P. Gabriel

Introduction au calcul différentiel

Table des matières

Introduction

On part du rappel qu'une fonction $f : \mathbb{R} \to \mathbb{R}$ est dérivable en $x \in \mathbb{R}$ si la limite du taux d'accroissement

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

existe. Cette limite, notée $f'(x)$, est appelée dérivée de f au point x. De manière équivalente on peut écrire

$$f(x+h) = f(x) + f'(x)\,h + o(h).$$

Autrement dit, f admet un développement limité à l'ordre 1 en x. La droite d'équation $h \mapsto f(x) + f'(x)h$ est la tangente au graphe de la fonction f au point x. On peut également écrire dans $\mathbb{R}^2$

$$\begin{pmatrix} x+h \\ f(x+h) \end{pmatrix} = \begin{pmatrix} x \\ f(x) \end{pmatrix} + \begin{pmatrix} h \\ f'(x)h \end{pmatrix} + \begin{pmatrix} 0 \\ o(h) \end{pmatrix} = \begin{pmatrix} x \\ f(x) \end{pmatrix} + h \begin{pmatrix} 1 \\ f'(x) \end{pmatrix} + o(h) \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Ce qui vient d'être dit sur les fonctions de $\mathbb{R}$ dans $\mathbb{R}$ s'étend sans difficulté aux fonctions de $\mathbb{R}$ dans $\mathbb{R}^2$. Une fonction $f : \mathbb{R} \to \mathbb{R}^2$, définie par ses coordonnées $f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix}$, est dérivable en x si chacune de ses coordonnées f_1 et f_2, qui sont des fonctions de $\mathbb{R}$ dans $\mathbb{R}$, sont dérivables en x. On note alors $f'(x) = \begin{pmatrix} f_1'(x) \\ f_2'(x) \end{pmatrix}$ la dérivée de f, qui n'est plus un nombre cette fois, mais un vecteur de $\mathbb{R}^2$. On a néanmoins toujours le développement limité

$$f(x+h) = f(x) + hf'(x) + o(h),$$

et la droite d'équation $h \mapsto f(x) + hf'(x)$ est toujours la tangente au graphe de la fonction f au point x. On peut également écrire dans $\mathbb{R}^3$

$$\begin{pmatrix} x+h \\ f_1(x+h) \\ f_2(x+h) \end{pmatrix} = \begin{pmatrix} x \\ f_1(x) \\ f_2(x) \end{pmatrix} + \begin{pmatrix} h \\ f_1'(x)h \\ f_2'(x)h \end{pmatrix} + \begin{pmatrix} 0 \\ o(h) \\ o(h) \end{pmatrix} = \begin{pmatrix} x \\ f_1(x) \\ f_2(x) \end{pmatrix} + h \begin{pmatrix} 1 \\ f_1'(x) \\ f_2'(x) \end{pmatrix} + h \begin{pmatrix} 0 \\ \varepsilon_1(h) \\ \varepsilon_2(h) \end{pmatrix},$$

où $\varepsilon_1(h) \to 0$ et $\varepsilon_2(h) \to 0$ quand $h \to 0$.

Regardons maintenant ce qui se passe pour les fonctions de $\mathbb{R}^2$ dans $\mathbb{R}$. Les graphes de telles fonctions sont des nappes, et la notion de droite tangente en un point est remplacée par celle de plan tangent. En d'autres termes, le développement limité d'une fonction $f : \mathbb{R}^2 \to \mathbb{R}$ en un point $x \in \mathbb{R}^2$ s'écrit sous la forme

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

où $L_x : \mathbb{R}^2 \to \mathbb{R}$ est une application linéaire, de sorte que le plan d'équation $h \mapsto f(x) + L_x(h)$ est le plan tangent au graphe de f en x. Si l'on écrit $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ et $h = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$, alors le développement limité peut s'écrire sous la forme

$$f(x_1 + h_1, x_2 + h_2) = f(x_1, x_2) + h_1 L_x(1,0) + h_2 L_x(0,1) + o\begin{pmatrix} h_1 \\ h_2 \end{pmatrix}.$$

On peut également écrire dans $\mathbb{R}^3$

$$\begin{pmatrix} x_1 + h_1 \\ x_2 + h_2 \\ f(x_1 + h_1, x_2 + h_2) \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ f(x_1, x_2) \end{pmatrix} + \begin{pmatrix} h_1 \\ h_2 \\ L_x(h_1, h_2) \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ o(h) \end{pmatrix}.$$

1 Différentielle

Dans toute la suite, $(E, \|\cdot\|_E)$ et $(F, \|\cdot\|_F)$ sont deux espaces vectoriels normés de dimension finie sur $\mathbb{R}$ (par exemple, mais pas seulement, $E = \mathbb{R}^n$ et $F = \mathbb{R}^p$), et Ω est un ensemble ouvert de E.

1.1 Définition et premières propriétés

Définition/Proposition 1.1. *Soit $x \in \Omega$ et soit $f : \Omega \to F$. On dit que f est différentiable en x s'il existe une application linéaire $L_x : E \to F$ telle que*

$$f(x+h) = f(x) + L_x(h) + o(h).$$

Si une telle application linéaire existe, elle est unique et on la note $L_x = \mathrm{d}f(x)$.

Preuve de l'unicité.

$$f(x+h) = f(x) + L_1(h) + o(h) = f(x) + L_2(h) + o(h)$$
$$\implies L_1(h) - L_2(h) = o(h) = \|h\|\varepsilon(h)$$

Pour h fixé on a pour tout $t > 0$ assez petit

$$L_1(th) - L_2(th) = t\|h\|\tilde{\varepsilon}(t)$$

et donc par linéarité

$$L_1(h) - L_2(h) = \|h\|\tilde{\varepsilon}(t) \xrightarrow[t\to 0]{} 0,$$

d'où $L_1(h) = L_2(h)$. □

Exemple important 1. Si $f :]a, b[\to \mathbb{R}$ dérivable est en x, $f(x+h) = f(x) + f'(x)h + o(h)$ donc f est différentiable en x, et $\mathrm{d}f(x) : h \mapsto f'(x)h$. Réciproquement si f est différentiable en x, alors f est dérivable en x et $f'(x) = \mathrm{d}f(x)(1)$.

Exemple important 2. Si $f : E \to F$ est une application linéaire, alors elle est différentiable en tout point x de E et $\mathrm{d}f(x)(h) = f(h)$.

Proposition 1.2. *Si f est différentiable en x, alors f est continue en x.*

Démonstration. En dimension finie, toutes les applications linéaires sont continues, d'où le résultat. □

Proposition 1.3. *On a les propriétés suivantes :*

i) Si $f : \Omega \to F$ et $g : \Omega \to F$ sont deux fonctions différentiables en $x \in \Omega$, alors pour tout $\alpha, \beta \in \mathbb{R}$ la fonction $\alpha f + \beta g$ est différentiable en x et

$$\mathrm{d}(\alpha f + \beta g)(x) = \alpha\, \mathrm{d}f(x) + \beta\, \mathrm{d}g(x).$$

ii) Si $f : \Omega \to \mathbb{R}$ et $g : \Omega \to \mathbb{R}$ sont deux fonctions différentiables en $x \in \Omega$, alors la fonction fg est différentiable en x et

$$\mathrm{d}(fg)(x) = f(x)\,\mathrm{d}g(x) + g(x)\,\mathrm{d}f(x).$$

iii) Soit $f : \Omega \to F$ et soit $g : \Omega' \to G$, où Ω' est un ouvert de F et G est un troisième evn de dimension finie sur $\mathbb{R}$. Si $f(\Omega) \subset \Omega'$, f est différentiable en x et g est différentiable en $f(x)$, alors $g \circ f$ est différentiable en x et

$$\mathrm{d}(g \circ f)(x) = \mathrm{d}g(f(x)) \circ \mathrm{d}f(x).$$

Exercice. *Pour $f : \Omega \to]0, +\infty[$ différentiable en x, montrer que $1/f$ est différentiable en x et calculer sa différentielle.*

1.2 Dérivées directionnelles

Si $f:\Omega\to F$ est différentiable en x, alors pour tout v fixé dans E on a

$$f(x+tv)=f(x)+\mathrm{d}f(x)(tv)+o(t)=f(x)+t\,\mathrm{d}f(x)(v)+t\varepsilon(t),$$

et donc

$$\frac{f(x+tv)-f(x)}{t}=\mathrm{d}f(x)(v)+\varepsilon(t)\xrightarrow[t\to 0]{}\mathrm{d}f(x)(v).$$

En d'autres termes, la fonction $g:t\mapsto f(x+tv)$ est dérivable en 0 et $g'(0)=\mathrm{d}f(x)(v)$.

