)$.

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

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

Preuve

L'objectif de la preuve est de montrer que

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

La fonction g étant différentiable en $f(x)$, il existe une fonction ε_1 définie sur un voisinage de 0 et à valeurs dans $\mathbb{R}^m$ telle que $\varepsilon_1(k) \to \varepsilon_1(0) = 0$ quand k tend vers 0 et

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

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 $\varepsilon_2(h) \to \varepsilon_2(0) = 0$ quand $h \to 0$. On a donc

$$\begin{aligned} g(f(x+h)) - g(f(x)) = {} & dg(f(x)) \cdot (f(x+h) - f(x)) \\ & + \varepsilon_2(h)\|f(x+h) - f(x)\|. \end{aligned}$$

Comme f est différentiable en x, il existe une fonction ε_3 définie sur un voisinage de 0 et à valeurs dans $\mathbb{R}^n$ telle que $\varepsilon_3(h) \to \varepsilon_3(0) = 0$ quand h tend vers 0 et

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

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,

$$\varepsilon(h) = \varepsilon_2(h)\|df(x) \cdot (h/\|h\|) + \varepsilon_3(h)\| + dg(f(x)) \cdot \varepsilon_3(h).$$

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

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

le résultat est donc acquis.

TODO; décomposer en règle de la somme et du facteur constant ? Omettre facteur constant et le voir comme un corollaire?

Linéarité de la différentielle

La combinaison linéaire $(x, y) \in \mathbb{R}^2 \mapsto \lambda x + \mu y \in \mathbb{R}$ est différentiable en tout point pour tous scalaires λ et μ et

$$d(\lambda x + \mu y) = \lambda dx + \mu dy.$$

Remarque

Si l'on note c l'application de combinaison linéaire, ce résultat signifie que pour tout couple (h_x, h_y) de réels, on a

$$d\mathrm{c}(x, y) \cdot (h_x, h_y) = \lambda h_x + \mu h_y.$$

Démonstration

TODO

■

Règle du produit

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

$$dxy = xdy + ydx$$

Remarque

Si l'on note p l'application produit, ce résultat signifie que pour tout couple (h_x, h_y) de réels, on a

$$d\mathrm{p}(x, y) \cdot (h_x, h_y) = xh_y + yh_x.$$

Démonstration

TODO: (cas matriciel pour le produit ? A un moment ?). En exercice ?

■

Jacobienne, dérivées partielles et directionnelles

Objectifs

TODO: à l'oral, insister sur différentielle comme point de départ et le reste (dérivées partielles, directionnelle, etc) s'ensuivent. Montrer que la démarche inverse ne marche pas (bien que la jacobienne puisse être formellement définie, la chain rule ne marche pas, donc on ne peut pas les multiplier)

Matrice Jacobienne

Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ où U est ouvert et soit 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 $df(x)$ de f en x qui est une application linéaire de $\mathbb{R}^m$ dans $\mathbb{R}^n$

TODO: à quelle moment est-ce que j'indique que

$$[df(x) \cdot e_j]_i = df_i(x) \cdot e_j?$$

Dérivée Partielle

Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ où U est un ouvert et soit $x \in U$. Lorsque la i-ème fonction partielle de f en x

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

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,

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

quand le second membre existe.

Dérivées partielles et différentielle

Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ où U est un ouvert et soit 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 $h \in \mathbb{R}^n$

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

Preuve

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 $t \neq 0$; substituer $h := te_j$ dans cette relation fournit

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

En exploitant la linéarité de la différentielle, on obtient donc

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

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

$$df(x)\cdot e_j = \lim_{t\to 0}\frac{f_i(x+te_j)-f_i(x)}{t} =: \partial_j f(x)$$

La différentielle pouvant être calculée composante par composante, on en déduit que

$$\partial_i f(x) df_i(x)\cdot e_j = \lim_{t\to 0}\frac{f_i(x+te_j)-f_i(x)}{t}.$$

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

$$h = (h_1, \dots, h_m) = h_1 e_1 + \dots + h_m e_m$$

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

$$df(x)\cdot h = df(x)\cdot\left(\sum_i h_i e_i\right) = \sum_i (df(x)\cdot e_i) h_i = \sum_i \partial_i f(x) h_i.$$

■

La dérivée partielle n'est qu'un cas particulier du concept de dérivée directionnelle, limitée aux directions de la base canonique de $\mathbb{R}^n$.

