Mémo de calcul différentiel

Frédéric Le Roux (adaptation : Frédéric Klopp) — SU 3M260

2021

Table des matières

I La différentielle

Dans ce chapitre, on étudie la situation suivante : on a une application f définie sur un ouvert Ω d'un espace vectoriel normé E, à valeurs dans un autre espace vectoriel normé F, et un point a de Ω. On veut étudier le comportement de f au voisinage de a. Le principe du calcul différentiel est d'approcher $f(a+h)$, pour des petites valeurs de h, à l'aide d'une application linéaire appelée différentielle de f au point a.

Noter que comme Ω est ouvert, $f(a+h)$ est bien défini pour tout h assez petit. Dans toute la suite, lorsqu'on écrit $f(a+h)$, on suppose implicitement que $a+h$ appartient à Ω.

I.1 Théorie

On rappelle que $\mathcal{L}(E,F)$ désigne l'ensemble des applications linéaires continues de E dans F. On utilisera les deux définitions suivantes.

• Une *application affine* est une application du type $x \mapsto b + L \cdot x$ où $L \in \mathcal{L}(E,F)$ et b est un élément de F, autrement dit la somme d'une application linéaire et d'une constante.

• Soit $o : O \to F$ une application définie sur un ouvert O de E contenant 0. On dira que $o(\vec{h})$ est *négligeable devant $\vec{h}$* si

$$\lim_{\vec{h}\to 0} \frac{\left\|o(\vec{h})\right\|}{\left\|\vec{h}\right\|} = 0,$$

autrement dit

$$\forall \varepsilon > 0 \;\; \exists \delta > 0 \;\; \forall \vec{h} \in O \;\;\; (\left\|\vec{h}\right\|_E < \delta \Rightarrow \left\|o(\vec{h})\right\|_F < \varepsilon \left\|\vec{h}\right\|).$$

On vérifie facilement que cette notion ne change pas lorsqu'on change la norme de E en une norme équivalente, ni quand on change la norme de F en une norme équivalente. Une conséquence importante est qu'en dimension finie, la notion de différentiabilité ne va pas dépendre du choix des normes.

Exercice 1.— **Vérifier** que lorsque $o(\vec{h})$ est négligeable devant $\vec{h}$, on a $\lim_{\vec{h}\to 0} o(\vec{h}) = 0$. **Montrer** que la réciproque est fausse.**Vérifier** que $\left\|\vec{h}\right\|^2$ est négligeable devant $\vec{h}$.

(a) Différentielle

On dit que f est *différentiable* au point a s'il existe une application linéaire continue $L \in \mathcal{L}(E,F)$ telle qu'on ait

$$f(a+\vec{h}) = f(a) + L \cdot \vec{h} + o(\vec{h})$$

où $o(h)$ est négligeable devant h. L'application L est alors appelée *différentielle* de f au point a et notée $Df(a)$ (parfois $D_a f$ ou même $f'(a)$). On dit aussi que f admet un *développement limité à l'ordre* 1 au point a. L'application affine $h \mapsto f(a) + Df(a).h$ s'appelle *partie principale* du développement limité, la quantité $o(h)$ est son *reste*.

Proposition. *L'application L est unique.*

Recette de preuve.— Soient L_1, L_2 deux applications linéaires continues vérifiant toutes les deux la définition de la différentielle, il s'agit de montrer que $L_1 = L_2$. **Montrer** d'abord que $L \cdot h := L_1 \cdot h - L_2 \cdot h$ est négligeable devant h. Il s'agit maintenant de voir qu'une application linéaire L telle que $L \cdot \vec{h}$ est négligeable devant $\vec{h}$ est l'application nulle. Nous allons utiliser l'homogénéité de la norme. Fixons un vecteur $\vec{v} \neq 0$, et posons $h(t) = t\vec{v}$. Lorsque t tend vers 0, $h(t)$ tend vers 0, et donc

$$\lim_{t \to 0} \frac{\|L \cdot (h(t))\|}{\|h(t)\|} = 0.$$

Montrer que cette quantité ne dépend en fait pas de t, **en déduire** qu'elle est nulle. **Conclure**.

Proposition. *Si f est différentiable au point a elle y est continue.*

Recette de preuve.— **Écrire** le DL à l'ordre 1. **Conclure** en utilisant le fait que $Df(a)$ est, par définition, une application continue, et que $o(\vec{h})$ tend vers 0 lorsque $\vec{h}$ tend vers 0.

On dit que f est *différentiable sur Ω* si elle l'est en tout point de Ω.

Attention, on a alors autant d'applications linéaires $Df(a)$ que de points a dans Ω !

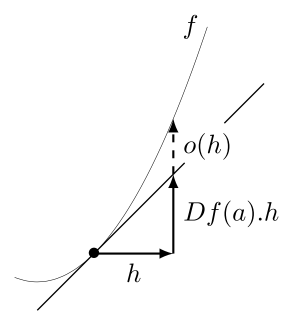
Exemple I : applications affines Soit $L : E \to F$ une application linéaire continue entre deux espacex vectoriels normés, et soit b un élément de F. L'application $f : x \mapsto b + L \cdot x$ est appelée application affine de E dans F. Elle est différentiable sur E, et pour tout a et tout h, $Df(a) \cdot h = L \cdot h$.

Dans ce cas très spécial, $Df(a) = L$, et la différentielle $Df(a)$ ne dépend pas de a.

Exercice 2.—Vérifier l'affirmation précédente. Que vaut la différentielle de la *translation* $x \mapsto x + b$ en un point a ?

Exemple II : fonctions de $\mathbb{R}$ dans $\mathbb{R}$ Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction dérivable en un point a. On a par définition

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



On peut écrire

$$f(a+h) = f(a) + f'(a).h + o(h)$$

en posant $o(h) = f(a+h) - (f(a) + f'(a).h)$; **vérifier** que $o(h)$ est négligeable devant h. Ainsi, on voit que f est différentiable en a, et sa différentielle est l'application $Df(a) : h \mapsto f'(a).h$, qui est linéaire de $\mathbb{R}$ dans $\mathbb{R}$. Réciproquement, on peut montrer que si f est différentiable au point a alors elle est dérivable au point a, et sa dérivée est $f'(a) = Df(a).1$. Ainsi, pour les fonctions de $\mathbb{R}$ dans $\mathbb{R}$, les notions de dérivabilité et différentiabilité coincident. Tout ceci se généralise immédiatement lorsque l'espace d'arrivée est un espace vectoriel normé quelconque F.

Exemple III : fonction de deux variables La formule $f(x,y) = xe^{3y}$ définit une fonction de $\mathbb{R}^2$ dans $\mathbb{R}$. Montrons qu'elle est différentiable au point $(2,1)$. On développe $f(2+h_1, 1+h_2)$ et on essaie de faire apparaître un développement limité à l'ordre 1, c'est-à-dire

— le terme constant $f(2,1)$,

— un terme linéaire qui correspondra à la différentielle au point $(2,1)$,

— un reste qui doit être négligeable devant $\|(h_1,h_2)\|$.

Effectuer le calcul, en remplaçant le terme "e^{3h_2}" par son développement limité en 0, $e^{3h_2} = 1 + 3h_2 + o(h_2)$, et **identifier** le terme constant, le terme linéaire et le reste. On obtient

$$f(2+h_1, 1+h_2) = f(2,1) + e^3(h_1 + 6h_2) + o(h_1, h_2).$$

avec un reste $o(h_1,h_2)$ formé de plusieurs termes. **Vérifier** que ce reste est négligeable devant $\|(h_1,h_2)\|$ en relisant la définition de *négligeable* et en utilisant

$$\lim_{h_2 \to 0} \frac{o(h_2)}{h_2} = 0 \quad \text{et} \quad |h_i| \leq \|(h_1,h_2)\|, \quad i = 1,2.$$

(On a le droit d'utiliser la norme qui nous convient, on prend par exemple la norme $\|.\|_\infty$ pour avoir la dernière majoration.) Puisque l'application $L : (h_1,h_2) \mapsto e^3(h_1 + 3h_2)$ est linéaire, on a bien obtenu un développement limité à l'ordre 1, ce qui prouve que f est différentiable au point $(2,1)$, et que sa différentielle en ce point est l'application L.

Cet exemple est là pour illustrer la définition de la différentielle, mais rassurez-vous, on aura bientôt des outils qui nous permettront de retrouver facilement la différentielle de n'importe quelle fonction donnée par une formule de ce type.

Exemple IV : inversion de matrice Nous allons utiliser des résultats sur l'algèbre des matrices (voir le chapitre sur les espaces vectoriels normés). Pour toute matrice H de norme matricielle $\|H\| < 1$, la matrice $\mathrm{Id} + H$ est inversible et on a

$$(\mathrm{Id} + H)^{-1} = \sum_{k=0}^{+\infty} (-H)^k = \mathrm{Id} - H + H^2 - H^3 + \cdots = \mathrm{Id} - H + o(H) \qquad (\star)$$

Ici on prend n'importe quelle norme d'opérateur sur l'espace vectoriel normé $\mathcal{M}_n(\mathbb{R})$ des matrices carrées.

en posant

$$o(H) = \sum_{k=2}^{+\infty} (-H)^k.$$

Exercice 3.— **Majorer** la norme de $o(H)$, par exemple pour tout $\|H\| < \frac{1}{2}$, pour montrer que $o(H)$ est négligeable devant H. On commencera par mettre H^2 en facteur.

On en déduit que l'application consistant à inverser une matrice est différentiable au point Id, et que sa différentielle est $H \mapsto -H$. L'approximation affine donnée par l'égalité $(\star)$ est très facile à calculer, comme par exemple

$$\begin{pmatrix} 1.09 & 0.07 & 0.05 \\ 0.1 & 0.95 & -0.02 \\ -0.08 & 0.01 & 0.92 \end{pmatrix}^{-1} \simeq \begin{pmatrix} 1 - 0.09 & -0.07 & -0.05 \\ -0.1 & 1 + 0.05 & 0.02 \\ 0.08 & -0.01 & 1 - 0.08 \end{pmatrix}$$

Dans cet exemple les coefficients de H sont de l'ordre de 10^{-1}. Un calcul plus précis donne le résultat

$$\begin{pmatrix} 0.919823 & -0.0672348 & -0.051452 \\ -0.0951178 & 1.05934 & 0.0281987 \\ 0.0810185 & -0.0173611 & 1.08218 \end{pmatrix}.$$

et on voit que l'écart avec les coefficients de notre approximation affine est de l'ordre de 10^{-2}, qui est "beaucoup plus petit" que $\|H\|$.

Exemple V : espaces de fonctions

On se place dans l'espace vectoriel normé $E = C([0,1],\mathbb{R})$, muni de la norme $\|.\|_\infty$. Pour tout élément f de E, l'application $f^2 : x \mapsto f(x)^2$ est encore un élément de E. Pour $f, \vec{h}$ dans E on a

$$(f+\vec{h})^2 = f^2 + 2f\vec{h} + \vec{h}^2$$

et $\left\|\vec{h}^2\right\|_\infty = \left\|\vec{h}\right\|_\infty^2$ est négligeable devant h, donc l'application

$$\begin{array}{ccc} E & \to & E \\ f & \mapsto & f^2 \end{array}$$

est différentiable en tout point f, et sa différentielle est l'application linéaire $\vec{h} \mapsto 2f\vec{h}$, qui est bien continue car $\left\|2f\vec{h}\right\|_\infty \leq 2\|f\|_\infty \left\|\vec{h}\right\|_\infty$.

Exercice 4.— Vérifier cette dernière inégalité, qui dit simplement que le "sup" du produit de deux fonctions positives est plus petit que le produit des "sup".

Voici un deuxième exemple en dimension infinie. On se place dans le sous-espace vectoriel E_1 de E formé des fonctions de classe C^1. On munit E_1 de la norme $\|f\| := \|f\|_\infty + \|f'\|_\infty$. Soit $f \in E_1$; la longueur du graphe de f est donnée par la formule

$$\mathrm{Long}(f) = \int_0^1 \left(1+f'(t)^2\right)^{\frac{1}{2}} dt.$$

Le terme $\left(1+f'(t)^2\right)^{\frac{1}{2}}$ est la norme du vecteur vitesse de la courbe $t \mapsto (t, f(t))$, qui décrit le graphe de f ; cette formule est un cas particulier de la longueur d'une courbe γ, qui est égale à $\int \|\gamma'(t)\|\, dt$.

En calculant la longueur du graphe de $f+h$, on montre que la "fonctionnelle" $\mathrm{Long} : E_1 \to \mathbb{R}^+$ est différentiable en tout point f, et sa différentielle est l'application linéaire continue

$$D\mathrm{Long}(f).h = \int_0^1 \left(1+f'(t)^2\right)^{-\frac{1}{2}} f'(t)h'(t)dt.$$

Exercice 5.— Montrer que la fonctionnelle Long n'est *pas* continue en 0 lorsque E_1 est muni de la norme $\|.\|_\infty$. *Aide : trouver une fonction h uniformément proche de la fonction nulle, mais qui oscille beaucoup, de façon à ce que son graphe ait une longueur très supérieure à celle du graphe de la fonction nulle (h est uniformément petite mais sa dérivée ne l'est pas).* Mieux : montrer que cette application est discontinue en tout élément de E_1. *A fortiori*, cette application n'est pas différentiable.