Définition 1.4. *Soient $x\in\Omega$, $v\in E$ et $f:\Omega\to F$. Si la quantité $\frac{1}{t}\big(f(x+tv)-f(x)\big)$ a une limite quand $t\to 0$, on dit que f admet une dérivée directionnelle en x dans la direction v. On notera cette limite $D_vf(x)$, qui est un vecteur de F.*

Proposition 1.5. *Si $f:\Omega\to F$ est différentiable en $x\in\Omega$, alors f admet des dérivées directionnelles en x dans toutes les directions $v\in E$, et on a $D_vf(x)=\mathrm{d}f(x)(v)$.*

⚠ La réciproque n'est absolument pas vraie. Considérer par exemple la fonction

$$f:\begin{array}{ccc}\mathbb{R}^2 & \to & \mathbb{R}\\ x & \mapsto & \begin{cases}1 & \text{si } x_2=x_1^2\neq 0\\ 0 & \text{sinon}\end{cases}\end{array}$$

qui a des dérivées directionnelles (nulles) en 0 dans toutes les directions mais n'est même pas continue en 0.

1.3 Dérivées partielles

On considère ici le cas $E=\mathbb{R}^n$. Cet espace admet une base dite *canonique* $(e_i)_{1\leqslant i\leqslant n}$ où

$$e_i=\begin{pmatrix}0\\ \vdots\\ 0\\ 1\\ 0\\ \vdots\\ 0\end{pmatrix}\begin{array}{l}\\ \\ \\ \leftarrow \text{position } i.\\ \\ \\ \\ \end{array}$$

Pourquoi "canonique" ?

Les éléments x de $\mathbb{R}^n$ s'écrivent naturellement $x=\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix}$ et donc $x=x_1e_1+x_2e_2+\cdots+x_ne_n$.

On a alors dans $\mathbb{R}^n$ des dérivées directionnelles particulières selon les vecteurs de la base canonique $D_{e_i}f(x)=\mathrm{d}f(x)(e_i)$. Comme nous allons le voir, ces dérivées directionnelles particulières sont en fait les dérivées partielles de la fonction f.

Soit $f:\Omega\to F$, où Ω est un ouvert de $\mathbb{R}^n$, et soit $x\in\Omega$. Pour $i\in\{1,\cdots,n\}$ on définit alors au voisinage de x_i la fonction $g_i:t\mapsto f(x_1,\cdots,x_{i-1},t,x_{i+1},\cdots,x_n)$, qui prend ses valeurs dans F.

Définition 1.6. *On dit que f admet une dérivée partielle en x par rapport à la i-ème variable si g_i est dérivable en x_i. On note cette dérivée partielle $\dfrac{\partial f}{\partial x_i}(x)$, qui est donc un vecteur de F.*

Proposition 1.7. *La fonction f admet une dérivée partielle par rapport à sa i-ème variable si et seulement si elle admet une dérivée directionnelle dans la direction e_i, et on a alors $\dfrac{\partial f}{\partial x_i}(x)=D_{e_i}f(x)$.*

Démonstration. Immédiat, puisque $\frac{1}{t}\big(g_i(x_i+t)-g_i(x_i)\big)=\frac{1}{t}\big(f(x+te_i)-f(x)\big)$. □

Remarque 1.8. *Si on note* $\mathrm{d}x_i$ *la forme linéaire* $h \mapsto h_i$*, alors on a la décomposition* $\mathrm{d}f(x) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x)\,\mathrm{d}x_i$*.*

On considère maintenant le cas où $F = \mathbb{R}^p$, en plus de $E = \mathbb{R}^n$. Une fonction $f : \Omega \to F$ s'écrit alors coordonnées par coordonnées

$$f(x) = f(x_1, \cdots, x_n) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_p(x) \end{pmatrix} = \begin{pmatrix} f_1(x_1, \cdots, x_n) \\ \vdots \\ f_p(x_1, \cdots, x_n) \end{pmatrix}.$$

Définition 1.9 (Jacobienne)**.** *Si* $f : \Omega \to \mathbb{R}^p$ *admet des dérivées partielles en un point* $x \in \Omega$*, on appelle matrice jacobienne de* f *en* x *la matrice*

$$\operatorname{Jac} f(x) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(x) & \cdots & \dfrac{\partial f}{\partial x_n}(x) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_p}{\partial x_1}(x) & \frac{\partial f_p}{\partial x_2}(x) & \cdots & \frac{\partial f_p}{\partial x_n}(x) \end{pmatrix} = \left(\frac{\partial f_i}{\partial x_j}(x) \right)_{\substack{1 \leqslant i \leqslant p \\ 1 \leqslant j \leqslant n}}.$$

Définition 1.10 (Gradient)**.** *Si* $f : \Omega \to \mathbb{R}$ *admet des dérivées partielles en un point* $x \in \Omega$*, on appelle gradient de* f *en* x *le vecteur*

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix}.$$

Proposition 1.11. *Si* $f : \Omega \to \mathbb{R}^p$ *est différentiable en* x*, alors* $\mathrm{d}f(x)(h) = \operatorname{Jac} f(x) \cdot h$*, où* $\cdot$ *désigne la multiplication matricielle.*

Si $f : \Omega \to \mathbb{R}$ *est différentiable en* x*, alors* $\mathrm{d}f(x)(h) = \langle \nabla f(x), h \rangle$*, où* $\langle \cdot, \cdot \rangle$ *désigne le produit scalaire usuel (canonique) de* $\mathbb{R}^n$*.*

Démonstration. Découle directement de la Remarque 1.8. □

2 Accroissements finis et applications

2.1 Préliminaires

Définition 2.1. *On dit qu'une fonction $f : \Omega \to F$ est de classe C^1 si elle est différentiable en tout point x de Ω et l'application $x \mapsto \mathrm{d}f(x)$ est continue de Ω dans $\mathcal{L}(E,F)$, l'espace des applications linéaires de E dans F muni de la norme d'applications linéaires.*

Exemple. *Si $f : E \to F$ est une application linéaire, alors elle est de classe C^1 sur E. En effet on a déjà vu qu'elle était différentiable en tout $x \in E$ avec $\mathrm{d}f(x)(h) = f(h)$.*

Proposition 2.2. *Soit $f : \Omega \to F$ et soit $g : \Omega' \to G$, où Ω' est un ouvert de F et G est un troisième evn de dimension finie sur $\mathbb{R}$. Si f est de classe sur Ω, que $f(\Omega) \subset \Omega'$ et que g est de classe C^1 sur Ω', alors $g \circ f$ est de classe C^1 sur Ω.*