Dérivée directionnelle

Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ où U est un ouvert et soit x un point de U. On appelle *dérivée directionnelle* de f en x dans la direction $h \in \mathbb{R}^n$ la valeur

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

quand elle existe.

Dérivée partielle et directionnelle

La fonction f admet une dérivée directionnelle en x dans la direction e_i si et seulement si sa i-ème dérivée partielle existe; on a alors

$$f'(x,h) = \partial_i f(x).$$

Preuve

Direct.

■

Matrice Jacobienne

$$Df(x) = [df(x)] = [\partial_j f_i(x)]_{ij} = \left[\begin{array}{c} [df_1(x)] \\ \vdots \\ [df_m(x)] \end{array}\right]$$

TODO.

TODO: dumb down “différentielles” partielles en “dérivées partielles”. Et c’est une “découverte”, on part de la représentation sous forme matricielle des différentielles, les dérivée partielles ne sont pas considérées comme un point de départ.

TODO: variables nommées, impact notation.

TODO. Eventuellement un mot sur le concept d’application partielle ? Exploiter la chain rule plutôt que la construction élémentaire de l’introduction ?

TODO

Fcts C^1 et réciproque partielle... autre section ? ICI ?

Inégalité des accroissement finis

Différentielle et Intégrale

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

Toutefois, dans cette situation, si f est différentiable sur tout le segment $[a, b]$, il est possible de relier $f(a + h)$ à $f(a)$.

Théorème

Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ où U est ouvert, soit $a \in U$ et $h \in \mathbb{R}^n$ tel que le segment $[a, a + h]$

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

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

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

Preuve

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

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

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

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

Par le théorème fondamental du calcul, on a donc

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

Inégalité des accroissements finis (fonction d'une variable réelle)

Soit $f : [a, b] \to \mathbb{R}^m$ où $a \in \mathbb{R}$, $b \in \mathbb{R}$, $a \leq b$ et $m \in \mathbb{N}$. Si f est dérivable sur $[a, b]$ et M est un majorant de $\|f'\|$, c'est-à-dire si

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

Alors

$$\|f(b) - f(a)\| \leq M(b - a)$$

Preuve

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

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

Elle est donc également intégrable au sens de Henstock-Kurzweil [**TODO: référence**]; 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 Riemann: pour tout $\varepsilon > 0$, il existe une subdivision pointée $\mathcal{D}$ de l'intervalle $[a, b]$ telle que

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

En exploitant l'inégalité triangulaire, on obtient donc

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

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

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

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

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

Finalement, comme $a = x_0 \leq x_1 \leq \dots x_p = b$,

$$\sum_{i=0}^{p-1} |x_{i+1} - x_i| = \sum_{i=0}^{p-1} (x_{i+1} - x_i) = x_p - x_0 = b - a$$

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

$\blacksquare$

Inégalité des accroissements finis

Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ où U est ouvert, 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

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

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

Preuve

Steps: define $\phi : t \in [0,1] \mapsto f(a+th)$. Deduce $\phi'(t) = df(a+th)(h)$ (chaine + rule + derivative/differential link). Conclude with above

■

Points critiques

Différentielles d'ordre supérieur

Do's and don't

Ne pas expliciter la correspondance avec les applis n-linéaires en général (l'isomorphisme de trop). Une notation serait probablement la bienvenue, mais la collection des $\cdot h \cdot h \dots$ en $(h, h, \dots)$ peut être ambigu (pourrait être lue comme la décomposition d'un vecteur ...). Trouver une solution ici. OK, on se contente de multiplier les dots, avec convention association à gauche ("greedy")

Ce qui importe:

- comprendre comment "passer à l'échelle" de la diff à la diff d'ordre 2, qu'il n'y a "rien de nouveau" si l'on a déja compris comment différencier une fonction à valeurs matricielle (/tensorielle).
- donc dvlper en préambule le calcul diff appliqué aux fcts à valeurs fonctionnelles/tensorielles. Ne pas faire l'équivalent pour les arguments, cela n'est pas nécessaire pour traiter du calcul différentiel d'ordre supérieur.
- comprendre comment calculer $d^2 f(x) \cdot k \cdot h$ quand on sait qu'il existe sans "monter dans les étages" (trick: différencier $df(x) \cdot h$).
- comprendre quel terme représente $d^2 f(x) \cdot k \cdot h$ en pratique, quelle approximation ce terme fait. (Nota: au passage c'est crucial pour établir la symmétrie !).
- représentation tensorielle, dérivées partielles. Application au Hessien.
- Sommes de Taylor (avec o, avec reste intégral)

Nota: peut-être opportun de minimiser le coté diff par les valeurs matricielles. Idées serait de caractériser df en vérifiant la différentiabilité de $x \mapsto df(x) \cdot h$ pour tout h: on ne "monte" pas en rang et on peut définit $d^2 f(x) \cdot k \cdot h := d(x \mapsto df(x) \cdot h)(x) \cdot k$

Différentielles d’ordre supérieur

Note

La démarche pour présenter les différentielles d’ordre supérieur a été simplifié, mais le narratif peut profiter des “échecs” qui mène à notre solution finale:

1. On a envie de définir $d^2 f(x)$ comme $d(x \mapsto df(x))(x)$. C’est *exactement* la même démarche que la dérivée, et c’est une démarche légitime que l’on adoptera pour le cadre de la dimension infinie. Seul “problème”, l’objet $df(x)$ appartient aux applis linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^m$ et à ce stade on ne sait différencier que des fonctions à valeurs dans $\mathbb{R}^p$.
2. On “patche” la démarche précédente: ok, $df(x)$ est linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^m$, mais c’est isomorphe (via les matrices, plus la “mise à plat”) à $\mathbb{R}^p$ pour $p = mn$. Si on note π cette correspondance, on peut étudier la diff de $\pi \circ df$ et quand ça existe, le seul pb est que l’objet associé produit des valeurs dans $\mathbb{R}^p$ au lieu de trucs dans $\mathbb{R}^n \stackrel{\ell}{\to} \mathbb{R}^m$, mais c’est pas grave, on peut inverser la transformation, ce qui donne comme définition
$$d^2 f(x) = \pi^{-1} \circ d(\pi \circ df)(x).$$
c’est-à-dire
$$d^2 f(x) \cdot h \cdot k = (\pi^{-1}(d(\pi \circ df)(x) \cdot h)) \cdot k$$
(attention ici, “$\cdot$” ou “$\circ$” deviennent dangereux à utiliser sans parenthèse, on a basculé dans du higher-order avec π; et la convention que je pensais utiliser en remplaçant $\circ$ par $\cdot$ quand l’application est linéaire déconne avec π parce qu’il y a des applications non-linéaire “plus bas”; le cadre ou “$\cdot$” fait le job sans ambiguité serait à restreindre/préciser …). Bon, voilà pourquoi je crois que même si c’est tentant sur le principe, il ne faut pas présenter les choses comme ça au final. Mais ça peut faire l’objet d’exercices intéressants.
3. La version final, hyper simple: on se refuse à différencier un objet fonctionnel, on l’évalue sur une direction / variation de l’argument et là on s’est ramené au cadre usuel, donc on requière la diff et on constate que le résultat est linéaire par rapport à la première variation choisie, et on en déduit l’“anatomie” de la différentielle d’ordre 2 (c’est donc moins une construction qu’une découverte …). Au passage, par linéarité, on peut se convaincre facilement que notre définition de la différentielle d’ordre 2 revient à vérifier que chaque dérivée partielle (d’une fonction différentiable) est différentiable. Donc on vérifie l’existence avec la différentielle, mais on peut utiliser les dérivées partielles pour les calculs intermédiaires.