(b) Dérivées partielles

On se place ici dans le cas où $E = \mathbb{R}^m$ et $F = \mathbb{R}^n$; on notera $(e_1, \dots, e_m)$ la base canonique de E. On considère une application

$$f : \left\{ \begin{array}{ccc} \Omega & \longrightarrow & \mathbb{R}^n \\ x & \longmapsto & (f_1(x), \dots f_n(x)). \end{array} \right.$$

Soit $a = (x_1, \dots, x_m)$ un point de Ω. On dira que la fonction f admet, au point a, une *dérivée partielle par rapport à la i-ème variable* si la fonction

$$t \mapsto f(x_1, \dots, x_{i-1}, x_i + t, x_{i+1}, \dots, x_m)$$

Noter que cette fonction de t est définie sur un voisinage de 0.

est dérivable en $t = 0$. La dérivée est alors notée

$$\frac{\partial f}{\partial x_i}(a) \quad := \quad \lim_{t \to 0} \frac{f(a + te_i) - f(a)}{t}.$$

On voit facilement que f admet une dérivée partielle par rapport à la i-ème variable si et seulement si chacune des fonctions f_i admet une dérivée partielle par rapport à la i-ème variable, et dans ce cas on a

$$\frac{\partial f}{\partial x_i}(a) = \left(\frac{\partial f_1}{\partial x_i}(a), \dots, \frac{\partial f_n}{\partial x_i}(a) \right)$$

Proposition. *Si f est différentiable au point a, elle admet en ce point des dérivées partielles par rapport à toutes les variables, et les dérivées partielles sont données par*

$$\frac{\partial f}{\partial x_i}(a) = Df(a).e_i.$$

La matrice de $Df(a)$ dans les bases canoniques est alors la matrices des dérivées partielles,

$$Jf(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_m}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(a) & \cdots & \frac{\partial f_n}{\partial x_m}(a) \end{pmatrix}.$$

Si f va de $\mathbb{R}^m$ dans $\mathbb{R}^n$, alors $Df(a)$ aussi, ce qui permet de se souvenir que les colonnes correspondent aux coordonnées $x_1, \dots, x_m$ (cette matrice doit pouvoir être multipliée par un vecteur h de $\mathbb{R}^m$).

La matrice $Jf(a)$ est appelée *matrice jacobienne* de f au point a.

Recette de preuve.— Pour la première partie, **écrire** le développement limité de f donné par la différentielle, **appliquer-le** au vecteur te_i, **en déduire** la limite voulue. Pour la seconde, décomposer un vecteur h quelconque dans la base canonique en écrivant $h = \sum h_i e_i$.

La définition de dérivée partielle se généralise au cadre où E est un espace vectoriel normé quelconque, de la façon suivante. Soit a un point de Ω et $\vec{h}$ un vecteur non nul de E. On considère l'application

$$\varphi : t \mapsto f(a + t\vec{h})$$

$t \mapsto a + t\vec{h}$ est le paramétrage de la droite de E passant par le point a au temps $t = 0$, et parcourue à vitesse constante $\vec{h}$.

On dit que f admet, au point a, une *dérivée selon le vecteur $\vec{h}$* si φ est dérivable en 0, autrement dit si la limite

$$\frac{\partial f}{\partial \vec{h}}(a) := \lim_{t \to 0} \frac{f(a + t\vec{h}) - f(a)}{t}$$

existe. On montre comme avant que si f est différentiable en a alors elle y admet une dérivée selon tout vecteur non nul $\vec{h}$ de E, qui est donnée par la relation

$$\frac{\partial f}{\partial \vec{h}}(a) = Df(a) \cdot \vec{h}.$$

Exercice 6.— **Calculer** les dérivées partielles et la matrice jacobienne de la fonction $f(x,y) = xe^{3y}$, introduite dans l'exemple III ci-dessus, au point $a = (2,1)$. **Vérifier** qu'on retrouve la différentielle de f en ce point.

Exercice 7.— Avec les notations du paragraphe précédent, **vérifier** que la dérivée selon le $i$ème vecteur de la base canonique est égal à la $i$ème dérivée partielle :

$$\frac{\partial f}{\partial e_i}(a) = \frac{\partial f}{\partial x_i}(a).$$

(c) Règles de calcul

Dans cette section, E, F, G désignent des espaces vectoriels normés.

Différentielle à valeurs dans un espace produit Soit Ω un ouvert de E. Une application $f = (f_1, \dots, f_n) : \Omega \to \mathbb{R}^n$ est différentiable au point a si et seulement si chaque fonction f_j est différentiable au point a, et dans ce cas on a simplement $Df(a) \cdot \vec{h} = (Df_1(a) \cdot \vec{h}, \dots, Df_n(a) \cdot \vec{h})$. Ceci suit rapidement de la définition.

Plus généralement, soient F_1, F_2 deux espace vectoriels normés, on peut considérer l'espace vectoriel produit $F = F_1 \times F_2$ et le munir par exemple de la norme $\|(v_1, v_2)\|_\infty := \mathrm{Max}(\|v_1\|, \|v_2\|)$. Alors une application $f = (f_1, f_2) : \Omega \to F_1 \times F_2$ est différentiable au point a si et seulement si f_1 et f_2 sont chacune différentiables en a, et dans ce cas on a $Df(a) \cdot \vec{h} = (Df_1(a) \cdot \vec{h}, Df_2(a) \cdot \vec{h})$.

Différentielle d'une somme, d'un produit, de l'inverse On considère deux applications $f_1, f_2 : \Omega \to F$ définies sur un ouvert Ω de E, et un point a de Ω.

Proposition. *On suppose que f_1 et f_2 sont toutes les deux sont différentiables au point a. Alors :*

- *L'application $f_1 + f_2$ est différentiable au point a, et on a $D(f_1 + f_2)(a) = Df_1(a) + Df_2(a)$.*
- *On suppose que $F = \mathbb{R}$. Alors l'application $f_1 f_2$ est différentiable au point a, et on a $D(f_1 f_2)(a) = f_1(a) Df_2(a) + f_2(a) Df_1(a)$.*

Recette de preuve.— À partir de développements limités à l'ordre 1 de f_1 et f_2, on essaie d'en obtenir un pour $f_1 + f_2$. **Faire la somme** des deux développements limités donne une écriture pour $(f_1 + f_2)(a + \vec{h})$, et il s'agit simplement de vérifier que la somme des deux restes $o_1(\vec{h}) + o_2(\vec{h})$ est encore négligeable devant h, ce qui est immédiat.

La démarche est analogue pour le produit : **faire** le produit des deux développements limités pour obtenir une écriture de $f_1(a + \vec{h}) f_2(a + \vec{h})$. Cette fois-ci le reste $o(h)$ est plus compliqué,

montrer néanmoins à nouveau, en majorant $\left|o(\vec{h})\right|$, qu'il est négligeable devant $\vec{h}$. On utilisera la continuité des différentielles $Df_1(a)$ et $Df_2(a)$.

On déduit de ces deux règles que toute fonction polynomiale $P : \mathbb{R}^m \to \mathbb{R}$ est différentiable : en effet une telle fonction est obtenue en effectuant des produits et des sommes à partir des fonctions coordonnées $p_i : (x_1, \dots, x_m) \mapsto x_i$, qui sont linéaires et donc différentiables en tout point.

Différentielle d'une composée On considère deux applications $f : \Omega \to F$, $g : \Omega' \to G$, où Ω est un ouvert de E et Ω' un ouvert de F.

Proposition. *Soit a un point de Ω tel que le point $b = f(a)$ est dans Ω'. Si f est différentiable en a et g différentiable en b alors $g \circ f$ est différentiable en a et la différentielle de $g \circ f$ au point a est la composée de la différentielle de f au point a et de la différentielle de g au point $f(a)$:*

$$D(g \circ f)(a) = Dg(b) \circ Df(a).$$

Recette de preuve.— Remarquons d'abord que la fonction $g \circ f$ est définie sur $\Omega \cap f^{-1}(\Omega')$; par continuité de f au point a, cet ensemble contient une boule ouverte centrée en a. En particulier $g \circ f$ est définie sur un ouvert de E contenant a, comme demandé dans la définition de la différentiabilité au point a.

Pour qu'on puisse évaluer $g(f(x))$, il faut notamment que $f(x)$ appartienne à l'ensemble de définition de g.

Écrire d'abord une “preuve approchée” en “faisant comme si”, dans les développements limités de f et g, les restes étaient nuls : ceci permet en particulier de retrouver rapidement la formule. Pour un argument précis, on écrit les développements limités à l'ordre 1 de f en a et de g en $b = f(a)$,

$$f(a + \vec{h}) = f(a) + Df(a) \cdot \vec{h} + o_1(\vec{h})$$
$$g(b + \vec{k}) = g(b) + Dg(b) \cdot \vec{k} + o_2(\vec{k}).$$

On pose $K(\vec{h}) = Df(a) \cdot \vec{h} + o_1(\vec{h})$, on remarque que $K(\vec{h})$ tend vers 0 lorsque $\vec{h}$ tend vers 0 (c'est exactement la continuité de f en a). **Reporter** alors le premier développement limité dans le second, plus précisément appliquer la deuxième égalité avec $\vec{k} = K(\vec{h})$. Ceci donne un développement limité de $g \circ f$ en a, à condition de savoir **montrer** que le reste

$$o_3(\vec{h}) = Dg(b) \cdot o_1(\vec{h}) + o_2(K(\vec{h}))$$

est négligeable devant $\vec{h}$; ceci va suivre du fait que les restes $o_1(\vec{h}), o_2(\vec{k})$ sont négligeables respectivement devant $\vec{h}$ et $\vec{k}$, et que le quotient $\left\|K(\vec{h})\right\| / \left\|\vec{h}\right\|$ est borné. Au passage, on utilise la majoration caractéristique des applications linéaires continues,

$$\|L \cdot \vec{v}\| \leq \|L\| \, \|\vec{v}\|$$

où $\|L\|$ désigne la norme dans l'espace vectoriel $\mathcal{L}(E, F)$ ou $\mathcal{L}(F, G)$.

Exercice 8.— Généraliser en écrivant la formule donnant la différentielle en un point a de la composée des 3 applications $f_3 \circ f_2 \circ f_1$. *Aide : on appliquera deux fois de suite la formule de composition en écrivant par exemple* $f_3 \circ f_2 \circ f_1 = (f_3 \circ f_2) \circ f_1$. Généraliser à la composée de n applications.

Exercice 9.— Que vaut la différentielle de $T_b \circ f$ lorsque T_b est la translation de F de vecteur b ? de $f \circ T_a$ lorsque T_a est la translation de E de vecteur a ? de $L \circ f$, lorsque L est une application linéaire ? De $f \circ L$?

Ce type de différentielles composées intervient dans les démonstrations, lorsqu'on veut se ramener au cas où $a = 0$ et $Df(a) = \mathrm{Id}$, par exemple dans l'énoncé sur la différentiabilité de f^{-1}.

Différentielle de la réciproque

Proposition. *Soit $f : \Omega_E \to \Omega_F$ un homéomorphisme entre un ouvert Ω_E de E et un ouvert Ω_F de F, et a un point de Ω_E. Si f est différentiable en a et si $Df(a)$ est inversible dans $\mathcal{L}(E, F)$, alors l'application réciproque f^{-1} est différentiable en $b := f(a)$ et la différentielle au point $f(a)$ de l'application f^{-1} est la réciproque de la différentielle de f au point a :*

$$D(f^{-1})(b) = (Df(a))^{-1}.$$

Cet énoncé nous servira au chapitre IV dans la preuve du théorème d'inversion locale.

Recette de preuve.— On traite d'abord le cas où $E = F$, $a = f(a) = 0$ et $Df(a) = \mathrm{Id}$. Dans ce cas, il s'agit de montrer que f^{-1} est différentiable en 0 et que sa différentielle en 0 est l'identité. Pour ceci, on écrit, pour tout $\vec{h}$ assez petit,

$$f(\vec{h}) = \vec{h} + o_1(\vec{h}).$$

Soit $\vec{k}$ assez petit. En appliquant cette égalité à $\vec{h} = f^{-1}(\vec{k})$, on obtient

$$f^{-1}(\vec{k}) = \vec{k} - o_1(f^{-1}(\vec{k}))$$

et il s'agit de montrer que le reste $o_2(\vec{k}) := -o_1(f^{-1}(\vec{k}))$ est négligeable devant $\vec{k}$, sachant que le reste $o_1(\vec{h})$ est négligeable devant $\vec{h}$. En utilisant l'hypothèse $Df(a) = \mathrm{Id}$, **montrer** que, au voisinage de 0, la norme de $f(\vec{h})$ est comparable à celle de $\vec{h}$: Pour tout $\left\|\vec{h}\right\|$ assez petit,

$$\frac{1}{2}\left\|\vec{h}\right\| < \left\|f(\vec{h})\right\| < \frac{3}{2}\left\|\vec{h}\right\|.$$

En remplaçant $\vec{h}$ par $f^{-1}(\vec{k})$, en déduire un encadrement similaire pour $f^{-1}(\vec{k})$ et $\vec{k}$: la norme de $f^{-1}(\vec{k})$ est comparable à celle de $\vec{k}$. Lorsque $\vec{h}$ tend vers 0, $o_1(\vec{h})/\vec{h}$ tend vers zéro, **en déduire** que $o_2(\vec{k})/\vec{k}$ tend aussi vers zéro, ce qu'on voulait.

Où a-t-on utilisé que f est un homéomorphisme ?

On déduit le cas général de ce cas particulier : on translate à la source et au but pour se ramener à $a = f(a) = 0$, puis on compose par $Df(a)^{-1}$ pour se ramener à une différentielle égale à l'identité. Ceci revient à considérer

$$g = Df(a)^{-1} \circ T_{-f(a)} \circ f \circ T_a$$

qui est un homéomorphisme comme composé d'homéomorphismes. **En déduire** f^{-1} en fonction de g^{-1}, puis le fait que f soit différentiable en a par composition (on pourra utiliser les exercices 8 et 9).

I.2 Commentaires

(a) La différentielle

On présente généralement la dérivée d'une fonction f en un point a comme un nombre qui mesure le taux de variation de $f(x)$ lorsque x passe au point a. Le but de cette section [1] est de mettre en avant une présentation un peu différente, plus générale. C'est l'idée de la différentiation vue comme une *approximation linéaire* [2], idée qui est au centre d'une grande partie des mathématiques actuelles.

1. Ce qui suit est adapté du merveilleux livre *The Princeton Companion to Mathematics*.
2. En toute rigueur il faudrait plutôt écrire ici *approximation affine*.

Intuitivement, dire que $f'(a) = m$ revient à dire que si on regarde à travers un puissant microscope le graphe de f dans une petite région autour du point $(a, f(a))$, ce que l'on voit est presque exactement une ligne droite de pente m. En d'autres termes, au voisinage de a, la fonction f est approximativement linéaire. On peut même écrire une formule pour la fonction linéaire g qui approxime f :

$$g(x) = f(a) + m(x - a).$$

Son graphe est la droite de pente m passant par le point $(a, f(a))$. Une façon, un peu plus claire, consiste à écrire

$$g(a + h) = f(a) + mh.$$

Dire que g approxime f au voisinage de a revient à dire que $f(a+h)$ est approximativement égal à $f(a) + mh$ lorsque h est petit.