Démonstration. On a déjà vu au point *iii)* de la Proposition 1.3 que $g \circ f$ est alors différentiable en tout x de Ω et que

$$\mathrm{d}(g \circ f)(x) = \mathrm{d}g(f(x)) \circ \mathrm{d}f(x).$$

Il reste à vérifier que cette application est continue de Ω dans $\mathcal{L}(E,G)$. Soit $x \in \Omega$ et soit $(x_n) \subset \Omega$ telle que $x_n \to x$ dans E. On veut montrer que $\mathrm{d}(g\circ f)(x_n)$ converge vers $\mathrm{d}(g\circ f)(x)$ dans $\mathcal{L}(E,G)$. Pour cela on écrit

$$\begin{aligned}
\|\mathrm{d}(g\circ f)(x_n) - \mathrm{d}(g\circ f)(x)\|_{\mathcal{L}(E,G)} &= \|\mathrm{d}g(f(x_n)) \circ \mathrm{d}f(x_n) - \mathrm{d}g(f(x)) \circ \mathrm{d}f(x)\|_{\mathcal{L}(E,G)} \\
&\leqslant \|[\mathrm{d}g(f(x_n)) - \mathrm{d}g(f(x))] \circ \mathrm{d}f(x_n)\|_{\mathcal{L}(E,G)} + \|\mathrm{d}g(f(x)) \circ [\mathrm{d}f(x_n) - \mathrm{d}f(x)]\|_{\mathcal{L}(E,G)} \\
&\leqslant \|\mathrm{d}g(f(x_n)) - \mathrm{d}g(f(x))\|_{\mathcal{L}(F,G)} \|\mathrm{d}f(x_n)\|_{\mathcal{L}(E,F)} + \|\mathrm{d}g(f(x))\|_{\mathcal{L}(F,G)} \|\mathrm{d}f(x_n) - \mathrm{d}f(x)\|_{\mathcal{L}(E,F)}.
\end{aligned}$$

Comme f est différentiable en x, elle est continue en x, et donc $f(x_n) \to f(x)$ dans F. Comme $y \mapsto \mathrm{d}g(y)$ est continue par hypothèse, on en déduit que $\mathrm{d}g(f(x_n)) \to \mathrm{d}g(f(x))$ dans $\mathcal{L}(F,G)$. D'autre part, toujours par hypothèse, $x \mapsto \mathrm{d}f(x)$ est continue donc $\mathrm{d}f(x_n) \to \mathrm{d}f(x)$ dans $\mathcal{L}(E,F)$. En particulier $\|\mathrm{d}f(x_n)\|_{\mathcal{L}(E,F)}$ est bornée. Finalement on en déduit bien que $\|\mathrm{d}(g\circ f)(x_n) - \mathrm{d}(g\circ f)(x)\|_{\mathcal{L}(E,G)} \to 0$. ☐

Corollaire 2.3. *Soit $f : \Omega \to F$ de classe C^1 et soient $x, y \in \Omega$ tels que $[x,y] := \{tx + (1-t)y,\ t \in [0,1]\} \subset \Omega$. Alors l'application $\varphi : [0,1] \to F$, $\varphi(t) = f(tx + (1-t)y)$ est de classe C^1 sur $[0,1]$ et*

$$\varphi'(t) = \mathrm{d}f(tx + (1-t)y)(x-y).$$

Démonstration. Il y a juste besoin de vérifier la formule de φ', puisque la fonction $g : t \mapsto tx + (1-t)y$ est clairement de classe C^1 car linéaire. Pour cela on écrit pour $h \in \mathbb{R}$

$$h\varphi'(t) = \mathrm{d}\varphi(t)(h) = [\mathrm{d}f(g(t) \circ \mathrm{d}g(t)](h) = \mathrm{d}f(g(t))(hg'(t)) = h\,\mathrm{d}f(tx + (1-t)y)(x-y),$$

ce qui donne le résultat. ☐

Lemme 2.4. *Soit $u : [a,b] \to F$ une fonction continue. Alors on a*

$$\left\| \int_a^b u(t)dt \right\| \leqslant \int_a^b \|u(t)\|\, dt.$$

Démonstration. On approche l'intégrale sur $[a,b]$ par des sommes de Riemann, ce qui est justifié pour une fonction continue sur $[a,b]$:

$$\int_a^b u(t)\, dt = \lim_{n\to\infty} \frac{b-a}{n} \sum_{i=1}^n u\Big(a + i\frac{b-a}{n}\Big),$$

et par l'inégalité triangulaire il vient

$$\left\| \frac{b-a}{n} \sum_{i=1}^n u\Big(a + i\frac{b-a}{n}\Big) \right\| \leqslant \frac{b-a}{n} \sum_{i=1}^n \left\| u\Big(a + i\frac{b-a}{n}\Big) \right\|.$$

On obtient alors le résultat en passant à la limite $n \to \infty$. ☐

2.2 Inégalité des accroissements finis

Théorème 2.5. *Soit $f : \Omega \subset E \to F$ de classe C^1 et soient $x, y \in \Omega$ tels que $[x, y] \subset \Omega$. Alors on a l'*inégalité des accroissements finis

$$\|f(x) - f(y)\|_F \leqslant \sup_{t\in[0,1]} \|\,\mathrm{d}f(tx + (1-t)y)\|_{\mathcal{L}(E,F)} \|x - y\|_E.$$

Démonstration. Soit φ la fonction définie par $\varphi(t) = f(tx + (1-t)y)$. On a vu que φ est de classe C^1 sur l'intervalle $[0, 1]$ et que $\varphi'(t) = \mathrm{d}f(tx+(1-t)y)(x-y)$, donc en partant de $\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)dt$, il vient que

$$\|f(x) - f(y)\|_F = \|\varphi(1) - \varphi(0)\|_F = \left\| \int_0^1 \varphi'(t)\, dt \right\|_F \leqslant \int_0^1 \|\varphi'(t)\|_F\, dt.$$

Or $\varphi'(t) = \mathrm{d}f(tx + (1-t)y)(x - y)$ et donc $\|\varphi'(t)\|_F \leqslant \|\,\mathrm{d}f(tx + (1-t)y)\|_{\mathcal{L}(E,F)}\|x - y\|_E$. Comme $z \mapsto \mathrm{d}f(z)$est continue sur le compact $[x, y]$, on a que $\sup_{t\in[0,1]} \|\,\mathrm{d}f(tx + (1-t)y)\|_{\mathcal{L}(E,F)} < \infty$, et le résultat du théorème découle d'une simple majoration de l'intégrale puisque

$$\sup_{t\in[0,1]} \|\varphi'(t)\|_F \leqslant \sup_{t\in[0,1]} \|\,\mathrm{d}f(tx + (1-t)y)\|_{\mathcal{L}(E,F)}\|x - y\|_E.$$

□

Remarque 2.6. *Il est possible, en modifiant la méthode de preuve, de relaxer l'hypothèse $f \in C^1(\Omega)$ en f différentiable en tout point de $[x, y]$.*

2.3 Applications

Théorème 2.7. *On suppose que Ω est un ouvert connexe de E. Si $f : \Omega \to F$ est différentiable en tout point et que $\mathrm{d}f(x) = 0$ pour tout $x \in \Omega$, alors f est constante sur Ω.*

Démonstration. Soit $x \in \Omega$ et soit $r > 0$ tel que $B(x, r) \subset \Omega$. Pour tout $y \in B(x, r)$ on a $[x, y] \subset B(x, r) \subset \Omega$ et le Théorème des accroissements finis entraîne alors que $f(y) = f(x)$. Ce fait assure que, pour un x_0 quelconque de Ω, l'ensemble $\{x \in \Omega,\ f(x) = f(x_0)\}$ est ouvert dans Ω. Comme il est également fermé (facile à vérifier) et que Ω est connexe, on en déduit qu'il est égal à Ω. Le résultat est démontré. □

Remarque 2.8. *La connexité est essentielle. Si Ω n'est pas connexe, les fonctions constantes sur chaque composante connexe ont bien une différentielle nulle partout mais ne sont pas nécessairement égale à une constante sur Ω tout entier.*

Si $E = \mathbb{R}^n$, il est facile de vérifier (le faire) que si $f : \Omega \to F$ est de classe C^1, alors ses dérivées partielles existent et sont continues sur Ω. On va montrer grâce au Théorème 2.5 que la réciproque est vraie, ce qui donne un moyen pratique de vérifier qu'une fonction est de classe C^1 dans le cas $E = \mathbb{R}^n$.