Note

Notre définition simplifie la vie en dimension finie en se ramenant directement à chaque étape de la façon la plus simple au cadre de la différentielle de fonctions de $\mathbb{R}^n$ dans $\mathbb{R}^m$. Mais elle n'est probablement pas adapté au cadre de la dimension finie; l'adaptation la plus simple consisterait à écrire l'expression qui fait que $df(x) \cdot h$ existe en terme de limite par rapport à un terme k, mais à requérir en plus que cette limite existe uniformément par rapport à l'argument h tant que h reste borné (voir par exemple ici).

A ce stade, le cadre abstrait classique devient probablement préférable, car simplificateur, mais

— un contre-exemple qui montre que notre définition ne "marche pas" en dimension infinie (absence d'équivalence avec la classique) serait intéressant
— une note / un exercice sur cette définition alternative au cadre abstrait, plus proche de la démarche que nous avons choisi pour la dimension finie pourrait être intéressant

Différentielle d'ordre 2

Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ 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 $d^2 f(x)$, est définie comme l'application linéaire telle que pour tout h dans $\mathbb{R}^n$,

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

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

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

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 $d^2 f(x)$ est linéaire et la notation "$\cdot$" 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 $\mathbb{R}^n \to \mathbb{R}^m$, donc la fonction $d^2 f(x)$ associe linéairement à un vecteur de $\mathbb{R}^n$ une application linéaire de $\mathbb{R}^n \to \mathbb{R}^m$. Autrement dit,

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

ce qui se décline successivement en

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

— Pour alléger ces notations, on pourra considérer que dans les notations d'espace fonctionnels, le symbole "$\to$" associe à droite, par exemple:

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

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

La convention associée: lors de l'application d'une fonction linéaire, le symbole "$\cdot$" associe à gauche, par exemple:

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

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

Variation de la différentielle

Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ une fonction deux fois différentiable en un point x de U. On a

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

TODO: en pratique, on combinera le résultat ci-dessus avec la symmétrie de la différentielle d'ordre 2 pour mémoriser le résultat. Mais ce résultat est lui-même utile dans la preuve de la symmétrie. Comment présenter les résultat au final ? Se débrouiller pour minorer l'impact de la forme "temporaire" dans l'exposé oral c'est clair, mais dans le poly comment faire pour casser la boucle ? Du coup, ce résultat serait un lemme, et le "vrai" théorème simplifié suivra. OK.

Remarque

Dans l'équation ci-dessus, le "$o(\|k\|)$" est inséré dans une équation entre applications linéaires de $\mathbb{R}^n \to \mathbb{R}^m$. Il doit donc être interprété comme

$$o(\|k\|) = E(k)\|k\| \;\text{ où }\; E(k) \in \mathbb{R}^n \stackrel{\ell}{\to} \mathbb{R}^m, \; \lim_{h \to 0} E(k) = E(0) = 0.$$

TODO?. "Sortir" les lemmes du théorème ? Voire les notation Δf ? Ce sont des résultats majeurs ? (ça éclaire des choses sur ce qu'est $d^2 f$ et comment la calculer alors pourquoi pas … ça pourrait aussi nous éviter des lemmes "nestés" quoi qu'il en soit dans la preuve du théorème. A la limite, le résultat sur la symmétrie de Δf peut rester dedans, c'est ça le coeur de la preuve. Et je sors l'autre sur l'approximation de $d^2 f$ par $\Delta^2 f$.)

Preuve

Par définition de la différentielle d'ordre 2 en x, pour tout vecteur h de $\mathbb{R}^n$ fixé, on a, pour tout vecteur k de $\mathbb{R}^n$,

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

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

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

Pour tout vecteur k non nul, on a

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

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

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

soit

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

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

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

ce qui prouve le résultat cherché.

TODO: définir $\Delta^2 f$; définir Δf au préalable qqpart, et rappeler ici.

Le théorème qui suit montre que $d^2 f(x) \cdot h \cdot k$ fournit une approximation de $\Delta^2 f(x, h, k)$ quand h et k sont petits.