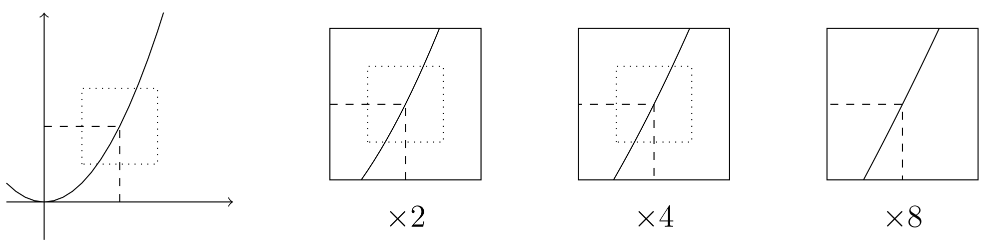


Figure 1 – Zooms successifs sur la fonction $x \mapsto x^2$

Ici, il faut faire un peu attention : après tout, si f est continue au point a alors quand h est petit, $f(a + h)$ sera proche de $f(a)$ et mh sera très petit, de sorte que $f(a+h)$ sera proche de $f(a) + mh$. Cette façon de voir semble marcher pour n'importe quelle valeur de m, et pourtant ce que nous nous voulons dire est très spécifique à la valeur $m = f'(a)$. Ce qui n'arrive qu'avec cette valeur de m, c'est que $f(a + h)$ est non seulement proche de $f(a) + mh$, mais tellement proche que la différence $o(h) = f(a+h) - f(a) - mh$ est petite *comparée à h*. Autrement dit,

$$\frac{o(h)}{h} \text{ tend vers 0 lorsque } h \text{ tend vers 0.}$$

Cette façon de voir peut se généraliser. Les fonctions qui apparaissent en mathématiques, et aussi dans les autres sciences, en ingénierie, en économie, *etc.*, sont souvent des fonctions de plusieurs variables, et peuvent donc être vues comme des fonctions définies sur un espace vectoriel de dimension strictement plus grande que 1. On peut alors se demander si, dans un petit voisinage d'un point, on peut les approcher par des applications linéaires. Lorsque c'est possible, cette approximation est extrêmement utile : une fonction générale peut *a priori* avoir un comportement très compliqué, mais si on peut l'approcher par une fonction linéaire, alors son comportement sera beaucoup plus facile à comprendre, au moins dans de petites régions de l'espace de dimension n. Dans ce cas on peut utiliser toute la machinerie de l'algèbre linéaire et des matrices, qui permet de faire des calculs, surtout si on dispose de l'aide d'un ordinateur.

Imaginez, par exemple, un météorologue s'intéressant à la façon dont la vitesse et la direction du vent varient d'un endroit à l'autre dans une certaine région de l'espace au-dessus de la surface de la Terre. À chaque point (x, y, z) de cette région (x et y représentent par exemple la latitude et la longitude et z l'altitude) on peut associer un vecteur (u, v, w) représentant la vitesse du vent en ce point : u, v, w sont les composantes du vecteur vitesse dans les directions x, y, z.

Déplaçons maintenant très légèrement le point (x, y, z) en choisissant trois petits nombres h, k, l et en considérant le point $(x + h, y + k, z + l)$. En ce nouveau point, nous nous attendons à ce que la vitesse du vent soit différente mais assez proche de celle au point (x, y, z) ; nous l'écrivons donc $(u+p, v+q, w+r)$. Comment la petite variation (p, q, r) du vecteur-vent dépend-elle de la petite variation (h, k, l) de la position du point ? En supposant que le vent n'est pas trop turbulent et que h,k, l sont suffisamment petits, nous nous attendons à ce que cette dépendance soit approximativement linéaire : c'est la façon dont la nature semble fonctionner. Autrement dit, nous nous attendons à ce qu'il existe une application linéaire T telle que (p, q, r) vale approximativement $T(h, k, l)$ lorsque h, k, l sont assez petits. Chacun des nombres p, q, r dépend de chacun des nombres h, k, l, et il nous faut donc 9 nombres pour exprimer cette dépendance linéaire. Sous forme matricielle, elle s'écrit

$$\begin{pmatrix} p \\ q \\ r \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \cdot \begin{pmatrix} h \\ k \\ l \end{pmatrix}.$$

Chaque entrée a_{ij} de la matrice exprime comment l'un des trois nombres p, q, r dépend de l'un des trois nombres h, k, l. Par exemple, si l'on fixe x et z, ce qui revient à poser $h = l = 0$, on obtient $p = a_{12}k$: le coefficient a_{12} représente donc le taux de variation de u lorsque y change. Techniquement, a_{12} est la dérivée partielle $\partial u/\partial y$ au point (x, y, z).

Ceci nous dit comment calculer la matrice de T (appelée matrice jacobienne), mais d'un point de vue conceptuel il vaut mieux éviter les coordonnées et raisonner de façon purement vectorielle. En notant $\mathbf{x}$ pour (x, y, z), $\mathbf{u}(\mathbf{x})$ pour (u, v, w) et $\mathbf{h}$ pour (h, k, l), tout ceci peut se résumer en écrivant la relation

$$\mathbf{u}(\mathbf{x} + \mathbf{h}) = \mathbf{u}(\mathbf{x}) + T(\mathbf{h}) + \mathbf{o}(\mathbf{h}),$$

où $\mathbf{o}(\mathbf{h})$ est petit comparé à $\mathbf{h}$. Ceci nous dit que si nous ajoutons un petit vecteur $\mathbf{h}$ à $\mathbf{x}$, la variation de $\mathbf{u}(\mathbf{x})$ sera approximativement $T(\mathbf{h})$. Cette formule est bien sûr très similaire à la formule du début, $f(a + h) = f(a) + mh + o(h)$.

Plus généralement, soit $\mathbf{u}$ une application de $\mathbb{R}^n$ dans $\mathbb{R}^m$. On dit que cette application est différentiable s'il existe une application linéaire $T : \mathbb{R}^n \to \mathbb{R}^m$ telle que, à nouveau, la formule $\mathbf{u}(\mathbf{x}+\mathbf{h}) = \mathbf{u}(\mathbf{x}) + T(\mathbf{h}) + \mathbf{o}(\mathbf{h})$ soit vérifiée avec $\mathbf{o}(\mathbf{h})$ petit devant $\mathbf{h}$. L'application linéaire T est appelée différentielle de $\mathbf{u}$ au point $\mathbf{x}$.

Le cas $m = 1$ est un cas particulier important. Si $f : \mathbb{R}^n \to \mathbb{R}$ est différentiable au point $\mathbf{x}$, alors la différentielle de f est une application linéaire T de $\mathbb{R}^n$ dans $\mathbb{R}$. La matrice de T est un vecteur ligne de taille n, qui est souvent noté $\nabla f(x)$ et appelé gradient de f au point $\mathbf{x}$. Ce vecteur pointe dans la direction dans laquelle

f augmente le plus vite, et sa longueur est égale au taux de variation dans cette direction.

(b) Vecteur vitesse d'une courbe

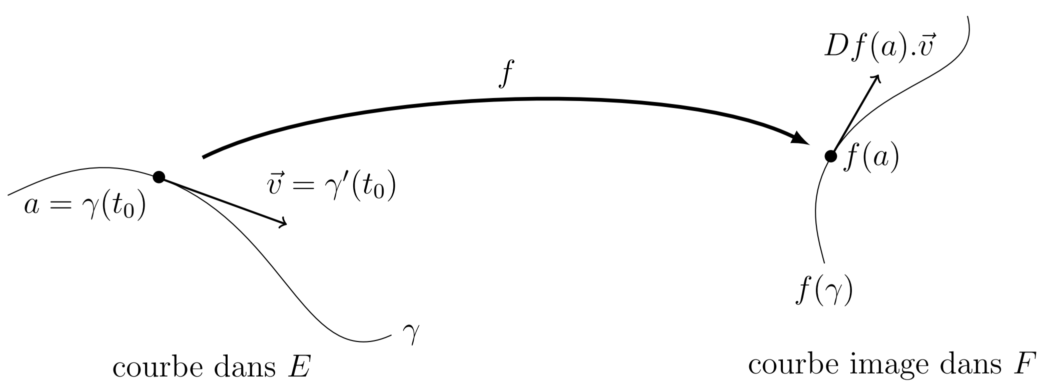
Une *courbe* dans un espace vectoriel normé E est une application $\gamma : I \to E$, définie sur un intervalle I de $\mathbb{R}$; il est utile de se représenter mentalement cette courbe comme le parcours d'un mobile dans E pendant l'intervalle de temps I, le point $\gamma(t)$ représentant la position du mobile au temps t ; on suppose généralement que γ est continue. On dit que γ est *dérivable* en un temps t_0 si la limite

$$\gamma'(t_0) := \lim_{h \to 0} \frac{\gamma(t_0 + h) - \gamma(t_0)}{h}$$

existe. La dérivée $\gamma'(t_0)$ est alors un vecteur de E qu'on appelle aussi *vecteur vitesse de la courbe γ au temps t_0*. Cette définition généralise le cas où $E = \mathbb{R}$. On montre (comme dans le cas où $E = \mathbb{R}$) que lorsque γ est dérivable au temps t_0, elle y est aussi différentiable, et sa différentielle est l'application $h \mapsto \gamma'(t_0).h$, qui est bien linéaire de $\mathbb{R}$ dans E.

Considérons maintenant une application $f : \Omega \to F$, où Ω est un ouvert de E contenant l'image $\gamma(I)$ de notre courbe. Supposons que la courbe γ est dérivable au temps t_0, et que f est différentiable au point $a := \gamma(t_0)$. Une application directe du théorème de composition nous dit que l'application $f \circ \gamma$, qui est une courbe dans F, est aussi dérivable au temps t_0. De plus, son vecteur vitesse est l'image, par la différentielle au point a, du vecteur vitesse de la courbe γ :

$$(f \circ \gamma)'(t_0) = Df(\gamma(t_0)).\gamma'(t_0).$$



Ceci nous donne une interprétation géométrique : *la différentielle au point a indique la façon dont l'application f transforme les vecteurs vitesse des courbes passant par le point a.*

Exercice 10.— **1.** Vérifier qu'une courbe $\gamma : I \to E$ qui est dérivable en t_0 est aussi différentiable en t_0, et que sa différentielle est bien $D\gamma(t_0) : h \mapsto \gamma'(t_0).h$. **2.** En déduire, à l'aide du théorème de composition, la formule $(f \circ \gamma)'(t_0) = Df(\gamma(t_0)).\gamma'(t_0)$. **3.** Que donne la formule dans le cas particulier où γ est le paramétrage d'une droite, $\gamma(t) = a + t\vec{h}$? Nous avons déjà rencontré cette formule dans le cours, pouvez-vous dire où ? Nous l'utiliserons très souvent, retenez-là !

(c) Dérivées partielles

Pour une fonction f de $\mathbb{R}$ dans $\mathbb{R}$, le nombre dérivée $f'(x)$ s'interprète de la façon suivante : *pour un petit accroissement h de la variable x, l'accroissement $f(x+h)-f(x)$ de la fonction est à peu près $f'(x).h$, autrement dit il est proportionnel à h (en première approximation) et $f'(x)$ est le facteur de proportionnalité.*

Pour une fonction f de $\mathbb{R}^m$ dans $\mathbb{R}$, les dérivées partielles peuvent s'interpréter de façon analogue : lorsqu'on fait subir à l'*une* des variables x_i un petit accroissement h_i, les autres variables restant inchangées, la fonction f subit un petit accroissement

$$f(x_1,...,x_{i-1},x_i+h_i,x_{i+1},...,x_m) - f(x_1,...,x_{i-1},x_i,x_{i+1},...,x_m)$$

qui est à peu près égal à

$$h_i \frac{\partial f}{\partial x_i}(x).$$

La formule donnant la différentielle en coordonnées,

$$f(x+h)-f(x) \simeq Df(x).h = \sum_{i=1}^{m} h_i \frac{\partial f}{\partial x_i}(x)$$

s'interprète alors en disant que lorsqu'on modifie *toutes les variables à la fois*, chacune d'une petite quantité, l'accroissement de f est la somme des accroissements dûs à chacune des variables prise individuellement.

(d) La fonctionnelle de longueur

La fonctionnelle de longueur, donnée en exemple plus haut, est très utilisée, dans différents contextes. Sur une surface lisse, ses points critiques sont les “géodésiques”, c'est-à-dire les plus courts chemins. En optique, le temps de parcourt d'un rayon lumineux est $\int n(\gamma(t))\,\|\gamma'(t)\|\,dt$, où $n(P)$ est l'indice de réfraction au point P ; c'est une variante de la fonctionnelle de longueur, et le principe de Fermat dit que le chemin suivi par la lumière est un point critique de la fonctionnelle “temps de parcours”.

I.3 Exercices

Exercice 11.—

1. Montrer que l'application “produit” $(x,y) \mapsto xy$, définie de $\mathbb{R}^2$ dans $\mathbb{R}$, est différentiable sur $\mathbb{R}^2$, et donner sa différentielle en un point (x,y).

2. Plus généralement, montrer que l'application “produit scalaire” $(x,y) \mapsto \langle x,y\rangle$, définie de $\mathbb{R}^N \times \mathbb{R}^N$ dans $\mathbb{R}$, est différentiable sur $\mathbb{R}^N \times \mathbb{R}^N$ et donner sa différentielle.

3. Encore plus généralement, on considère trois espaces vectoriels normés de dimensions finies E_1, E_2, F, et une application bilinéaire $B : E_1 \times E_2 \to F$. On rappelle que, grâce l'hypothèse de dimension finie, B est automatiquement continue : il existe une constante C telle que, pour tout $x \in E_1$ et $y \in E_2$, $B(x,y) \leq C\,\|x\|\,\|y\|$. Montrer que B est différentiable sur $E = E_1 \times E_2$, et que sa différentielle au point $a = (a_1,a_2)$ est l'application linéaire

$$(h_1,h_2) \mapsto B(h_1,a_2) + B(a_1,h_2).$$

On pourra utiliser sur E la norme $\|(h,k)\| = \max\{\|h\|, \|k\|\}$.