Théorème 2.9. *Soit $f : \Omega \subset \mathbb{R}^n \to F$. Si les dérivées partielles de f existent et sont continues sur Ω, alors f est de classe C^1 sur Ω.*

Démonstration dans le cas $n = 2$ (cas général par récurrence). Il s'agit de montrer que f est différentiable en tout $x \in \Omega$, de différentielle donnée par $\mathrm{d}f(x) = \displaystyle\sum_{i=1}^{2} \frac{\partial f}{\partial x_i}(x)\, \mathrm{d}x_i$. Pour cela, on écrit

$$\begin{aligned} f(x+h) - f(x) - h_1\frac{\partial f}{\partial x_1}(x) - h_2\frac{\partial f}{\partial x_2}(x) &= f(x_1+h_1, x_2+h_2) - f(x_1, x_2) - h_1\frac{\partial f}{\partial x_1}(x) - h_2\frac{\partial f}{\partial x_2}(x) \\ &= \underbrace{f(x_1+h_1, x_2+h_2) - f(x_1+h_1, x_2) - h_2\frac{\partial f}{\partial x_2}(x)}_{(I)} + \underbrace{f(x_1+h_1, x_2) - f(x_1, x_2) - h_1\frac{\partial f}{\partial x_1}(x)}_{(II)}. \end{aligned}$$

Par définition de la dérivée partielle par rapport à x_1 au point x, on a que

$$(II) = f(x_1 + h_1, x_2) - f(x_1, x_2) - h_1\frac{\partial f}{\partial x_1}(x) = o(h_1) = o(h).$$

Pour la partie (I), on utilise le Théorème 2.5. La fonction $g : t \mapsto f(x_1 + h_1, t) - (t - x_2)\frac{\partial f}{\partial x_2}(x)$ est de classe C^1 au voisinage de x_2, de différentielle donnée par $\mathrm{d}g(t)(k) = k\frac{\partial f}{\partial x_2}(x_1 + h_1, t) - k\frac{\partial f}{\partial x_2}(x)$. Le Théorème 2.5 appliqué à g sur le segment $[x_2, x_2 + h_2]$ assure alors que

$$\left\| f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) - h_2\frac{\partial f}{\partial x_2}(x) \right\|_F \leqslant \sup_{t\in[0,1]} \left\| \frac{\partial f}{\partial x_2}(x_1 + h_1, x_2 + th_2) - \frac{\partial f}{\partial x_2}(x) \right\|_F |h_2|.$$

Par continuité de $x \mapsto \frac{\partial f}{\partial x_2}(x)$, on a que

$$\sup_{t\in[0,1]} \left\| \frac{\partial f}{\partial x_2}(x_1 + h_1, x_2 + th_2) - \frac{\partial f}{\partial x_2}(x) \right\|_F \xrightarrow[h\to 0]{} 0,$$

et on en déduit donc que

$$(I) = f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) - h_2\frac{\partial f}{\partial x_2}(x) = o(h_2) = o(h),$$

ce qui finit la démonstration. $\square$

3 Différentielle seconde et optimisation

3.1 Différentielle seconde

Si $f : \Omega \to F$ est différentiable en tout point de Ω et si $\mathrm{d}f : \Omega \to \mathcal{L}(E, F)$ est différentiable en $x \in \Omega$, la différentielle $\mathrm{d}(\mathrm{d}f)(x)$ est une application linéaire de E dans $\mathcal{L}(E, F)$, c'est-à-dire $\mathrm{d}(\mathrm{d}f)(x) \in \mathcal{L}(E, \mathcal{L}(E, F))$.

Définition 3.1. *Soit $f : \Omega \to F$ différentiable en tout point de Ω. On dit que f est deux fois différentiable en $x \in \Omega$ si $\mathrm{d}f : \Omega \to \mathcal{L}(E, F)$ est différentiable en x. On appelle alors différentielle seconde de f en x l'application $\mathrm{d}^2 f(x)$ définie pour tous $h, k \in E$ par*

$$\mathrm{d}^2 f(x)(h, k) = [\mathrm{d}(\mathrm{d}f)(x)(h)](k).$$

On dit que f est de classe C^2 si $\mathrm{d}f$ est de classe C^1.

On vérifie facilement que $\mathrm{d}^2 f(x) : E \times E \to F$ est bilinéaire, ce que l'on note $\mathrm{d}^2 f(x) \in \mathcal{L}(E \times E, F)$. Une fonction f deux fois différentiable sur Ω est alors de classe C^2 si et seulement si $\mathrm{d}^2 f$ est continue de Ω dans $\mathcal{L}(E \times E, F)$. Dans le cas où $E = \mathbb{R}^n$, on peut caractériser les fonctions de classe C^2 à l'aide des dérivées partielles.

Théorème 3.2. *Soit $f : \Omega \subset \mathbb{R}^n \to F$ de classe C^1. Alors f est de classe C^2 si et seulement si ses dérivées partielles sont de classe C^1 sur Ω, et on a alors*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \mathrm{d}^2 f(x)(e_i, e_j).$$

Démonstration. D'une part on a, d'après le Théorème 2.9, que $\mathrm{d}f$ est de classe C^1 si est seulement si pour tout i la dérivée partielle $\frac{\partial}{\partial x_i}(\mathrm{d}f)$ existe et est continue de Ω dans $\mathcal{L}(\mathbb{R}^n, F)$. D'autre part on a, toujours d'après le Théorème 2.9, que $\frac{\partial f}{\partial x_j}$ est de classe C^1 si et seulement si pour tout i la dérivée partielle $\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}$ existe et est continue de Ω dans F.

Il suffit donc de vérifier, pour tout i, que $\left[\frac{\partial}{\partial x_i}(\mathrm{d}f)\right](x)$ existe si et seulement si $\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}(x)$ existe pour tout j, et qu'on a alors

$$\left[\frac{\partial}{\partial x_i}(\mathrm{d}f)\right](x) = \sum_{j=1}^{n} \frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}(x)\, \mathrm{d}x_j.$$

Supposons que $\left[\frac{\partial}{\partial x_i}(\mathrm{d}f)\right](x)$ existe. On a alors

$$\mathrm{d}f(x + te_i) = \mathrm{d}f(x) + t\left[\frac{\partial}{\partial x_i}(\mathrm{d}f)\right](x) + o(t)$$

dans $\mathcal{L}(\mathbb{R}^n, F)$, ou de manière équivalente

$$\mathrm{d}f(x + te_i)(h) = \mathrm{d}f(x)(h) + t\left[\frac{\partial}{\partial x_i}(\mathrm{d}f)\right](x)(h) + o(t)$$

dans F pour tout $h \in \mathbb{R}^n$. On a donc en particulier, pour tout j,

$$\mathrm{d}f(x + te_i)(e_j) = \mathrm{d}f(x)(e_j) + t\left[\frac{\partial}{\partial x_i}(\mathrm{d}f)\right](x)(e_j) + o(t),$$

qui s'écrit également

$$\frac{\partial f}{\partial x_j}(x + te_i) = \frac{\partial f}{\partial x_j}(x) + t\left[\frac{\partial}{\partial x_i}(\mathrm{d}f)\right](x)(e_j) + o(t).$$