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

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

Preuve

TODO: h et k assez petits pour que les expressions soient toutes définies.

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 le théorème des accroissements finis: g est différentiable sur le segment $[0, h]$ et

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

Comme

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

par le théorème controllant la variation de la différentielle, 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 accroissement finis fournit

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

Symmétrie de la différentielle d'ordre 2

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

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

Preuve

Notons $\Delta f(x,h) = f(x+h) - f(x)$ la variation de f en x associée pour une variation h de l'argument et $\Delta^2 f(x,h,k)$ la variation de $\Delta f(x,h)$ en x pour une variation k de l'argument:

$$\Delta^2 f(x,h,k) = (f(x+k+h) - f(x+k)) - (f(x+h) - f(x)).$$

On peut remarquer que la variation d'ordre 2 de f en x est symmétrique: lorsque $\Delta^2 f(x,h,k)$ est définie, $\Delta^2 f(x,k,h)$ également et

$$\Delta^2 f(x,h,k) = \Delta^2 f(x,k,h).$$

La conclusion s'imposera alors: en effet, si h et k sont des vecteurs de $\mathbb{R}^n$ en exploitant la symmétrie de $\Delta^2 f$, on obtient

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

En substituant th à h et tk à tk dans cette expression, puis en faisant tendre t vers 0, on peut rendre th et tk arbitrairement proches de 0 et donc s'assurer que pour tout $\varepsilon > 0$,

$$\begin{aligned}\|d^2 f(x) \cdot h \cdot k - d^2 f(x) \cdot k \cdot h\| &= \frac{1}{t^2}\|d^2 f(x) \cdot th \cdot tk - d^2 f(x) \cdot tk \cdot th\| \\ &\leq \frac{1}{t^2} 2\varepsilon(\|th\| + \|tk\|)^2 \\ &= 2\varepsilon(\|h\| + \|k\|)^2,\end{aligned}$$

ce qui nous assure que $d^2 f(x) \cdot h \cdot k - d^2 f(x) \cdot k \cdot h$.

Variation de la différentielle

Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ une fonction deux fois différentiable en un point x de U. On a

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

Preuve

Par le lemme sur la variation de la différentielle, 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 symmétrique,

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

et par conséquent

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

qui est l'égalité cherchée.

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

Différentielle d'ordre k

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

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

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 $\mathbb{R}^n$,

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

ou de façon équivalente

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

Remarque

On a

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

Fonctions à valeurs matricielles/tensorielles

Objectif: étendre les constructions du calcul différentielle aux fonctions $f : U \subset \mathbb{R}^p \to \mathbb{R}^{m \times n}$ (après valeurs scalaires et vectorielles, matricielles).

Etape 1: valeurs interprétée indifférement comme une matrice de taille $m \times n$ ou comme une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^m$.

Examples

On peut associer à tout vecteur x non nul de $\mathbb{R}^n$ la projection orthogonale sur x; c'est une application linéaire $P(x)$ qui a tout vecteur y de $\mathbb{R}^n$ associe le vecteur

$$P(x) \cdot y = \left\langle \frac{x}{\|x\|}, y \right\rangle \frac{x}{\|x\|} = \frac{x}{\|x\|} \cdot \left(\frac{x}{\|x\|} \right)^t \cdot y$$

Produit scalaire, exp matrice, etc ?

Définition

TODO: motiver la nature de dF quand F est à valeurs fonctionnelles.

Si $F : U \subset \mathbb{R}^n \to \mathbb{R}^{p \times m}$, la différentielle de F au point $x \in \mathbb{R}^n$ est l'application $dF(x)$ telle que $dF(x) \cdot h$ soit la meilleure approximation, linéaire en h, de $F(x+h) - F(x)$ pour de petites valeurs de h

$$F(x+h) = F(x) + dF(x) \cdot h + o(h)$$

L'application $dF(x)$ est donc une application linéaire de $\mathbb{R}^n$ dans les applications linéaires de $\mathbb{R}^m$ dans $\mathbb{R}^p$:

$$dF(x) : \mathbb{R}^n \stackrel{\ell}{\to} (\mathbb{R}^m \stackrel{\ell}{\to} \mathbb{R}^p)$$