Exercice 12.—

1. Soient $f_1, f_2 : E \to \mathbb{R}$ différentiables au point a. **a.** Exprimer l'application produit $x \mapsto f_1(x)f_2(x)$ comme une composée de deux applications différentiables. **b.** Retrouver ainsi la formule de la différentielle d'un produit. (On pourra utiliser l'exercice 11).

2. Plus généralement, soient $f_1 : E \to F_1$ et $f_2 : E \to F_2$ différentiables au point a, et $B : F_1 \times F_2 \to F$ une application bilinéaire. On considère l'application $f : x \mapsto B(f_1(x), f_2(x))$, dont on veut montrer qu'elle est différentiable en a et calculer sa différentielle. **a.** Expliquer pourquoi la question précédente était un cas particulier cette question. **b.** Résoudre cette deuxième question en s'inspirant de la première.

3. Application : donner la différentielle au point a de $f : x \mapsto \langle f_1(x), f_2(x)\rangle$ lorsque f_1, f_2 sont deux applications de E dans $\mathbb{R}^n$ différentiable en a.

Exercice 13.— Calculer la différentielle de l'application $f \mapsto \int_0^1 (f(t))^2 dt$, définie de $C([0,1], \mathbb{R})$ dans $\mathbb{R}$, en un point α quelconque.

II Extrema : conditions d'ordre 1

Dans cette partie, on considère une fonction $f : \Omega \to \mathbb{R}$, où Ω est un ouvert d'un espace vectoriel normé E. Le calcul différentiel fournit des outils pour déterminer les extrema (maxima ou minima) de la fonction f sur Ω, ou, lorsque $E = \mathbb{R}^m$, sur une partie S de Ω donnée par un système d'équations.

II.1 Théorie

(a) Extrema libres

Soit X un espace métrique, $f : X \to \mathbb{R}$ une fonction, et x_0 un élément de X. On dit que

— f *admet un maximum en* x_0 si pour tout x de X, $f(x) \leq f(x_0)$,

— f *admet un minimum en* x_0 si pour tout x de X, $f(x) \geq f(x_0)$,

— f *admet un maximum local en* x_0 s'il existe $\varepsilon > 0$ tel que pour tout élément x de la boule $B(x_0, \varepsilon)$, $f(x) \leq f(x_0)$,

— f *admet un minimum local en* x_0 s'il existe $\varepsilon > 0$ tel que pour tout élément x de la boule $B(x_0, \varepsilon)$, $f(x) \geq f(x_0)$.

Exercice 14.— **Dessiner** l'allure du graphe d'une fonction de $\mathbb{R}$ dans $\mathbb{R}$ qui admet exactement deux minimum locaux dont l'un est un minimum, et un maximum local qui n'est pas un maximum.

Principe de Fermat en dimension 1 Soit f une fonction dérivable de $\mathbb{R}$ dans $\mathbb{R}$. Supposons que f admette un maximum ou un minimum local en un point a. Le principe de Fermat dit que dans sous cette hypothèse, $f'(a) = 0$.

Démontrons ceci en détail. On a le développement limité

$$f(a+t) = f(a) + tf'(a) + o(t)$$

avec $o(t)$ négligeable devant t. Mettons-le sous la forme (pour tout $t \neq 0$ et assez petit)

$$f(a+t) = f(a) + t\left(f'(a) + \frac{o(t)}{t}\right).$$

Puisque $o(t)$ est négligeable devant t, lorsque t tend vers 0, le terme entre parenthèses tend vers $f'(a)$. Supposons que $f'(a) > 0$. Alors pour tout $t \neq 0$ assez petit le terme entre parenthèses est strictement positif, et par conséquent on a

$$\begin{array}{ll} f(a+t) > f(a) & \text{pour tout } t > 0 \text{ assez petit,} \\ f(a+t) < f(a) & \text{pour tout } t < 0 \text{ assez petit.} \end{array}$$

Ceci montre que, si $f'(a) > 0$, f n'admet ni minimum local ni maximum local au point a. Un raisonnement symétrique conduit à la même conclusion lorsque $f'(a) < 0$. Par contraposée, on a la propriété voulue.

Ce raisonnement, bien que n'impliquant qu'une fonction de $\mathbb{R}$ dans $\mathbb{R}$, sera le point clé dans les preuves des théorèmes d'extrema en dimension supérieure.

Le théorème suivant généralise le principe de Fermat en dimension supérieure.

Théorème. *(condition d'ordre 1 sur les extrema) Soit $f : \Omega \to \mathbb{R}$ une fonction définie sur un ouvert d'un espace vectoriel normé E, et a un point de Ω en lequel f est différentiable. Si f admet un maximum local ou un minimum local en a, alors la différentielle en ce point est nulle :*

$$Df(a) = 0_{\mathcal{L}(E,\mathbb{R})}.$$

Un point en lequel la différentielle est l'application nulle, comme dans la conclusion du théorème, est appelé *point critique*.

Recette de preuve.— On considère f comme dans l'énoncé, présentant un maximum ou un minimum local en un point a, on veut montrer que $Df(a) = 0$. Fixons un vecteur non nul $\vec{h}$, il s'agit de montrer que $Df(a) \cdot \vec{h} = 0$.

Donnons deux preuves. La première consiste à se ramener au cas de la dimension 1 traité avant l'énoncé, en regardant la restriction de f aux droites passant par a. On considère l'application $t \mapsto f(a + t\vec{h})$. **Expliquer** précisément pourquoi cette application est définie sur un ouvert de $\mathbb{R}$ contenant 0, et admet un maximum ou minimum local en 0. **En déduire** que $Df(a) \cdot \vec{h} = 0$ (au besoin, relire la section sur les dérivées partielles).

La deuxième preuve consiste à généraliser la preuve de dimension 1. Supposons, par exemple que $Df(a) \cdot \vec{h} > 0$. **Ecrire** le développement limité à l'ordre 1 donné par la différentielle en a, l'appliquer au vecteur $\vec{k} = t\vec{h}$ et faire tendre t vers 0 ($\vec{h}$ étant fixé). En utilisant la définition de "négligeable devant $\vec{k}$" (et notamment la définition de la limite), **montrer** que $f(a + t\vec{h}) > f(a)$ pour tout réel $t > 0$ assez petit, et que $f(a + t\vec{h}) < f(a)$ pour tout réel $t < 0$ assez petit. Conclure.

La conclusion du théorème équivaut à :

$$\forall \vec{h}, \quad Df(a) \cdot \vec{h} = 0.$$

Lorsque f est à valeurs réelles, remarquons que $Df(a) \cdot \vec{h}$ est un nombre (le supposer strictement positif, comme on le fait plus bas, a donc bien un sens !).

(b) Vecteur gradient dans $\mathbb{R}^m$

Soit $f : \Omega \to \mathbb{R}$ une fonction définie et différentiable sur un ouvert Ω de $E = \mathbb{R}^m$, et a un point de Ω. La différentielle $Df(a)$ est alors une application linéaire de $\mathbb{R}^m$ dans $\mathbb{R}$,

$$Df(a) : (h_1, \dots h_m) \longmapsto h_1 \frac{\partial f}{\partial x_1}(a) + \cdots + h_m \frac{\partial f}{\partial x_m}(a),$$

et la matrice jacobienne est la matrice ligne

$$Jf(a) = \left(\tfrac{\partial f}{\partial x_1}(a), \dots, \tfrac{\partial f}{\partial x_m}(a) \right).$$

La transposée de cette matrice est appelé *vecteur gradient* de f au point a, vecteur que l'on note

$$\nabla_a f = \begin{pmatrix} \frac{\partial f}{\partial x_1}(a) \\ \vdots \\ \frac{\partial f}{\partial x_m}(a) \end{pmatrix}.$$

Pour tout vecteur $\vec{h}$, on a alors

$$Df(a) \cdot \vec{h} = \sum_{i=1}^{m} h_i \frac{\partial f}{\partial x_i}(a) = \langle \nabla_a f, \vec{h} \rangle$$

où $\langle \vec{v}, \vec{w} \rangle$ désigne le produit scalaire canonique des deux vecteurs $\vec{v}, \vec{w}$ dans $\mathbb{R}^m$. On voit notamment que $Df(a) \cdot \vec{h}$ est nul si et seulement si $\vec{h}$ est orthogonal au vecteur $\nabla_a f$.

On dit que le vecteur $\nabla_a f$ est *dual* de la forme linéaire $Df(a)$ relativement au produit scalaire.

Interprétation géométrique du gradient *Parmi les vecteurs $\vec{h}$ de norme $\varepsilon > 0$ fixé, le vecteur qui réalise le maximum de $Df(a)\cdot\vec{h}$ est celui qui est colinéaire au vecteur gradient et de même sens* : de façon condensée,

$$\sup_{\|\vec{h}\|=\varepsilon} Df(a)\cdot\vec{h} = Df(a).\vec{h}_0 \quad \text{où} \quad \vec{h}_0 = \varepsilon\frac{\nabla_a f}{\|\nabla_a f\|}$$

(voir l'exercice qui suit). Souvenons-nous que $Df(a)\cdot\vec{h}$ représente l'accroissement à l'ordre 1 de la fonction f lorqu'on se déplace du point a au point $a+\vec{h}$:

$$f(a+\vec{h}) - f(a) = Df(a)\cdot\vec{h} + o(\vec{h}).$$

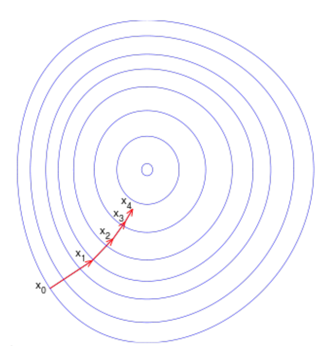
Autrement dit, *le vecteur $\nabla_a f$ indique la direction dans laquelle la fonction f croît le plus vite, lorsqu'on se déplace un tout petit peu en partant du point a.*

Exercice 15.— Soit ∇ un vecteur non nul de $\mathbb{R}^N$, $\varepsilon > 0$ fixé, et $\vec{h}$ un vecteur de norme ε. **1.** Rappeler la majoration de $\langle\nabla,\vec{h}\rangle$ donnée par l'inégalité de Cauchy-Schwarz. **2.** En utilisant le cas d'égalité, trouver le vecteur $\vec{h}$ qui maximise $\langle\nabla,\vec{h}\rangle$ parmi tous les vecteurs de norme ε. Trouver de même celui qui minimise cette quantité. **3.** Faire un dessin représentant les deux vecteurs (maximisant et minimisant), les vecteurs $\vec{h}$ tels que $\langle\nabla,\vec{h}\rangle = 0$, et indiquer les vecteurs $\vec{h}$ pour lesquels cette quantité est strictement positive.

II.2 Commentaires

(a) Gradient et optimisation

Comment programmer un ordinateur pour rechercher les extrema d'une fonction ? La notion de vecteur gradient est à la base d'algorithmes de recherche du maximum d'une fonction, comme l'algorithme du gradient, cf Wikipedia. Le principe est simple : il consiste à partir d'un point au hasard, et à se déplacer d'une certaine longueur (appelée *pas*) dans la direction indiquée par le vecteur gradient en ce point. Si le pas n'est pas trop grand, on se retrouve en un point où la valeur de la fonction est supérieure (conformément à l'interprétation géométrique du gradient). On se dirige à nouveau dans la direction indiquée par le vecteur gradient au nouveau point. On recommence le procédé tant que la norme du vecteur gradient est supérieure à un certain seuil. En pratique, la méthode fournit une bonne approximation d'un maximum local. Pour espérer trouver un maximum absolu, il faut la relancer un grand nombre de fois en partant à chaque fois d'un point différent choisi aléatoirement.



Source Wikipedia

III Applications de classe C^1

Dans ce qui suit, comme avant, E et F sont des espaces vectoriels normés.

III.1 Théorie

(a) L'inégalité des accroissements finis

Théorème (Inégalité des accroissements finis)**.** *Soit $\gamma : [a, b] \to F$ une application continue sur $[a, b]$ et dérivable sur $]a, b[$. On suppose qu'il existe M tel que*

$$\forall t \in]a, b[, \quad \|\gamma'(t)\| \leq M.$$

Alors

$$\|\gamma(b) - \gamma(a)\| \leq M(b - a).$$

Recette de preuve.— *(à sauter en première lecture)* Soit $\varepsilon > 0$. On considère l'ensemble I des nombres $t \in [a, b]$ vérifiant

$$\forall s \in [a, t], \quad \|\gamma(s) - \gamma(a)\| \leq (M + \varepsilon)(s - a) + \varepsilon \quad (\star).$$

On veut montrer que b appartient à I. **Vérifier** que si $t \in I$, alors $[a, t] \subset I$. **Vérifier** que $a \in I$, et qu'il existe $\delta > 0$ tel que I contient l'intervalle $[a, \delta]$. **Vérifier** que I est fermé (on écrira I comme l'image réciproque du fermé $]-\infty, 0]$ par une application continue).

Si vous vous demandez pourquoi on considère cet ensemble I, lisez la preuve jusqu'au bout et reposez-vous ensuite la question. En particulier, à quoi sert ε ?

L'ensemble I est donc un intervalle fermé contenant a et non réduit à a : il existe $t_0 \in]a, b]$ tel que $I = [a, t_0]$. On veut montrer que $t_0 = b$, on suppose par l'absurde que $t_0 < b$. Puisque $t_0 \in]a, b[$, l'hypothèse du théorème nous dit que γ est dérivable en t_0. Ecrivons le développement limité que nous donne cette propriété : on a, pour tout $h \neq 0$ assez petit,

$$\gamma(t_0 + h) = \gamma(t_0) + h\gamma'(t_0) + o(h) = \gamma(t_0) + h\left(\gamma'(t_0) + \frac{o(h)}{h}\right)$$

avec $o(h)$ négligeable devant h. En utilisant la définition de "négligeable devant h" avec notre ε, **en déduire** qu'il existe $\delta > 0$ tel que $[t_0, t_0 + \delta] \subset [a, b]$ et pour tout $h \in [0, \delta]$,

$$\|\gamma(t_0 + h) - \gamma(t_0)\| \leq (M + \varepsilon).h.$$

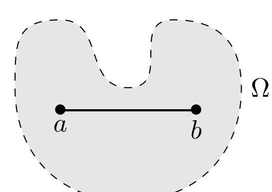
Puisque t_0 appartient à I, il vérifie l'inégalité $(\star)$. **En déduire** que, pour tout $h \in [0, \delta]$, le nombre $t_0 + h$ vérifie aussi cette inégalité. On conclut que $t_0 + \delta$ appartient aussi à I. Ceci contredit l'hypothèse que $I = [a, t_0]$.