On en déduit donc que $\frac{\partial f}{\partial x_j}$ admet une dérivée directionnelle dans la direction i en x, et que celle-ci est donnée par

$$\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}(x) = \left[\frac{\partial}{\partial x_i}(\mathrm{d}f)\right](x)(e_j).$$

Réciproquement, si $\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}(x)$ existe, on a

$$\frac{\partial f}{\partial x_j}(x+te_i) = \frac{\partial f}{\partial x_j}(x) + t\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}(x) + o(t),$$

qui s'écrit aussi

$$\mathrm{d}f(x+te_i)(e_j) = \mathrm{d}f(x)(e_j) + t\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}(x) + o(t),$$

et qui donne par linéarité

$$\mathrm{d}f(x+te_i)(h) = \mathrm{d}f(x)(h) + t\sum_{j=1}^{n}\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}(x)\,\mathrm{d}x_j(h) + o(t)$$

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

$$\mathrm{d}f(x+te_i) = \mathrm{d}f(x) + t\sum_{j=1}^{n}\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}(x)\,\mathrm{d}x_j + o(t).$$

En d'autres termes, $\mathrm{d}f$ admet une dérivée directionnelle dans la direction e_i, donnée par

$$\left[\frac{\partial}{\partial x_i}(\mathrm{d}f)\right](x) = \sum_{j=1}^{n}\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}(x)\,\mathrm{d}x_j.$$

□

Théorème 3.3 (Schwarz)**.** *Soit $f : \Omega \to F$ deux fois différentiable en $x \in \Omega$. Alors l'application $\mathrm{d}^2 f(x)$ est symétrique, c'est-à-dire que pour tous $h, k \in E$,*

$$\mathrm{d}^2 f(x)(h,k) = \mathrm{d}^2 f(x)(k,h).$$

Dans le cas où $\Omega \subset \mathbb{R}^n$, on a donc $\dfrac{\partial^2 f}{\partial x_i \partial x_j}(x) = \dfrac{\partial^2 f}{\partial x_j \partial x_i}(x)$ pour tous i, j si f est deux fois différentiable en x. Attention, on suppose bien ici que f est deux fois différentiable en x, et pas seulement que f admet des dérivées partielles secondes en x. Avant de prouver le théorème de Schwarz, on commence par un lemme.

Lemme 3.4. *Soit $f : \Omega \to F$ deux fois différentiable en $x \in \Omega$. Pour $h, k \in E$ suffisamment petits, on pose*

$$G(h,k) = f(x+h+k) - f(x+h) - f(x+k) + f(x).$$

Alors G vérifie

$$\frac{\|G(h,k) - \mathrm{d}^2 f(x)(h,k)\|_F}{\|h\|_E^2 + \|k\|_E^2} \xrightarrow[(h,k)\to(0,0)]{} 0.$$

Démonstration du Lemme 3.4. La fonction Q définie par $Q(h,k) = G(h,k) - \mathrm{d}^2 f(x)(h,k)$ est différentiable par rapport à sa deuxième variable (c'est-à-dire k). On note $\partial_2 Q$ cette différentielle partielle, qui est donnée par

$$\partial_2 Q(h,k)(\tilde{k}) = \mathrm{d}f(x+h+k)(\tilde{k}) - \mathrm{d}f(x+k)(\tilde{k}) - \mathrm{d}^2 f(x)(h,\tilde{k}).$$

Comme f est deux fois différentiable en x on a

$$\mathrm{d}f(x+h+k)(\tilde{k}) - \mathrm{d}f(x+k)(\tilde{k}) = \mathrm{d}^2 f(x)(h+k,\tilde{k}) - \mathrm{d}^2 f(x)(k,\tilde{k}) + \|h+k\|\varepsilon_1(h+k)\|\tilde{k}\| + \|k\|\varepsilon_2(k)\|\tilde{k}\|$$

et on en déduit par linéarité de $\mathrm{d}^2 f(x)$ par rapport à sa première variable que

$$\partial_2 Q(h,k)(\tilde{k}) = \big(\|h\| + \|k\|\big)\varepsilon_3(h,k)\|\tilde{k}\|.$$

L'inégalité des accroissements finis assure alors que

$$\|Q(h,k) - Q(h,0)\| \leqslant \sup_{t\in[0,1]} \|\partial_2 Q(h,tk)\|_{\mathcal{L}(E,F)}\|k\| \leqslant \big(\|h\| + \|k\|\big)\|k\|\|\varepsilon_3(h,k)\|.$$

Comme $Q(h,0) = 0$, on en déduit que

$$\frac{\|Q(h,k)\|}{\|h\|^2 + \|k\|^2} = \frac{\big(\|h\| + \|k\|\big)\|k\|\|\varepsilon_3(h,k)\|}{\|h\|^2 + \|k\|^2} \xrightarrow[(h,k)\to(0,0)]{} 0,$$

ce qui est le résultat demandé. □

Nous sommes maintenant armés pour démontrer le théorème de Schwarz.

Démonstration du Théorème 3.3. Soient $h, k \in E$ fixés. Pour $t > 0$ assez petit on a $G(th, tk) = G(tk, th)$ et donc, d'après le Lemme 3.4,

$$\| \mathrm{d}^2 f(x)(th, tk) - \mathrm{d}^2 f(x)(tk, th)\| \leqslant \| \mathrm{d}^2 f(x)(th, tk) - G(th, tk)\| + \|G(tk, th) - \mathrm{d}^2 f(x)(tk, th)\| = t^2 \varepsilon(t).$$

Par bilinéarité de $\mathrm{d}^2 f(x)$ on en déduit que

$$\| \mathrm{d}^2 f(x)(h, k) - \mathrm{d}^2 f(x)(k, h)\| \leqslant \varepsilon(t)$$

et on obtient le résultat en faisant tendre t vers 0. □

Définition 3.5 (Hessienne)**.** *Si $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ admet des dérivées partielles secondes en un point $x \in \Omega$, on appelle matrice hessienne de f en x la matrice*

$$\operatorname{Hess} f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix} = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right)_{1 \leqslant i,j \leqslant n}.$$

Si f est deux fois différentiable en x (par exemple si f est de classe C^2 sur Ω), on a alors l'expression

$$\mathrm{d}^2 f(x)(h, k) = \langle h, \operatorname{Hess} f(x) \cdot k \rangle.$$

De plus, le théorème de Schwarz assure que si f est deux fois différentiable en x, alors la Hessienne de f en x est une matrice symétrique. Comme elle est réelle, ses valeurs propres sont réelles et elle est diagonalisable dans une base orthonormée de $\mathbb{R}^n$.

3.2 Formule de Taylor-Young

Théorème 3.6 (Taylor-Young)**.** *Si $f : \Omega \to F$ est deux fois différentiable en $x \in \Omega$, alors on a le développement*

$$f(x + h) = f(x) + \mathrm{d}f(x)(h) + \frac{1}{2} \mathrm{d}^2 f(x)(h, h) + o(\|h\|^2).$$

Démonstration. La fonction F définie pour h suffisamment petit par

$$F(h) = f(x + h) - f(x) - \mathrm{d}f(x)(h) - \frac{1}{2} \mathrm{d}^2 f(x)(h, h)$$

est différentiable et, en utilisant le théorème de Schwarz,

$$\mathrm{d}F(h)(k) = \mathrm{d}f(x + h)(k) - \mathrm{d}f(x)(k) - \mathrm{d}^2 f(x)(h, k).$$

Par définition de la différentielle seconde de f on a alors que

$$\mathrm{d}F(h) = \mathrm{d}f(x + h) - \mathrm{d}f(x) - \mathrm{d}(\mathrm{d}f)(x)(h) = \|h\| \varepsilon(h).$$

On en déduit par l'inégalité des accroissements finis que

$$\|F(h) - F(0)\|_F \leqslant \sup_{t \in [0,1]} \| \mathrm{d}F(th)\|_{\mathcal{L}(E,F)} \|h\| \leqslant \|h\|^2 \|\varepsilon(h)\|_{\mathcal{L}(E,F)},$$

ce qui est le résultat recherché. □

Pour une fonction $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$, cette formule de Taylor-Young s'écrit également

$$f(x + h) = f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} \langle \operatorname{Hess} f(x) \cdot h, h \rangle + o(\|h\|^2).$$

On voit alors que le signe des valeurs propres de la Hessienne de f va jouer un rôle crucial dans la recherche d'optima de la fonction f.