On peut associer à cette application un tenseur de rang 3, **TODO, def etc.**

$$[dF(x) \cdot e_k \cdot e_j]_i$$

TODO $\cdot$ désigne la contraction tensorielle, composition, etc. Généraliser le cas matriciel, montrer les correspondances avec le cadre fonctionnel. Isomorphisme

$$\mathbb{R}^{m \times n \times n} \; \simeq \; \mathbb{R}^m \stackrel{\ell}{\leftarrow} \mathbb{R}^n \stackrel{\ell}{\leftarrow} \mathbb{R}^n$$

TODO: régle du produit:

$H(x) = G(x) \cdot F(x)$,

$$dH(x) \cdot h = (dG(x) \cdot h)F(x) + (dF(x) \cdot h)G(x)$$

Misc.

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

$df : U \subset \mathbb{R}^n \to (\mathbb{R}^n \stackrel{\ell}{\to} \mathbb{R}^m)$

$$d^2 f : U \subset \mathbb{R}^n \to (\mathbb{R}^n \stackrel{\ell}{\to} (\mathbb{R}^n \stackrel{\ell}{\to} \mathbb{R}^m))$$

Tensor stuff

TODO: définition. Il va falloir être malin …

Théorème

Si $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ est deux fois différentiable en x, pour tout $h \in \mathbb{R}^n$ et $k \in \mathbb{R}^n$,

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

Preuve

TODO

Théorème

TODO: approximation concrète de $d^2 f(x) \cdot h \cdot k$.

Exercices

Vecteurs, vecteurs colonnes, vecteurs lignes

Soit $x = (x_1, \cdots, x_n)$ un vecteur de $\mathbb{R}^n$.

1. Le vecteur colonne X associé à x

$$X = \left[\begin{array}{c} x_1 \\ \vdots \\ x_n \end{array} \right] \in \mathbb{R}^{n \times 1}.$$

   représente une application linéaire. Laquelle ?

2. Le vecteur colonne ligne X^t associé à x

$$X^t = \left[\begin{array}{ccc} x_1 & \cdots & x_n \end{array} \right] \in \mathbb{R}^{1 \times n}.$$

représente une application linéaire. Laquelle ?

Réponses

1. Par définition, le vecteur colonne associé à x représente l'application linéaire A de $\mathbb{R}$ dans $\mathbb{R}^n$ telle que pour tout $h \in \mathbb{R}$ et tout $i = 1, \dots, n$,

$$(Ah)_i = \sum_{k=1}^{1} X_{ik} h = x_i h,$$

soit $Ah = h$.

2. Par définition, le vecteur ligne associé à x représente l'application linéaire B de $\mathbb{R}^n$ dans $\mathbb{R}$ telle que pour tout $h = (h_1, \dots, h_n) \in \mathbb{R}^n$

$$Bh = \sum_k x_i h_i,$$

soit $Bh = \langle x, h \rangle$ ou $\langle \cdot, \cdot \rangle$ désigne le produit scalaire dans $\mathbb{R}^n$.

Dérivée sur un intervalle fermé

TODO: deux options: extension globale ou locale, y repenser.

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 g sur $]a - \varepsilon, b + \varepsilon[$ tel que g soit dérivable.

Montrer qu'alors, $f' = g'|_{[a,b]}$.

Dérivation en chaîne

Montrer que la règle de dérivation en chaîne ci-dessous, concernant les fonctions d'une variable, se déduit de la règle générale de différentiation en chaîne.