On a donc $t_0 = b$, et en particulier

$$\|\gamma(b) - \gamma(a)\| \leq (M + \varepsilon)(b - a) + \varepsilon.$$

Puisque cette inégalité est vraie pour tout $\varepsilon > 0$, un passage à la limite donne l'inégalité $\|\gamma(b) - \gamma(a)\| \leq M(b - a)$ recherchée.

Soit maintenant $f : \Omega \to F$ différentiable sur un ouvert Ω de E. On considère deux points a, b de Ω, et on suppose que le segment $[a, b]$ est inclus dans Ω.



Corollaire 1.

1. *S'il existe une constante M telle que, en tout point x de $[a, b]$, on a $\|Df(x)\| \leq M$, alors*

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

En particulier, si $x \mapsto \|Df(x)\|$ est majoré par M sur Ω et si Ω est convexe, alors f est M-lipschitzienne sur Ω.

2. *S'il existe une constante M telle que, en tout point x de $[a, b]$, on a $\|Df(x) - Df(a)\| \leq M$, alors*

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

Préciser, pour chacune des trois normes apparaissant dans l'énoncé, l'espace sur lequel elle est définie.

Recette de preuve.— On rappelle que l'application $t \mapsto (1-t)a + tb$ donne un paramétrage du segment $[a, b]$, parcouru à la vitesse constante $(b - a)$ entre le temps $t = 0$ et le temps $t = 1$. Pour démontrer le premier point, composons ce paramétrage avec l'application f, autrement dit considérons l'application

$$\gamma : \left\{ \begin{array}{ccc} [0,1] & \longrightarrow & F \\ t & \longmapsto & f((1-t)a + tb). \end{array} \right.$$

Calculer la dérivée $\gamma'(t)$ (on pourra consulter la section sur la vitesse d'une courbe dans le chapitre I). On rappelle que par définition de la norme d'une application linéaire, pour tout point x sur le segment $[a, b]$ et pour tout vecteur $\vec{v}$, on a

$$\|Df(x) \cdot \vec{v}\| \leq \|Df(x)\| \|\vec{v}\| \leq M \|\vec{v}\|.$$

En déduire une majoration du vecteur vitesse de γ. **Conclure** à l'aide du théorème précédent. **Vérifier** la seconde phrase du premier point (critère de lipschitziennité).

Pour le second point, **appliquer** le premier point à la fonction g définie par $g(x) = f(x) - Df(a)(x - a)$.

Corollaire 2. *Soit $f : \Omega \to F$ une application différentiable sur un ouvert connexe Ω. Supposons que $Df(x) = 0$ pour tout $x \in \Omega$. Alors f est constante sur Ω.*

Recette de preuve.— On se place sous les hypothèses de l'énoncé. En appliquant les résultats précédents, **montrer** que l'application f est constante sur toute boule $B(x, \varepsilon)$ incluse dans Ω. On dit que f est *localement constante*. La fin de la preuve consiste à montrer que sur ouvert connexe, toute application localement constante est constante.

Soit maintenant c un nombre, et $\Omega_c = \{x \in \Omega \mid f(x) = c\}$ (cet ensemble est appelé *ligne de niveau c* de la fonction f). En écrivant cet ensemble comme image réciproque d'un fermé par une application continue, **expliquer** pourquoi Ω_c est fermé dans Ω. En utilisant que f est localement constante, **montrer** que cet ensemble est aussi ouvert. **Conclure**.

Exercice 16.— Dans cet exercice, on montre que si la vitesse le long d'une courbe reste proche d'un vecteur $\vec{v}$, alors la courbe reste proche de la droite parcourue à vitesse $\vec{v}$. Plus précisément, soit t un réel positif, et $\gamma : [0, t] \to F$ une application continue, on suppose que γ est dérivable sur $]0, t[$. Soit $\vec{v}$ un vecteur de F, et M un réel tel que

$$\forall s \in [0, t], \|\gamma'(s) - \vec{v}\| \leq M.$$

Montrer l'inégalité

$$\|\gamma(t) - (\gamma(0) + t \cdot \vec{v})\| \leq M |t| \quad (\star).$$

Indication : appliquer le théorème à la fonction $f(s) = \gamma(s) - s\vec{v}$.

Cet exercice est utilisé dans la preuve du théorème de caractérisation des applications de classe C^1.

(b) Applications de classe C^1

Lorsque $f : \Omega \to F$ est différentiable en tout point de Ω, on peut considérer l'application

$$x \mapsto Df(x)$$

qui va de Ω dans l'espace vectoriel normé $\mathcal{L}(E, F)$. Si elle est continue, on dira que f est *de classe C^1 sur Ω*. On montre immédiatement qu'une somme, composition, produit, inverse d'applications de classe C^1 est de classe C^1.

Les énoncés qui suivent fournissent un critère pratique, en particulier lorsque f est donnée par une formule : pour montrer qu'elle est de classe C^1, il suffit de calculer ses dérivées partielles et de vérifier qu'elles sont continues.

Théorème. *On suppose ici que $E = \mathbb{R}^m$. Soit $f : \Omega \to F$ définie sur un ouvert Ω de E. Supposons*

— *que f admet des dérivées partielles en tout point de Ω,*

— *que ces dérivées partielles sont des fonctions continues sur Ω.*

Alors f est différentiable sur Ω.

Recette de preuve.—*(à sauter en première lecture)* Pour simplifier on se place dans $E = \mathbb{R}^2$, on note (x_1, x_2) les coordonnées des points de $\mathbb{R}^2$. On utilise la norme $\|.\|_\infty$ sur E. On se place sous les hypothèses du théorème. Soit $a = (a_1, a_2)$ un point de Ω, et considérons l'application linéaire L définie par

Liste des objets introduits dans la preuve : le point a, l'application L...

$$L(h_1, h_2) = h_1 \frac{\partial f}{\partial x_1}(a) + h_2 \frac{\partial f}{\partial x_2}(a).$$

On cherche à montrer que f est différentiable au point a, et que $Df(a) = L$. Pour ceci, pour tout vecteur $\vec{h} = (h_1, h_2)$ assez petit, on pose $o(\vec{h}) = f(a + \vec{h}) - f(a) - L \cdot \vec{h}$, on veut voir que cette quantité est négligeable devant $\vec{h}$; rappelons que ceci signifie que $o(\vec{h}) / \left\|\vec{h}\right\|$ tend vers 0 lorsque $\vec{h}$ tend vers 0. **Vérifier** que $o(\vec{h})$ est la somme des deux quantités

... un "petit" vecteur $\vec{h}$, les applications o, o_1, o_2...

$$o_1(h_1, h_2) = f(a_1 + h_1, a_2) - \left(f(a_1, a_2) - h_1 \frac{\partial f}{\partial x_1}(a_1, a_2) \right)$$

et

$$o_2(h_1, h_2) = f(a_1 + h_1, a_2 + h_2) - \left(f(a_1 + h_1, a_2) - h_2 \frac{\partial f}{\partial x_2}(a_1, a_2) \right).$$

On va montrer que chacune d'elle est négligeable devant $\vec{h}$. Pour estimer $o_1(\vec{h})$, on introduit la courbe $\gamma_1 : s \mapsto f((a_1 + s, a_2))$. **Calculer**, pour tout s, le vecteur vitesse $\gamma_1'(s)$. Fixons un $\varepsilon > 0$. On utilise maintenant l'hypothèse de continuité des dérivées partielles : il existe $\delta > 0$ tel que pour tout x dans la boule $B(a, \delta)$,

... la courbe γ_1, un nombre $\varepsilon > 0$, un nombre $\delta > 0$. Et c'est tout !

$$\left\| \frac{\partial f}{\partial x_1}(x) - \frac{\partial f}{\partial x_1}(a) \right\| < \varepsilon.$$

On suppose désormais que $\left\|\vec{h}\right\| < \delta$. Le segment entre les points a et $a + (h_1, 0)$ est alors contenu dans la boule $B(a, \delta)$, **en déduire** que les vecteurs vitesse de la courbe γ_1 vérifient, pour tout $s \in [0, h_1]$,

$$\|\gamma_1'(s) - \gamma_1'(0)\| < \varepsilon.$$

Les vecteurs vitesse de γ_1 étant "proches" du vecteur $\gamma_1'(0)$, on en déduit que γ_1 "n'est pas trop loin" de la droite $s \mapsto \gamma(0) + s\gamma'(0)$: plus précisément, on applique le résultat de l'exercice 16 ci-dessus, avec $v = \gamma_1'(0)$; l'inégalité $(\star)$ donne alors

$$\|\gamma_1(h_1) - \gamma_1(0) - \gamma_1'(0)\| \leq \varepsilon \, |h_1| \, .$$

Remplacer γ_1 par sa définition pour obtenir la majoration

$$\left\| o_1(\vec{h}) \right\| \leq \varepsilon \left| h_1 \right| \leq \varepsilon \left\| \vec{h} \right\|.$$

En relisant ce qui précède, **vérifier** qu'on a montré ceci :

$$\forall \varepsilon > 0 \;\; \exists \delta > 0 \;\; \forall \vec{h} \in B(0, \delta), \quad \left\| o_1(\vec{h}) \right\| \leq \varepsilon \left\| \vec{h} \right\|.$$

Vérifier que ceci correspond à la définition de "$o_1(\vec{h})$ est négligeable devant $\vec{h}$".

On estime $o_2(\vec{h})$ de façon tout à fait analogue, en utilisant la courbe

$$\gamma_2 : s \mapsto f((a_1 + h_1, a_2 + s).$$

Écrire les détails de cette estimation, en vous inspirant de ce qui précède. Ceci termine la preuve en dimension 2. La preuve du cas général est très similaire, on écrit $o(\vec{h})$ comme la somme de m fonctions $o_1(\vec{h}), \ldots, o_m(\vec{h})$.

Corollaire. *Lorsque $E = \mathbb{R}^m$, f est de classe C^1 sur Ω si et seulement si les dérivées partielles de f existent et sont continues sur Ω.*

En particulier, lorsque $F = \mathbb{R}^n$, $Df(x)$ dépend continûment de x si et seulement si les coefficients de la matrice jacobienne de f dépendent continûment de x.

Recette de preuve.— Lorsque les dérivées partielles de f existent et sont continues sur Ω, on a déjà vu que f est différentiable sur Ω, il reste à voir que $Df(x)$ dépend continûment de x. Pour cela, **montrer** l'inégalité

$$\|Df(x) - Df(y)\| \leq \max_{i=1,\ldots,N} \left\| \frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_i}(y) \right\|$$

(on a muni E de la norme $\|h\|_1 = |h_1| + \cdots + |h_1|$). Si F est de dimension finie, on peut utiliser l'argument alternatif suivant. Par hypothèses, les coefficients de la matrice jacobienne $Jf(x)$ dépendent continûment de x. Ceci montre que l'application $x \mapsto Jf(x)$ est continue. D'autre part l'application qui associe à une application linéaire sa matrice est un isomorphisme d'espaces vectoriels, c'est donc un homéomorphisme puisqu'en dimension finie toutes les applications linéaires sont continues. Par composition, $x \mapsto Df(x)$ est continue.

Exemple Reprenons la fonction $f(x, y) = xe^{3y}$ dont on avait montré au premier chapitre la différentiabilité par une méthode directe. D'après les théorèmes de calcul différentielle en une variable, cette fonction admet des dérivée partielles, et on a

$$\frac{\partial f}{\partial x}(x, y) = e^{3y} \quad \text{et} \quad \frac{\partial f}{\partial x}(x, y) = 3xe^{3y}$$

qui sont des fonctions continues de (x, y). On en déduit que f est différentiable en tout point de $\mathbb{R}^2$, et sa différentielle est donnée par

$$Df(x, y)(\vec{h}) = e^{3y} h_1 + 3xe^{3y} h_2.$$

III.2 Commentaires

(a) Interprétation physique de l'inégalité des accroissements finis

Si $\gamma(t)$ est la position au temps t d'une voiture, $\|\gamma'(t)\|$ est la vitesse indiquée au compteur au temps t ; $\|\gamma(b) - \gamma(a)\|$ est la distance (à vol d'oiseau) entre le point de départ et le point d'arrivée ; $b - a$ est le temps de parcours, et l'inégalité du théorème des accroissements finis ne dit rien d'autre que ceci : en roulant pendant un temps T avec une vitesse au compteur qui ne dépasse jamais la valeur V, on ne peut pas se retrouver à une distance supérieure à VT du point de départ.

(b) Continuité de Df

Le "raisonnement" suivant est faux :

≪ *En dimension finie, toutes les applications linéaires sont continues, donc la différentielle d'une application différentiable* $f : \mathbb{R}^m \to \mathbb{R}^n$ *est continue, et* f *est automatiquement de classe* C^1 ≫.

Où est l'erreur ?...

La différentielle Df est une application qui prend en entrée une première variable a, puis une seconde variable h. Ces deux variables jouent des rôles très différents. La première variable, a, représente le point en lequel on calcule la différentielle. La différentielle $Df(a)$ est alors une application linéaire, elle associe au vecteur h, qui représente une petite variation du point a, le vecteur $Df(a).h$. Pour un point a fixé, l'application $Df(a) : h \mapsto Df(a).h$ est linéaire et toujours continue (même en dimension infinie, parce que ça fait partie de la définition de différentiabilité). L'application $a \mapsto Df(a)$, elle, n'est pas linéaire, et n'est en général pas continue, même en dimension finie.