3.3 Optimisation

Dans toute cette partie, Ω désigne un ouvert de $\mathbb{R}^n$ et $f : \Omega \to \mathbb{R}$ est une fonction au minimum continue sur Ω.

Définition 3.7. *On dit que $f \in C(\Omega, \mathbb{R})$ admet un minimum local (resp. maximum local) en $x_0 \in \Omega$ s'il existe $r > 0$ tel que*

$$\forall x \in B(x_0, r), \quad f(x_0) \leqslant f(x) \qquad (\textit{resp. } f(x_0) \geqslant f(x)).$$

Proposition 3.8 (Condition nécessaire du premier ordre)**.** *Soit $f \in C(\Omega, \mathbb{R})$. Si f admet un extremum local en $x_0 \in \Omega$) (i.e. un minimum ou maximum local) et si f est différentiable en x_0, alors* $\mathrm{d}f(x_0) = 0$.

Comme $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$, on peut écrire de manière équivalente $\operatorname{Jac} f(x_0) = 0$ ou encore $\nabla f(x_0) = 0$.

Démonstration. Faisons la preuve dans le cas d'un minimum. On sait que pour tout h suffisamment petit, $f(x_0 + h) \geqslant f(x_0)$. On en déduit, en utilisant la différentiablilité de f en x_0, que

$$0 \leqslant \langle \nabla f(x_0), h \rangle + \|h\| \varepsilon(h)$$

avec $\varepsilon(h) \to 0$ quand $h \to 0$. Fixons maintenant $v \in \mathbb{R}^n$ et considérons $h = tv$ pour $t > 0$ suffisamment petit. On obtient alors, en divisant par t puis en passant à la limite $t \to 0$, que

$$0 \leqslant \langle \nabla f(x_0), v \rangle.$$

En changeant v en $-v$ <(ou en prenant $t < 0$) on obtient l'inégalité inverse, et finalement on a donc $\langle \nabla f(x_0), v \rangle = 0$ pour tout $v \in \mathbb{R}^n$. En prenant $v = \nabla f(x_0)$ on en déduit que $\|\nabla f(x_0)\| = 0$, et donc $\nabla f(x_0) = 0$. ☐

Définition 3.9. *Un point $x_0 \in \Omega$ est appelé point critique de f si f est différentiable en x_0 et $\nabla f(x_0) = 0$.*

Définition 3.10 (Rappel)**.** *Une matrice symétrique réelle $A \in \mathcal{S}_n(\mathbb{R})$ est dite positive, et on note $A \geqslant 0$, si $\langle Ax, x \rangle \geqslant 0$ pour tout vecteur $x \in \mathbb{R}^n$. Elle est dite définie positive si $\langle Ax, x \rangle > 0$ pour tout vecteur $x \in \mathbb{R}^n$ non nul.*

Remarque 3.11. *On rappelle également qu'une matrice $A \in \mathcal{S}_n(\mathbb{R})$ est positive si et seulement si toutes ses valeurs propres sont positives ou nulles, et définie positive si et seulement si elles sont strictement positives.*

Proposition 3.12 (Condition nécessaire du second ordre)**.** *Soit $f \in C(\Omega, \mathbb{R})$. Si f admet un minimum local en $x_0 \in \Omega$ (resp. maximum local) et si f est deux fois différentiable en x_0, alors* $\operatorname{Hess} f(x_0) \geqslant 0$ *(resp.* $\operatorname{Hess} f(x_0) \leqslant 0$*).*

Démonstration. Comme $\nabla f(x_0) = 0$, le développement de Taylor-Young donne, pour un minimum local,

$$0 \leqslant \frac{1}{2} \langle \operatorname{Hess} f(x_0)\, h, h \rangle + \|h\|^2 \varepsilon(h).$$

En raisonnant comme dans la preuve de la condition du premier ordre, on en déduit que

$$0 \leqslant \langle \operatorname{Hess} f(x_0)\, v, v \rangle$$

pour tout $v \in \mathbb{R}^n$. Autrement dit $\operatorname{Hess} f(x_0) \geqslant 0$. ☐

Proposition 3.13 (Condition suffisante)**.** *Soit $f \in C(\Omega)$ et soit $x_0 \in \Omega$ un point en lequel f est deux fois différentiable. Si $\nabla f(x_0) = 0$ et* $\operatorname{Hess} f(x_0) > 0$ *(resp.* $\operatorname{Hess} f(x_0) < 0$*), alors f admet un minimum local en x_0 (resp. un maximum local).*

Démonstration. On part de la formule de Taylor-Young

$$f(x_0 + h) = f(x_0) + \langle \nabla f(x_0), h \rangle + \frac{1}{2} \langle \operatorname{Hess} f(x_0)\, h, h \rangle + \|h\|^2 \varepsilon(h).$$

Soit $\nu > 0$ la plus petite valeur propre de la Hessienne au point x_0. Alors on a

$$\langle \operatorname{Hess} f(x_0)\, h, h\rangle \geqslant \nu \|h\|^2$$

et, comme $\nabla f(x_0) = 0$, cela entraîne que

$$f(x_0 + h) \geqslant f(x_0) + \Big(\frac{\nu}{2} + \varepsilon(h)\Big)\|h\|^2.$$

Puisque $\varepsilon(h) \to 0$ quand $h \to 0$, il existe $r > 0$ tel que $\|h\| < r \implies \|\varepsilon(h)\| \leqslant \nu/2$. On a alors $f(x_0+h) \geqslant f(x_0)$ pour tout $h \in B(0, r)$, ce qui revient à dire que f admet un minimum local en x_0. $\square$

Remarque 3.14. *En dimension $n = 2$, il est très facile de connaître le signe des valeurs propres d'une matrice A à l'aide de son déterminant et de sa trace. En effet* $\det(A) = \lambda_1\lambda_2$ *et* $\operatorname{tr}(A) = \lambda_1 + \lambda_2$*. On a donc trois cas de figure :*

1. $\det(A) = 0$*. Alors* 0 *est valeur propre et la trace donne la deuxième valeur propre.*
2. $\det(A) > 0$*. Alors A a deux valeurs propres de même signe et la trace nous donne ce signe.*
3. $\det(A) < 0$*. Alors A a deux valeurs propres de signes opposés.*

Dans le cas où $\nabla f(x_0) = 0$ et $\operatorname{Hess} f(x_0)$ *a deux valeurs propres de signes opposés, on dit que f admet un point selle en x_0.*

4 Inversion locale et application

4.1 Homéomorphismes et difféomorphismes

Définition 4.1. *Soient $A \subset E$ et $A' \subset F$ deux ensembles. On dit que $f : A \to A'$ est un homéomorphisme si f est continue, bijective, et que son application réciproque f^{-1} est continue (on dit que f est bicontinue).*

Lemme 4.2. *Soit $f : K \to K'$ une application bijective et continue sur un compact K. Alors f est un homéomorphisme de K sur K' (qui est compact).*

Démonstration. Le fait que K' soit compact résulte du fait que $K' = f(K)$ avec f continue et K compact.