Soit $f : U \subset \mathbb{R} \to \mathbb{R}$ et $g : V \subset \mathbb{R} \to \mathbb{R}$ 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

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

Réponse

Les fonction f et g sont dérivables donc différentiables (cf. Différentielle et Dérivée). Par application de la [règle de différentiation en chaîne][Règle de dérivation en chaîne], leur composée $g \circ f$ est donc différentiable. C'est une fonction d'une variable, elle est donc dérivable, à nouveau en invoquant le lien entre différentielle et dérivée. Pour ces trois fonctions, on obtient la dérivée en appliquant la différentielle à 1; La règle de différentiation en chaîne fournissant

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

on en déduit

$$\begin{aligned}
(g \circ f)'(x) &= (d(g \circ f)(x)) \cdot 1 \\
&= (dg(f(x)) \cdot df(x)) \cdot 1 \\
&= dg(f(x)) \cdot (df(x) \cdot 1) \\
&= dg(f(x)) \cdot (f'(x)1) \\
&= (dg(f(x)) \cdot 1)f'(x) \\
&= g'(f(x))f'(x)
\end{aligned}$$

Calcul Méca

Faire les calculs menant à $C(q, \dot{q})\dot{q}$ en mécanique lagrangienne ?

Dérivée directionnelle d'Hadamard

Source: (Shapiro 1990)

Rappel. Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ où U est ouvert et $x \in U$. La fonction f est *directionnellement dérivable* si pour tout vecteur $h \in \mathbb{R}^n$, la dérivée directionnelle

$$f'(x, h) = (t \mapsto f(x + th))'(0)$$

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 $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$, défini sur un intervalle ouvert I contenant 0, telle que $\gamma(I) \subset U$, $\gamma(0) = x$ et $\gamma'(0)$ existe, la dérivée $(f \circ \gamma)'(0)$ existe.

1. Montrer que si f est directionnellement dérivable au sens de Hadamard en x, alors f est directionnellement dérivable au sens classique.
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
$$(f \circ \gamma)'(0) = f'(x, \gamma'(0)).$$

3. **Dérivation en chaîne.** Soit $f : U \subset \mathbb{R}^p \to \mathbb{R}^n$ et $g : V \subset \mathbb{R}^n \to \mathbb{R}^m$ 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 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 en x et

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

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

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

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

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.

Réponses

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

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

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.

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 $\beta : J \to \mathbb{R}^n$, 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 $\alpha :]-\varepsilon, \varepsilon[\to \mathbb{R}^n$ par

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

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

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

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

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

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

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

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

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

3. Soit $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$, 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

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

4. Tout d'abord, si la limite

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

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

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

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

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

Le chemin γ étant dérivable en 0,

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

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

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

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

$$\lim_{i\to+\infty} \frac{f(x+t_ik_i)-f(x)}{t_i}$$

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 $t\neq 0$,

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

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

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

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,

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

existe.

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 $k \to h$ et le second terme du membre de droite tend vers 0 quand t et k tendent vers 0, donc

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

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, elle 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

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

ou de façon équivalente, que

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

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

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

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 $k_i''\to h$. Par hypothèse de dérivabilité au sens de Hadamard, on aurait

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

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

Asymptotique

Comportement asymptotique de $f(x+2h)-2f(x+h)+f(x)$ (par approximation de variation d'ordre 2 par d^2f.)

Mean Value Theorem

(version avec avec enveloppe convexe ? A voir. L'idée est éventuellement d'étendre le cas scalaire au cas des fonctions à valeurs vectorielles ...) Cf McLeod “Mean Value Theorem for Vector-Valued Functions”.

Analycité

Borne sur $f^{(n)}$ et analycité ?

Arguments Matriciels

Différentielle d'objects comme det A ?

Exploiter https://terrytao.wordpress.com/2013/01/13/matrix-identities-as-derivatives-of-determinant-identities/#comment-514937

Convexité

Lien convexité et différentielle d'ordre 2.

Oloid

cf http://www.heldermann-verlag.de/jgg/jgg01_05/jgg0113.pdf, par exemple calcul plan tangent ?

Formes, Fonction Distance, Squelette

TODO: équivalence entre $(d_A(x))^2$ différentiable et x pas sur le squelette de A (deux projections sur $\overline{A}$).

Pousser le bouchon avec $(d_A(x))^2$ convexe et A convexe ?

cf Zolésio.

Références

.

Shapiro, A. 1990. "On Concepts of Directional Differentiability." *Journal of Optimization Theory and Applications* 66 (3): 477–87.