III.3 Exercices

Exercice 17.—

1. L'inégalité des accroissements finis n'a d'intérêt que si la différentielle est bornée sur le segment $[a, b]$. Expliquer pourquoi c'est le cas lorsque γ est de classe C^1.

2. En supposant Ω convexe et $\|Df(x)\| \leq M$ pour tout x de Ω, rappeler pourquoi f est M-lipschitzienne sur Ω.

3. Soit $f : \Omega \to F$ de classe C^1, et K un compact convexe de Ω (par exemple une boule fermée). Montrer que f est lipschitzienne sur K. Plus difficile : montrer que c'est encore vrai lorsque K est un compact connexe.

IV Inversion locale, fonctions implicites

Le théorème d'inversion locale et le théorème des fonctions implicites permettent tous les deux, à partir d'informations sur la différentielle d'une application en un certain point, d'obtenir des renseignements sur le comportement de l'application au voisinage de ce point.

Soit $f : E \to F$ une application, et $b = f(a)$ un point de F dans l'image de f. Est-ce que les points de F proches de b ont aussi un antécédent par f ? Le théorème d'inversion locale répond par l'affirmative, dès que la différentielle $Df(a)$ est bijective. De plus, dans ce cas, tout point assez proche de b a un *unique* antécédent proche de a.

Le théorème des fonctions implicites concerne une équation du type $f(x_1, \dots, x_m) = 0$, où f est une fonction de $\mathbb{R}^m$ dans $\mathbb{R}$. Soit P un point de $\mathbb{R}^m$ qui est une solution de cette équation. Y a-t-il d'autres solutions proches de P ? Le théorème donne une réponse très précise, du moment que la dérivée partielle de f par rapport à la dernière variable ne soit pas nulle au point P. Dans ce cas, si l'on modifie suffisamment peu les $m-1$ premières coordonnées du point P, il existe une unique façon de modifier un petit peu la dernière coordonnée de façon à trouver une nouvelle solution de l'équation. Autrement dit, au voisinage du point P, l'équation détermine la dernière coordonnée comme une fonction des $m-1$ autres. De plus, le théorème s'étend aux systèmes d'un nombre quelconque d'équations.

IV.1 Théorie

Dans cette section, les espaces vectoriels normés E, F sont supposés être des espaces de Banach. La complétude nous permettra de faire appel au théorème du point fixe de Banach-Picard.

(a) Difféomorphismes

Une application $\Phi : U \to V$ entre un ouvert U de E et un ouvert V de F est un *C^1-difféomorphisme* si elle est de classe C^1, elle est bijective, et sa bijection réciproque est aussi de classe C^1. La composée de deux C^1-difféomorphismes est un C^1-difféomorphisme, la réciproque d'un C^1-difféomorphisme est un C^1-difféomorphisme (c'est immédiat). En particulier, l'ensemble $\mathrm{Diff}^1(\mathbb{R}^n)$ des C^1-difféomorphismes de $\mathbb{R}^n$ dans $\mathbb{R}^n$ est un groupe pour la loi de composition.

Exemple 1 Toute application linéaire inversible de $\mathbb{R}^n$ dans $\mathbb{R}^n$ est un difféomorphisme. Plus généralement, une application linéaire continue de E dans F est un difféomorphisme si et seulement si elle admet une réciproque qui est une application linéaire continue. On dit alors que l'application linéaire continue est *inversible* (en dimension finie la continuité de la réciproque est automatique, mais pas en dimension infinie). Dans ce cas, on voit que E et F doivent avoir la même dimension (éventuellement infinie).

Exemple 2 Soit $f : \mathbb{R}^2 \to \mathbb{R}^2$ l'application définie par $f(x, y) = (y, x^2)$. Elle n'est pas injective puisque par exemple $f(-1, 0) = f(1, 0)$; elle n'est pas non plus surjective puisque les points (a,b) avec $b<0$ n'ont pas d'antécédant. Notons $U^- = \{(x,y) \in \mathbb{R}^2 \mid x<0\}$, $U^+ = \{(x,y) \in \mathbb{R}^2 \mid x>0\}$, et $V = \{(a,b) \in \mathbb{R}^2 \mid b > 0\}$. On a, pour tout (x,y) dans U^- et tout (a,b) dans V,

$$f(x,y) = (a,b) \Leftrightarrow (x,y) = g(a,b)$$

en posant $g(a,b) = (-\sqrt{b}, a)$. Ceci montre que la restriction $f_{U^-} : U^- \to V$ est bijective, et que son application réciproque est g. Puisque f et g sont de classe C^1, f_{U^-} est un C^1-difféomorphisme entre U^- et V. De même, $f_{U^+} : U^+ \to V$ est un C^1-difféomorphisme.

Exemple 3 L'exponentielle complexe, $z \mapsto e^z$, est une application de $\mathbb{C}$ dans $\mathbb{C}$. En coordonnées, elle s'écrit

$$\exp(x,y) = (e^x \cos(y), e^x \sin(y)).$$

Ce n'est pas un difféomorphisme de $\mathbb{R}^2$ dans $\mathbb{R}^2$ car elle n'est pas injective : pour tout (x,y), on a

$$\exp(x, y+2\pi) = \exp(x,y),$$

et plus généralement, deux points ont la même image si et seulement si ils diffèrent d'une translation verticale de longueur multiple de 2π. On voit donc que la restriction de l'exponentielle à la bande $U = \mathbb{R}\times]0,2\pi[$ est injective. Cette restriction $f = \exp_{|U}$ est en fait un difféomorphisme entre la bande U et l'ouvert V du plan complémentaire du demi-axe des x positifs. Pourquoi la bijection réciproque f^{-1} est-elle de classe C^1 ? On peut donner des formules pour f^{-1} et vérifier sur les formules. Une autre option consiste à calculer d'abord la différentielle de f. En identifiant $\mathbb{R}^2$ et $\mathbb{C}$, on trouve que

Voir cette note d'un article de Michèle Audin pour une représentation graphique de $z \mapsto e^z$.

$$Jf(z) = e^x \begin{bmatrix} \cos(y) & -\sin(y) \\ \sin(y) & \cos(y) \end{bmatrix}$$

qui est une matrice inversible (son déterminant vaut e^{2x}). Le fait que f soit un difféomorphisme découle alors du théorème d'inversion local ci-dessous, et de son corollaire 2.

(b) Théorème d'inversion locale

Soit maintenant f un C^1-difféomorphisme quelconque. En différenciant l'égalité $f^{-1} \circ f = \mathrm{Id}$ en un point a dont l'image est notée b, on obtient

$$D(f^{-1})(b) \circ Df(a) = \mathrm{Id}$$

et on voit que la différentielle de f au point a est inversible (voir l'exemple 1 plus haut). L'un des objets de cette section est de comprendre dans quelle mesure la réciproque est vraie. Le théorème d'inversion locale dit qu'elle est vraie "localement" : si $Df(a)$ est inversible, alors f est un difféomorphisme au voisinage de a.

Théorème. *Supposons que $f : \Omega \to F$ est de classe C^1 sur l'ouvert Ω, et que a est un point de Ω en lequel la différentielle $Df(a)$ est une application linéaire continue inversible d'inverse continu. Alors il existe un ouvert U de E contenant a, et un ouvert V de F contenant $f(a)$, tels que la restriction de f à U soit un C^1-difféomorphisme entre U et V.*

Le point le plus frappant du théorème dit que si $Df(a)$ est injective, alors f est également injective sur un petit voisinage U du point a.

Corollaire 1. *(de l'application ouverte) Soit $f : \Omega \to F$ une application de classe C^1. Supposons que pour tout point a de Ω, la différentielle $Df(a)$ est inversible dans $\mathcal{L}(E, F)$. Alors f est ouverte : l'image par f de tout ouvert $O \subset \Omega$ est un ouvert de F.*

Corollaire 2. *(d'inversion globale) Sous les hypothèses du corollaire précédent, si de plus f est injective, alors c'est un C^1-difféomorphisme sur son image.*

On commence par déduire du théorème les deux corollaires.

Recette de preuve.— Sous les hypothèses du premier corollaire, montrons d'abord que $f(\Omega)$ est un ouvert de F. **Prendre** un point y de $f(\Omega)$, que cherche-t-on ? **Appliquer** la définition de l'image d'un ensemble pour trouver un point x dans Ω. On peut maintenant appliquer le théorème d'inversion locale, **écrire** les objets qu'il nous fournit. On a ainsi trouvé un ouvert V de F, **vérifier** que $y \in V \subset f(\Omega)$. **Trouver** enfin le ε recherché.

Soit maintenant O un ouvert inclus dans Ω. La restriction $f_{|O}$ vérifie les hypothèses du corollaire. On lui applique ce qu'on vient de montrer, et on en déduit que son image $f(O)$ est un ouvert.

Recette de preuve.— Sous les hypothèses du second corollaire, f est une bijection de Ω vers $f(\Omega)$. Il reste à voir que sa réciproque f^{-1} est de classe C^1. Le théorème d'inversion locale s'applique : f est localement un C^1-difféomorphisme au voisinage de n'importe quel point a de Ω. En particulier f^{-1} est aussi un C^1-difféomorphisme au voisinage de n'importe quel point b de $f(\Omega)$, donc sa différentielle en b existe et dépend continûment de b.

La preuve du théorème d'inversion locale est difficile, mais toutes les idées sont déjà présentes dans l'exercice suivant, qui est plutôt facile. On va utiliser des ingrédients très variés du cours de topologie et de calcul différentiel :

1. l'inégalité des accroissements finis,
2. toute application linéaire continue $\mathrm{Id} + M$ avec $\|M\| < 1$ est inversible,
3. le théorème du point fixe de Banach-Picard,
4. la différentiabilité de l'application réciproque d'un homéomorphisme de classe C^1 dont la différentielle est inversible (cf chapitre I, dernière proposition de la section 1.(c)).

Exercice 18.— Soit E un espace de Banach. On considère une application $g : E \to E$ de la forme $g = \mathrm{Id} + \phi$, avec ϕ de classe C^1 vérifiant

$$\|D\phi(a)\| < \frac{1}{2}$$

pour tout point a de E. On va montrer que g est alors un C^1-difféomorphisme. *Chacune des quatre questions ci-dessous utilise l'un des quatre points rappelés ci-dessus.*

1. Montrer que ϕ est $\frac{1}{2}$-lipschitzienne.

2. Dans cette question, on veut montrer que g est une bijection, autrement dit que pour tout y de E il existe un unique x de E tel que $g(x) = y$. Fixons un point y de E. **a.** En utilisant que $g = \mathrm{Id} + \phi$, traduire l'équation $g(x) = y$, d'inconnue x, en une recherche de point fixe pour une certaine application $T : E \to E$. **b.** Montrer que T est contractante. **c.** Conclure.

3. Puisque g est une bijection, elle admet une bijection réciproque g^{-1}. Montrer que g^{-1} est 2-lipschitzienne. L'application g est donc un homéomorphisme.

4. Montrer que, pour tout point a de E, la différentielle $Dg(a)$ est inversible.

5. Montrer que g^{-1} est différentiable en tout point b de E.

Recette de preuve.— *(à sauter en première lecture)* Démontrons le théorème d'inversion locale. Soit $T : x \mapsto x + a$ la translation de E qui envoie 0 sur a, T' la translation de F qui envoie $f(a)$ sur 0, on considère l'application

$$g = (Df(a))^{-1} \circ T' \circ f \circ T.$$

Vérifier que g est définie d'un ouvert de E contenant 0 dans un autre ouvert de E, et qu'on a $g(0) = 0$ et $Dg(0) = \mathrm{Id}$. **Exprimer** aussi f en fonction de g. L'application g est évidemment de classe C^1. Nous allons montrer que g est un difféomorphisme au voisinage de 0 ; il en découlera immédiatement que f sera un difféomorphisme au voisinage de a (comme composée de difféomorphismes).

Le développement limité à l'ordre 1 de g en 0 s'écrit alors

$$g(x) = x + o(x)$$

avec $o(x)$ négligeable devant x. On se donne un y dans E, et on définit $T(x) = y - o(x)$. Comme dans l'exercice, un point fixe de T est un antécédant de y par g. On cherche donc à montrer que, si y est assez proche de 0, l'application T a un unique point fixe proche de 0. Dans l'exercice on appliquait le théorème du point fixe à une application de E dans E. Ici T n'est pas définie de E dans E, et il va d'abord falloir trouver une partie fermée de E qui est stable par T.

Comme f est de classe C^1, l'application o l'est aussi, **que vaut** $Do(0)$? **En déduire** l'existence d'un $\delta > 0$ tel que

$$\|Do(x)\| \leq \frac{1}{2}$$

pour tout $x \in B(0, \delta)$. D'après l'inégalité des accroissements finis, l'application o est 1/2-lipschitzienne sur cette boule ; **vérifier** que T l'est aussi. On suppose désormais que

$$y \in V := B\left(0, \frac{\delta}{2}\right).$$

Notons

$$B = \{x \mid \|x\| \leq \delta\}$$

la boule fermée de rayon δ centrée en 0. **Montrer** que $T(B) \subset B(0, \delta)$. En particulier, $T(B) \subset B$. **Expliquer** pourquoi B est complet. On peut maintenant appliquer le théorème du point fixe contractant à l'application $T_{|B} : B \to B$. La conclusion de tout ceci est : *Pour tout* $y \in V = B(0, \frac{\delta}{2})$ *il existe un unique* $x \in B(0, \delta)$ *tel que* $g(x) = y$.

Notons $h(y)$ ce point x. On a ainsi défini une fonction $h : V \to B(0, \delta)$, et on a $g(h(y)) = y$ pour tout y de V. On a également $h(g(x')) = x'$ pour tout x' de l'ouvert

$$U := B(0, \delta) \cap g^{-1}(V).$$

En effet, un tel x' a son image $y' = g(x')$ dans V, et $h(y')$ est alors (par définition) l'unique point x de $B(0, \delta)$ tel que $g(x) = y'$: comme x' satisfait cette égalité, par unicité on a $x' = h(y') = h(g(x'))$.