Maintenant, soit $(y_n) \subset K'$ une suite telle que $y_n \to \ell \in K'$. On veut montrer que $f^{-1}(y_n) \to f^{-1}(\ell')$. On pose $x_n = f^{-1}(y_n) \in K$ et $\ell = f^{-1}(\ell') \in K$, c'est-à-dire $y_n = f(x_n)$ et $\ell' = f(\ell)$. Comme K est compact, on peut extraire de (x_n) une suite sous-suite $x_{\varphi(n)} \to x \in K$. Comme f est continue, $f(x_{\varphi(n)}) \to f(x)$ et donc $f(x) = f(\ell)$, ce qui entraîne que $x = \ell$ puisque f est bijective. La seule valeur d'adhérence de la suite (x_n) est donc ℓ et cela assure que c'est toute la suite (x_n) qui converge vers ℓ. On a donc montré que $f^{-1}(y_n) \to \ell = f^{-1}(\ell')$. La fonction f^{-1} est donc continue et f est un homéomorphisme. □

Définition 4.3. *Soient $\Omega \subset E$ et $\Omega' \subset F$ deux ouverts. On dit que $f : \Omega \to \Omega'$ est un difféomorphisme de classe C^1, ou C^1-difféomorphisme, si f est un homéomorphisme et de plus f et f^{-1} sont de classe C^1.*

Le lemme suivant va nous montrer que s'il existe un difféomorphisme de $\Omega \subset \mathbb{R}^n$ dans $\Omega' \subset \mathbb{R}^p$, alors nécessairement $n = p$.

Lemme 4.4. *Soient $\Omega \subset E$ et $\Omega' \subset F$ deux ouverts, $f : \Omega \to \Omega'$ un homéomorphisme, et $x \in \Omega$, $y \in \Omega'$ deux points tels que $y = f(x)$ et $\mathrm{d}f(x)$ et $\mathrm{d}f^{-1}(y)$ existent. Alors $\mathrm{d}f(x)$ est inversible et on a*

$$\mathrm{d}f^{-1}(y) = \big[\mathrm{d}f(x)\big]^{-1} = \big[\mathrm{d}f(f^{-1}(y))\big]^{-1}.$$

En particulier, $\dim E = \dim F$.

Démonstration. Il suffit d'écrire la différentielle de $f \circ f^{-1} = \mathrm{Id}$ en y qui donne

$$\mathrm{d}f(f^{-1}(y)) \circ \mathrm{d}f^{-1}(y) = \mathrm{Id},$$

ainsi que la différentielle de $f^{-1} \circ f = \mathrm{Id}$ en x qui donne

$$\mathrm{d}f^{-1}(f(x)) \circ \mathrm{d}f(x) = \mathrm{Id}.$$

Évidemment, $\mathrm{d}f(x)$ étant une application linéaire inversible de E dans F, on a forcément $\dim E = \dim F$. □

Corollaire 4.5. *Si $f : \Omega \to \Omega'$ est un homéomorphisme de classe C^1 dans Ω tel que f^{-1} est différentiable en tout point de Ω', alors f est un difféomorphisme de classe C^1.*

Démonstration. Il n'y a que la continuité de $\mathrm{d}f^{-1}$ à vérifier, ce qui découle du lemme puisqu'on a alors $\mathrm{d}f^{-1}(y) = \big[\mathrm{d}f(f^{-1}(y))\big]^{-1}$. □

4.2 Le théorème d'inversion locale

Théorème 4.6 (Inversion locale)**.** *Soient $f : \Omega \to F$ une fonction de classe C^1 et $x_0 \in \Omega$. Si $\mathrm{d}f(x_0)$ est inversible, alors il existe deux voisinages $U \subset \Omega$ et $V \subset F$ de x_0 et $f(x_0)$ respectivement tels que f est un C^1-difféomorphisme de U sur V.*

Démonstration. On procède en plusieurs étapes.

Étape 1 : *Réduction du problème.*
On se ramène au cas où $x_0 = 0$, $f : \Omega \subset E \to E$, $f(x_0) = 0$ et $\mathrm{d}f(x_0) = \mathrm{Id}$ en introduisant la fonction

$$g(x) = [\mathrm{d}f(x_0)]^{-1}(f(x + x_0) - f(x_0)).$$

On voit facilement que g envoie des éléments de E dans E, que $g(0) = 0$ et que $\mathrm{d}g(0) = \mathrm{Id}$ (utiliser soit la définition soit le théorème de dérivation composée). De plus

$$y = f(x) \quad \Longleftrightarrow \quad [\mathrm{d}f(x_0)]^{-1}(y - f(x_0)) = g(x - x_0).$$

On voit donc que résoudre l'équation $y = f(x)$ pour y proche de $f(x_0)$, en cherchant x proche de x_0, revient à résoudre l'équation $\tilde{y} = g(\tilde{x})$ pour $\tilde{y}$ proche de 0. Si on prouve que g est un C^1-difféomorphisme de $\widetilde{U} \subset E$ sur $\widetilde{V} \subset E$, alors on aura immédiatement que f est un C^1-difféomorphisme de $U = \widetilde{U} + \{x_0\} \subset E$ sur $V = \mathrm{d}f(x_0)(\widetilde{V}) + \{f(x_0)\} \subset F$. Donc f est inversible et $f^{-1} = h$ est de classe C^1 puisque g^{-1} l'est.

On prouve maintenant le théorème dans ce nouveau cadre en revenant à la notation f (à la place de g).

Étape 2 : *Résolution de l'équation $y = f(x)$.*
On va utiliser le théorème des applications contractantes (Banach-Picard) via le lemme suivant.

Lemme 4.7. *Soient $r > 0$ et $T : \overline{B(0,r)} \subset E \to E$ une application telle que $\|T(0)\| \leqslant \frac{r}{2}$ et*

$$\|T(x) - T(x')\| \leqslant \frac{1}{2}\|x - x'\| \qquad \text{pour tous} \quad x, x' \in \overline{B(0,r)}.$$

Alors T admet un unique point fixe dans $\overline{B(0,r)}$.

La preuve de ce lemme est une conséquence immédiate du théorème de point fixe de Banach-Picard dès lors qu'on a remarqué que, grâce aux hypothèses, T envoie $\overline{B(0,r)}$ dans $\overline{B(0,r)}$.

On souhaite appliquer ce lemme à

$$T(x) = x - (f(x) - y),$$

où y est un point fixé dans une voisinage de 0 à déterminer. Si on y parvient, on aura bien résolu notre équation puisque

$$y = f(x) \qquad \Longleftrightarrow \qquad T(x) = x.$$

Pour vérifier les hypothèses du lemme, on remarque d'abord que

$$\mathrm{d}T(0) = \mathrm{Id} - \mathrm{d}f(0) = 0,$$

et comme $\mathrm{d}T$ est continue puisque f est de classe C^1, il existe $r > 0$ tel que $\|\mathrm{d}T\|_{\mathcal{L}(E)} \leqslant 1/2$ dans la boule $\overline{B(0,r)}$. En appliquant l'inégalité des accroissements finis à T entre deux points $x, x' \in \overline{B(0,r)}$, on a bien la propriété de contraction demandée sur T. Il reste à calculer

$$T(0) = 0 - (f(0) - y) = y$$

et donc si $\|y\| \leqslant r/2$ le lemme s'applique. En d'autres termes, pour tout $y \in \overline{B(0,r/2)}$ il existe une unique solution x dans $\overline{B(0,r)}$ à l'équation $y = f(x)$.

On remarque qu'en changeant r en $r/2$ on obtient que pour tout $y \in B(0,r/4)$ l'unique solution x à l'équation $y = f(x)$ appartient à $B(0,r)$.