Les ensembles U et V sont ouverts. On a $h(V) = U$, $g(U) = V$, et donc $g_{|U} : U \to V$ et $h : V \to U$ sont des bijections réciproques. La première est clairement continue, la seconde l'est

Terence Tao, médaillé Fields 2006, est un mathématicien aux capacités de travail étonnantes. Le 9 septembre 2011, il poste un message sur le forum Mathoverflow demandant s'il existe une version de ce théorème pour les applications qui sont seulement différentiables sur Ω (et pas de classe C^1). Le 12 septembre, à 19h21, il reçoit une réponse lui indiquant un article de Jean Saint-Raymond de 18 pages. Le soir même, à 00h10, il indique qu'il a posté un billet sur son blog expliquant la démonstration.

aussi : **montrer** en effet que h est 2-lipschitzienne, en utilisant que o est $\frac{1}{2}$-lipschitzienne et que h est la réciproque de $x \mapsto x + o(x)$. On a montré que g est un homéomorphisme local.

Montrons enfin que, quitte à restreindre U, et V, $g_{|U}$ est un C^1-difféomorphisme. Soit

$$U' = \{x \in U \mid Dg(x) \text{ inversible}\}.$$

L'ensemble des éléments inversibles de $\mathcal{L}(E, E)$ étant un ouvert, U' est ouvert ; il contient 0. On pose aussi $V' = g(U')$; puisque $g_{|U} : U \to V$ est un homéomorphisme, V' est aussi un ouvert. D'après la différentiabilité de l'application réciproque d'un homéomorphisme de classe C^1, h est différentiable en tout point y de V', et $Dh(y) = (Dg(x))^{-1}$ pour tout $y = g(x)$ dans V'. Comme l'inversion est continue dans $GL(E)$ et que $Dg : U' \to GL(E)$ est continue, Dh est une application continue sur V'. L'application $g : U' \to V'$ est donc un C^1-difféomorphisme.

(c) Exemples d'application du théorème d'inversion locale

Le théorème d'inversion locale permet de montrer l'existence et l'unicité d'une solution d'un système d'équations, même lorsqu'on ne sait pas trouver une formule pour cette solution. Voyons successivement un exemple dans $\mathbb{R}^2$, un exemple dans l'espace des matrices, et un exemple en dimension infinie.

Exercice 19.—(adapté de l'examen deuxième session 2013) Montrer que pour tout a assez proche de 0 et tout b assez proche de 1, le système d'équations

$$\begin{cases} xe^y + 2y &= a \\ 1 + \sin(3x + 4y) &= b \end{cases}$$

a une unique solution (x, y) proche de $(0, 0)$. *On pourra introduire la fonction $f : \mathbb{R}^2 \to \mathbb{R}^2$ définie par $f(x, y) = (xe^y + 2y, 1 + \sin(3x + 4y))$ et remarquer que $f(0, 0) = (0, 1)$.*

Exercice 20.— On définit l'exponentielle d'une matrice M par la série absolument convergente

$$\exp(M) = \mathrm{Id} + M + \frac{1}{2!}M^2 + \frac{1}{3!}M^3 + \cdots .$$

1. Montrer que toute matrice M assez proche de la matrice identité peut s'écrire comme l'exponentielle d'une matrice N proche de la matrice nulle. *On montrera que* $\exp : \mathcal{M}_n(\mathbb{R}) \to \mathcal{M}_n(\mathbb{R})$ *est différentiable en* 0 *et que sa différentielle est inversible.* **2.** En posant $N = \log(M)$, donner un développement limité de log à l'ordre 1 au point Id.

Exercice 21.— On se place dans l'espace vectoriel normé $E = C([0, 1], \mathbb{R})$, muni de la norme $\|.\|_\infty$. **1.** Montrer que l'application $f \mapsto f^2$ n'est pas un C^1-difféomorphisme sur son image. **2.** Montrer qu'elle n'est pas un C^1-difféomorphisme sur aucun voisinage de la fonction nulle. **3.** Montrer que, par contre, sa restriction à l'ouvert des fonctions strictement positives est un difféomorphisme (on pourra relire le paragraphe sur la différentielle de cette application).

(d) Le Théorème des Fonctions Implicites dans $\mathbb{R}^2$

Considérons l'équation $x^2+y^3=1$. Tout nombre réel possédant une unique racine cubique, l'équation équivaut à $y=(1-x^2)^{\frac{1}{3}}$: on dit que l'équation *détermine y en fonction de x*. Par contre, l'équation $x^2+y^2=1$ ne détermine pas y en fonction de x, pour deux raisons : (1) si $1-x^2$ est strictement négatif, aucun nombre y ne satisfait l'équation, et (2) si $1-x^2$ est strictement positif, les deux nombres $y=\pm\sqrt{1-x^2}$ satisfont l'équation.

Généralisons ce vocabulaire. On considère une application $f : E \to F$, où E est un un espace vectoriel normé produit, c'est-à-dire qu'il s'écrit $E = X \times Y$ avec X, Y deux espaces vectoriels normés, et F un espace vectoriel normé. (Dans les exemples précédents, on avait $E=\mathbb{R}^2=\mathbb{R}\times\mathbb{R}$ et $F=\mathbb{R}$.) Soit c un élément fixé de F (une "constante"). On dit qu'une équation du type

$$f(x,y)=c$$

détermine y en fonction de x sur un domaine $U \times V$ de $E = X \times Y$ si, pour tout $x \in U$ donné, il existe un unique $y \in V$ tel que $f(x,y)=c$. Dans ce cas, en notant ϕ la fonction de U dans V qui associe à x cet unique y, on a

$$\forall (x,y) \in U \times V, \quad f(x,y)=c \Leftrightarrow y=\phi(x).$$

Dans ce cas, l'ensemble des couples (x,y) solutions de l'équation $f(x,y)=c$ coincide donc avec le graphe de la fonction ϕ.

Le théorème des fonctions implicites permet de montrer qu'une équation détermine, au moins localement, y en fonction de x.

Théorème (Théorème des fonctions implicites dans $\mathbb{R}^2$: une équation, deux inconnues)**.** *Soit $f : \mathbb{R}^2 \to \mathbb{R}$ de classe C^1, c un nombre, et (a,b) un point tel que $f(a,b)=c$. Supposons que*

$$\frac{\partial f}{\partial y}(a,b) \neq 0.$$

alors il existe $\alpha, \beta > 0$ tels que l'équation $f(x,y)=c$ détermine y en fonction de x sur $U \times V$, avec $U =]a-\alpha, a+\alpha[$ et $V =]b-\beta, b+\beta[$:

$$\forall x \in U \quad \exists! y \in V \quad f(x,y)=c.$$

De plus, la fonction $\phi : U \to V$ définie par

$$f(x,y)=c \Leftrightarrow y=\phi(x)$$

est de classe C^1.

On dit que l'équation $f(x,y)=c$ *détermine localement y en fonction de x au voisinage du point (a,b)*.

Autrement dit, l'ensemble des solutions de l'équation $f(x,y)=c$ coïncide localement avec le graphe d'une fonction de classe C^1.

Exercice 22.— Soit f définie par $f(x,y)=x^2+y^2-1$. Déterminer les points (x_0,y_0) du cercle d'équation $f(x,y)=0$ en lesquels l'hypothèse du théorème est vérifié. Donner des valeurs de α et β qui conviennent. Soit $(x_0,y_0)=(1,0)$; l'équation détermine-t-elle localement y en fonction de x au voisinage de ce point ? Montrer que l'équation détermine localement x en fonction de y au voisinage de ce point.

Exercice 23.— Soit f définie par $f(x,y) = x^3 + y^3 - 3xy$. Montrer que l'ensemble $\mathcal{C}$ des solutions de l'équation $f(x,y) = 0$ est localement le graphe d'une fonction (y fonction de x ou x fonction de y) au voisinage de tout point autre que $(0,0)$.

Exercice 24.— Sous les hypothèses générales du théorème, calculer $\phi'(x_0)$. *Aide : dériver la relation* $f(x, \phi(x)) = 0$.

Recette de preuve.— Le théorème précédent est la version dans $\mathbb{R}^2$ du théorème des fonctions implicites énoncé et démontré plus bas, qui découle lui-même du théorème d'inversion locale. Cependant, en dimension deux, on peut en donner une preuve élémentaire, qui repose sur le calcul différentiel en une variable. Nous allons démontrer que l'équation détermine localement y en fonciton de x, mais pas que la fonction implicite ϕ est de classe C^1 : pour cette propriété, se reporter à la preuve du théorème général.

On se place sous les hypothèses de l'énoncé. La seconde dérivée partielle de f en (x_0, y_0) n'est pas nulle ; pour fixer les idées, supposons qu'elle est strictement positive (le cas négatif se traite de façon symétrique). **Montrer** d'abord qu'on peut trouver deux nombres strictement positifs, α, β tels que

$$\frac{\partial f}{\partial y}(x,y) > 0$$

pour tout $(x,y) \in]x_0 - \alpha, x_0 + \alpha[\times [y_0 - \beta, y_0 + \beta]$. Remarquons que ce nombre $\frac{\partial f}{\partial y}(x,y)$ est la dérivée de l'application $y \mapsto f(x,y)$, qui est une fonction d'une variable. Par conséquent, pour chaque x fixé, $f(x,y)$ croît strictement lorsque y varie entre $y_0 - \beta$ et $y_0 + \beta$. En appliquant cette remarque une première fois pour $x = x_0$, on en déduit

$$f(x_0, y_0 - \beta) < c < f(x_0, y_0 + \beta).$$

Considérons maintenant l'ensemble I des nombres $x \in]x_0 - \alpha, x_0 + \alpha[$ tels que

$$f(x, y_0 - \beta) < c < f(x, y_0 + \beta).$$

Montrer que c'est un ouvert contenant x_0. Quitte à diminuer α, on peut donc supposer que ces deux inégalités ont lieu pour tout $x \in]x_0 - \alpha, x_0 + \alpha[$. Fixons un tel x. **Déduire** de ces deux inégalités qu'il existe un unique y dans $]y_0 - \beta, y_0 + \beta[$ tel que $f(x,y) = c$. Ceci termine la preuve.

Exercice 25.— On peut généraliser l'argument précédent. Considérons $f : \mathbb{R}^m \to \mathbb{R}$ de classe C^1, c un nombre, $a = (a_1, ..., a_m)$ un point tel que $f(a) = c$. Supposons que

$$\frac{\partial f}{\partial x_m}(a) \neq 0.$$

Montrer que l'équation $f(x) = c$ détermine localement x_n comme fonction des autres coordonnées $(x_1, \ldots, x_{m-1})$. On suivra le plus fidèlement possible l'argument donné dans la preuve précédente.

(e) Le théorème des fonctions implites, version générale

On considère une application $f : \Omega \to F$, où Ω est un ouvert l'espace vectoriel normé produit $E = X \times Y$, et F est un espace vectoriel normé.

Théorème (Théorème des fonctions implicites)**.** *Soit $f : \Omega \to F$ de classe C^1, c un élément de F, et (a, b) un point de $X \times Y = E$ tel que $f(a, b) = c$. Supposons que l'application linéaire continue*

$$Df(a, b)_{|\{0\}\times Y} : Y \to F$$

est inversible. Alors l'équation $f(x, y) = c$ détermine localement y en fonction de x au voisinage du point (a, b) : autrement dit, il existe un ouvert U de X contenant a, un ouvert V de Y contenant b tels que $U \times V \subset \Omega$, et une fonction $\phi : U \to V$ de classe C^1 telle que, pour tout $(x, y) \in U \times V$,

Ici x et y ne sont pas des nombres mais des éléments des espaces vectoriels X et Y.

$$f(x, y) = c \quad \Leftrightarrow y = \phi(x).$$

• Il faut noter qu'on a $\phi(a) = b$ puisque $f(a, b) = c$. La conclusion dit qu'au voisinage du point a, l'ensemble L_c des solutions de l'équation $f(\omega) = c\}$ est le graphe d'une application de X dans Y de classe C^1 ; plus précisément,

$$L_c \cap (U \times V) = \{(x, \phi(x)) \mid x \in U\}.$$

• Dans la situation la plus simple qui correspond à l'énoncé précédent, $E = \mathbb{R}^2 = \mathbb{R}\times\mathbb{R}$, X est l'axe des abscisses et Y l'axe des ordonnées, et $F = \mathbb{R}$. L'application $Df(a)$ s'écrit en coordonnées

$$(h, k) \mapsto \frac{\partial f}{\partial x}(a)h + \frac{\partial f}{\partial y}(a)k$$

sa restriction à Y est simplement $k \mapsto \frac{\partial f}{\partial y}(a)k$ et la condition d'inversibilité équivaut à

$$\frac{\partial f}{\partial y}(a) \neq 0.$$

On retrouve ainsi l'énoncé dans $\mathbb{R}^2$ vue à la section précédente.

• Comment se traduit, en pratique, l'hypothèse "$Df(a, b)_{|Y}$ inversible" ? Supposons que X soit de dimension m et Y de dimension n, choisissons des coordonnées $(x_1, \dots, x_m)$ sur X et $(y_1, \dots, y_n)$ sur Y. Puisqu'il existe une application linéaire inversible de Y vers F, ces deux espaces vectoriels ont la même dimension ; choisissons des coordonnées sur F et écrivons $(f_1, \dots, f_n)$ les coordonnées de f. Avec ces notations, la matrice de la différentielle $Df(a, b)$ s'écrit

$$\left(\begin{array}{ccc|ccc} \frac{\partial f_1}{\partial x_1}(a,b) & \cdots & \frac{\partial f_1}{\partial x_m}(a,b) & \frac{\partial f_1}{\partial y_1}(a,b) & \cdots & \frac{\partial f_1}{\partial y_n}(a,b) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(a,b) & \cdots & \frac{\partial f_n}{\partial x_m}(a,b) & \frac{\partial f_n}{\partial y_1}(a,b) & \cdots & \frac{\partial f_n}{\partial y_n}(a,b) \end{array}\right).$$

La matrice a l'air plus impressionnante que d'habitude, mais c'est juste parce que la fonction f a deux sortes de variables, les x_i et les y_j. Le trait vertical sert juste à séparer la matrice en deux blocs, l'un correspondant aux variables x_i et l'autre aux y_j.