Une autre remarque importante pour la suite est que pour tout $x \in \overline{B(0,r)}$ on a $\mathrm{d}f(x) = \mathrm{Id} - \mathrm{d}T(x)$ avec $\|\mathrm{d}T(x)\|_{\mathcal{L}(E)} \leqslant 1/2$, ce qui assure que $\mathrm{d}f(x)$ est inversible.

Étape 3 : *Continuité de la solution x par rapport à y.*
On veut prouver que x dépend continûment de y. On raisonne comme dans la preuve du Lemme 4.2. Soit $y \in \overline{B(0,r/2)}$ et soit $(y_n) \subset \overline{B(0,r/2)}$ une suite qui converge vers y. La suite (x_n) des solutions de $y_n = f(x_n)$ vit dans l'ensemble compact $\overline{B(0,r)}$ et a pour seule valeur d'adhérence possible la solution x de $y = f(x)$ (en vertu de la continuité de f). La suite (x_n) converge donc vers x.

Étape 4 : *Définition et différentiabilité de l'application réciproque.*
Pour raisonner avec des ouverts, on pose $V = B(0,r/4)$ et $U = f^{-1}(V) \cap B(0,r)$. Par continuité de f, l'ensemble U est bien un ouvert. Clairement, f est une bijection de U sur V en vertu de l'étape 2.

Pour prouver que f^{-1} est différentiable sur V, on utilise le fait que si f^{-1} est différentiable en $y \in V$ alors sa différentielle est $\mathrm{d}f^{-1}(y) = [\mathrm{d}f(x)]^{-1}$ où $x = f^{-1}(y)$. On s'intéresse alors à la quantité

$$A(h) = f^{-1}(y + h) - f^{-1}(y) - [\mathrm{d}f(x)]^{-1}(h)$$

et on veut montrer que $A(h) = o(h)$. Pour h suffisamment petit, il existe $k \in E$ tel que $y + h = f(x + k)$. Autrement dit, on considère $k = f^{-1}(y+h) - x = f^{-1}(y+h) - f^{-1}(y)$. Par l'étape 3, on a que $k \to 0$ quand $h \to 0$. Par différentiabilité de f en x on a par ailleurs

$$h = f(x + k) - f(x) = \mathrm{d}f(x)(k) + \|k\|\varepsilon(k).$$

On en déduit d'une part que

$$A(h) = (x + k) - x - [\mathrm{d}f(x)]^{-1}\big(df(x)(k) + \|k\|\varepsilon(k)\big) = -\|k\|[\mathrm{d}f(x)]^{-1}(\varepsilon(k))$$

et d'autre part que

$$k = [\mathrm{d}f(x)]^{-1}(h) + \|k\|[\mathrm{d}f(x)]^{-1}(\varepsilon(k)).$$

De la deuxième égalité on déduit que

$$\|k\| \leqslant \|[\mathrm{d}f(x)]^{-1}\|_{\mathcal{L}(F,E)}\big(\|h\| + \|k\|\,\|\varepsilon(k)\|\big).$$

En prenant h suffisamment petit pour que $\|\varepsilon(k)\| \leqslant \frac{1}{2}\|[\mathrm{d}f(x)]^{-1}\|_{\mathcal{L}(F,E)}$ on obtient que

$$\|k\| \leqslant 2\|[\mathrm{d}f(x)]^{-1}\|_{\mathcal{L}(F,E)}\|h\|.$$

En injectant dans la première égalité on obtient que

$$\|A(h)\| \leqslant 2\|[\mathrm{d}f(x)]^{-1}\|^2_{\mathcal{L}(F,E)}\|h\|\|\varepsilon(k)\|,$$

et donc on a bien $A(h) = o(h)$ puisque $\varepsilon(k) \to 0$ quand $h \to 0$.

Étape 5 : *f est un C^1-difféomorphisme de U sur $V = f(U)$.*
Il reste juste à vérifier que $y \mapsto \mathrm{d}f^{-1}(y)$ est continue, ce qui est évident puisqu'on a montré à l'étape 4 que $\mathrm{d}f^{-1}(y) = \big[\mathrm{d}f(f^{-1}(y))\big]^{-1}$. □

Théorème 4.8 (Inversion globale)**.** *Soient $f : \Omega \to F$ une fonction de classe C^1 qui est injective et telle que $\mathrm{d}f(x)$ est inversible pour tout $x \in \Omega$. Alors f est un C^1-difféomorphisme de Ω sur $f(\Omega)$.*

Démonstration. La fonction f étant injective et surjective sur $f(\Omega)$, elle est bijective de Ω sur $f(\Omega)$. Comme elle est aussi de classe C^1, il reste juste à vérifier que f^{-1} est aussi de classe C^1.

Pour $y \in f(\Omega)$ on pose $x = f^{-1}(y)$ et on applique le théorème d'inversion locale : comme $\mathrm{d}f(x)$ est inversible, il existe deux voisinages $U \subset \Omega$ de x et $V \subset F$ de y tels que f est un C^1-difféomorphisme de U sur V. Le difféomorphisme réciproque (de V sur U) coïncide bien sûr avec f^{-1} qui est donc, en particulier, différentiable en y. □

4.3 Le théorème des fonctions implicites

Théorème 4.9 (Fonctions implicites)**.** *Soit $f : \Omega \subset E \times F \to G$ une fonction de classe C^1 et soit $(x_0, y_0) \in \Omega$ tel que $f(x_0, y_0) = 0$ et la différentielle partielle $\partial_2 f(x_0, y_0)$ est inversible. Alors il existe un voisinage U de (x_0, y_0) et une fonction g définie sur un voisinage V de x_0 tels que*

$$(x, y) \in U,\ f(x, y) = 0 \quad\Longleftrightarrow\quad x \in V,\ y = g(x).$$

De plus g est de classe C^1 sur V et

$$\mathrm{d}g(x) = -\big[\partial_2 f(x, g(x))\big]^{-1} \circ \partial_1 f(x, g(x)).$$

Démonstration. Soit $h : \Omega \to E \times G$ définie par $h(x, y) = (x, f(x, y))$. Alors $h(x_0, y_0) = (x_0, 0)$ et

$$\mathrm{d}h(x, y) = \begin{pmatrix} \mathrm{Id}_E & 0 \\ \partial_1 f(x, y) & \partial_2 f(x, y) \end{pmatrix}.$$

En particulier, $\mathrm{d}h(x_0, y_0)$ est inversible et d'après le théorème d'inversion locale il existe un voisinage U de (x_0, y_0) dans Ω, un voisinage $W = h(U)$ de $h(x_0, y_0) = (x_0, 0)$ tels que h est un C^1-difféomorphisme de U sur

W. Donc $h^{-1} = (h_1^{-1}, h_2^{-1})$ est une fonction de classe C^1 sur W et on a $h^{-1}(x, f(x,y)) = h^{-1} \circ h(x,y) = (x,y)$. On a alors

$$(x,y) \in U,\ f(x,y) = 0 \iff (x,y) = h^{-1}(x,0) \iff y = h_2^{-1}(x,0),$$

ce qui définit une fonction g de classe C^1 sur un voisinage $V = \{x \in E,\ (x,0) \in W\}$ de x_0 par $g(x) = h_2^{-1}(x,0)$.

Ensuite, pour tout $x \in V$ on a $(x, g(x)) \in U$ et $f(x, g(x)) = 0$. En différenciant par rapport à x il vient

$$\partial_1 f(x, g(x)) + \partial_2 f(x, g(x)) \circ \mathrm{d}g(x) = 0.$$

Comme d'autre part $\mathrm{d}h$ est inversible sur U, la différentielle partielle $\partial_2 f$ est également inversible sur U et on obtient le résultat recherché en composant par $\left[\partial_2 f(x, g(x))\right]^{-1}$. $\square$