Puisque Y correspond aux vecteurs dont les coordonnées x_i sont toutes nulles, la matrice de la restriction $Df(a, b)_{|Y}$ correspond au bloc des coordonnées y_j (situé à droite du trait de séparation, en bleu dans la version électronique). L'hypothèse équivaut donc au fait que la matrice carrée formée par la partie droite de la matrice précédente,

$$\begin{pmatrix} \frac{\partial f_1}{\partial y_1}(a) & \cdots & \frac{\partial f_1}{\partial y_n}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial y_1}(a) & \cdots & \frac{\partial f_n}{\partial y_n}(a) \end{pmatrix}$$

est inversible.

Recette de preuve.— *(à sauter en première lecture)* On déduit le théorème des fonctions implicites du théorème d'inversion locale, de la façon suivante. Sous les hypothèses de l'énoncé, on définit $\overline{f} : \Omega \to X \times F$ en posant $\overline{f}(x, y) = (x, f(x, y))$. **Vérifier** que la différentielle de $\overline{f}$ au point a est

$$D\overline{f}(a, b)(h, k) = (h, Df(a, b).h + Df(a, b).k) = (h, Df(a, b)_{|X}.h + Df(a, b)_{|Y}.k).$$

Soient $(x, z) \in X \times F$, **montrer** que le système $D\overline{f}(a)(h, k) = (x, z)$, d'inconnues (h, k), a une unique solution. Ceci montre que $D\overline{f}(a)$ est bijective, c'est donc une application linéaire inversible.

On peut donc appliquer le théorème d'inversion locale à l'application $\overline{f}$ et au point (a, b) : $\overline{f}$ se restreint en un C^1-difféomorphisme entre un ouvert O contenant (a, b) et un ouvert O' contenant $\overline{f}(a)$. Quitte à diminuer O, on peut supposer qu'il est de la forme $U_0 \times V$, avec U_0 un ouvert contenant a et V un ouvert contenant b. Notons $\overline{g} : O' \to U \times V$ la réciproque de ce difféomorphisme. On définit

$$U = \{x \in U_0, (x, c) \in O'\}.$$

On a, pour tout $(x, y) \in U_0 \times V$,

$$f(x, y) = c \Leftrightarrow \overline{f}(x, y) = (x, c) \Leftrightarrow (x, y) = \overline{g}(x, c) \Leftrightarrow x = \overline{g}_2(x, c).$$

en posant $\overline{g}_2 = \pi_2 \circ \overline{g}$ avec $\pi_2(x, y) = y$. La fonction $\phi = \overline{g}_2$ convient.

(f) Exemples d'utilisation du théorème des fonctions implites

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction de classe C^1. On s'intéresse à l'équation $f(x_1, \ldots, x_n) = 0$. Soit P un point de $\mathbb{R}^n$ qui est solution de l'équation ; le théorème donne une condition pour que l'équation détermine localement x_n en fonction de $x_1, \ldots, x_{n-1}$ au voisinage de P ; comment s'écrit-elle en coordonnées ? Pour simplifier supposons que $n = 3$. Soit X le plan ("horizontal") contenant tous les vecteurs du type $x = (x_1, x_2, 0)$, et Y l'axe ("vertical") contenant les vecteurs du type $(0, 0, x_3)$. La matrice de la différentielle de f au point P est

$$\begin{pmatrix} \frac{\partial f}{\partial x_1}(P) & \frac{\partial f}{\partial x_2}(P) & \frac{\partial f}{\partial x_3}(P) \end{pmatrix}.$$

Sa restriction à la droite Y est simplement $k \mapsto \frac{\partial f}{\partial x_3}(P)k$, et l'hypothèse du théorème est simplement

$$\frac{\partial f}{\partial x_3}(P) \neq 0.$$

Plus généralement, dans $\mathbb{R}^n$, la condition du théorème dit que la dérivée partielle par rapport à x_n est non nulle.

Exercice 26.— On considère la sphère unité de $\mathbb{R}^3$, notée $\mathbb{S}^2$, d'équation $x^2 + y^2 + z^2 = 1$. **1.** En quels points de la sphère l'équation détermine-t-elle localement z fonction de x, y ? On pourra répondre d'abord par un calcul direct, puis comparer avec l'hypothèse du théorème des fonctions implicites. **2.** Montrer qu'en tout point de la sphère, l'équation détermine z fonction de (x,y), ou y fonction de (x,z), ou z fonction de (x,y).

Exercice 27.— On considère l'équation

$$z^2 e^{zx} + 2zy^2 - 1 = 0.$$

1. Trouver toutes les solutions du type $(0,0,z)$.

2. Montrer que pour toutes valeurs assez petites de x et de y, il existe une unique solution (x,y,z) avec z proche de 1. Autrement dit, l'équation définit localement z comme une fonction $z = \phi(x,y)$ au voisinage de la soluion $(0,0,1)$.

3. En dérivant la relation $f(x,y,\varphi(x,y)) = 0$, calculer la différentielle (c'est-à-dire les dérivées partielles) de φ au point $(0,0)$.

4. En déduire une valeur approchée d'une solution avec $x = 0,03$ et $y = -0,04$ (si elle existe...).

On considère maintenant une partie $\mathcal{C}$ de $\mathbb{R}^3$ donnée par un système de deux équations, disons $f_1(x,y,z) = 0$ et $f_2(x,y,z) = 0$. Ce système détermine-t-il localement y et z comme fonction de x ? Le système des deux équations peut s'écrire de façon condensée

$$\begin{pmatrix} f_1(x,y,z) \\ f_2(x,y,z) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

ou encore $f(x,y,z) = c$ en posant $f = (f_1, f_2)$ et $c = (0,0)$. Cette fois-ci, on cherche à appliquer le théorème avec pour X l'axe des abscisses et pour Y le plan vertical contenant les vecteurs du type $(0,y,z)$. En un point P de $\mathcal{C}$, la matrice de $Df_{|Y}(a,b)$ est

$$\begin{pmatrix} \frac{\partial f_1}{\partial y}(P) & \frac{\partial f_1}{\partial z}(P) \\ \frac{\partial f_2}{\partial y}(P) & \frac{\partial f_2}{\partial z}(P) \end{pmatrix}.$$

Le théorème nous dit que si cette matrice est inversible, alors le système détermine localement y et z comme des fonctions de x.

Dans les cas simples, quand on parvient à "résoudre" le système, on utilise la première équation pour exprimer z en fonction de x et de y, puis, avec cette expression, on utilise la deuxième équation pour exprimer y en fonction de x ; on en déduit enfin l'expression de z en fonction de x. Ceci explique pourquoi on s'attend à ce qu'un système de deux équations à trois inconnues permettent d'exprimer deux d'entre elles comme fonction de la troisième.

Exercice 28.— (Deux équations) On considère l'intersection de la sphère $\mathbb{S}^2$ avec le cylindre d'axe vertical passant par le point $(1,0,0)$ et de rayon 1, qui a pour d'équation

$$(x-1)^2 + y^2 = 1.$$

Montrer que le système de deux équations détermine localement y et z en fonction de x, sauf en quatre points à déterminer. Esquisser le dessin de cette intersection et interpréter graphiquement le résultat du calcul.

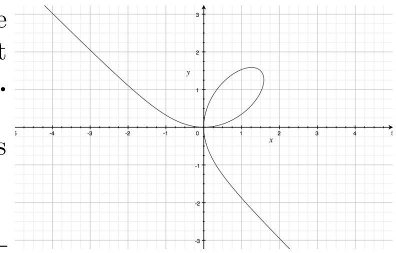
Exercice 29.— **1.** Montrer que l'équation matricielle $M^3 + N^3 - 3MN = \mathrm{Id}$ définit localement N en fonction de M au voisinage du couple solution $(\mathrm{Id}, 0)$. Autrement dit, pour toute matrice M dont les coefficients assez proches de ceux de l'identité, il existe une unique matrice $N = \Phi(M)$ dont les coefficients sont proches de 0, telle que $M^3 + N^3 - 3MN = \mathrm{Id}$. **2.** Calculer la différentielle de Φ au point Id, et écrire le développement limité à l'ordre 1 en ce point.

IV.2 Commentaires

(a) Dessins

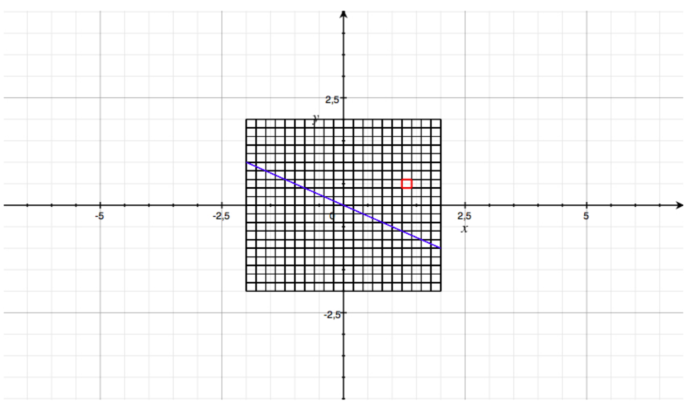
Exercice 30.— Pour chacune de ces fonctions f, l'équation $f(x, y) = 0$ détermine-t-elle y en fonction de x sur le domaine de définition de f ? Sinon, trouver un domaine plus petit où c'est le cas. **1.** $f(x, y) = y - x^2$. **2.** $f(x, y) = y^2 - x$. **3.** $f(x, y) = x^2 + y^2 - 1$ **4.** $f(x, y) = x^3 + y^3 - 3xy$ (pour cette dernière équation, on s'aidera de l'ensemble des solutions dessiné ci-contre). **5.** $f(x, y) = y^3 - x$.

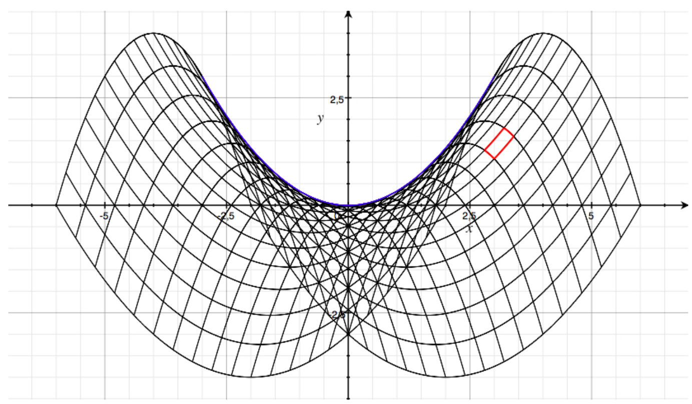
A quelle condition sur a, b, c l'équation $ax + by + c = 0$ détermine-t-elle y en fonction de x ?



Exercice 31.—(Illustration du théorème d'inversion locale)

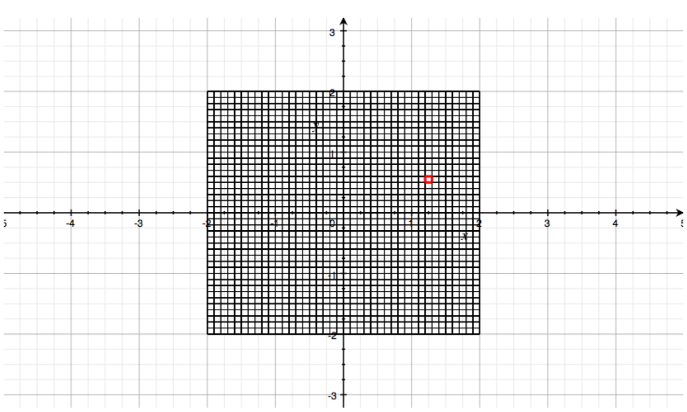
1. On considère l'application $F_1 : \mathbb{R}^2 \to \mathbb{R}^2$ définie par $F_1(x, y) = (2x + y, x^2 - y^2)$. Le dessin suivant représente une grille et son image par l'application F. La droite en bleu, à gauche, est envoyé sur la parabole en bleu, à droite. **a.** Déterminer les "bons" points a, ceux en lesquels l'hypothèse du théorème d'inverison locale est vérifiée. **b.** Soit a un "mauvais" point. L'application F_1 est-elle localement injective en a ? Localement surjective en $F_1(a)$? On pourra s'aider du dessin.

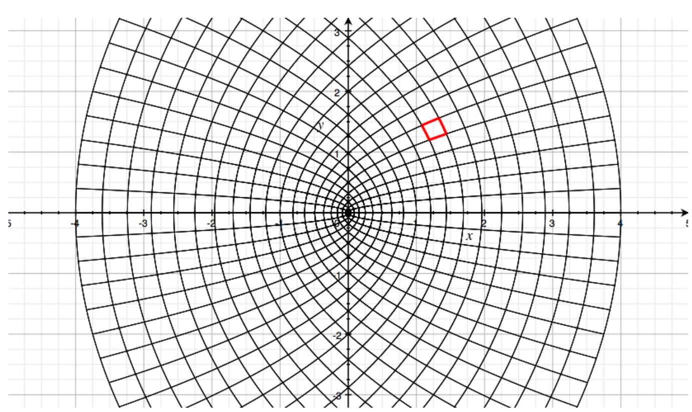




2. Mêmes question pour l'application $F_2 : \mathbb{R}^2 \to \mathbb{R}^2$ définie par $F_2(x, y) = (x^2 - y^2, 2xy)$, représentée ci-dessous.

3. Mêmes questions pour l'application $x \mapsto x^3$, de $\mathbb{R}$ dans $\mathbb{R}$.





Pour en savoir plus, on peut lire le joli article Le pli et la fronce sur le site Images des mathématiques.

Exercice 32.— **1.** On considère l'application $M \mapsto \exp(M)$ de $\mathcal{M}_N(\mathbb{R})$ dans $GL_N(\mathbb{R})$. Montrer qu'elle est différentiable en l'identité, et calculer sa différentielle. On admet que cette application est de classe C^∞ ; en déduire que toute matrice N assez proche de l'identité est l'exponentielle d'une matrice M, et que M est unique si on la suppose assez proche de l'identité.

2. Montrer de même que toute matrice N assez proche de l'identité est le carré d'une unique matrice $M(N)$ proche de l'identité. Donner un développement limité à l'ordre 1 de $M(Id + H)$ lorsque H tend vers 0.