CHAPITRE 3

APPLICATIONS DIFFÉRENTIABLES

7. Définitions et premières propriétés

[1] En guise d'introduction, commençons par un rappel sur les fonctions dérivables $\mathbb{R} \to \mathbb{R}$. Soient $a \in \mathbb{R}$ et $f : \mathbb{R} \to \mathbb{R}$ une fonction définie sur un intervalle ouvert I contenant a. On dit que f est *dérivable* en a si la limite

$$\lim_{\substack{h\to 0\\ h\neq 0}} \frac{f(a+h)-f(a)}{h}$$

existe ; dans ce cas elle est notée $f'(a)$ et l'on a $f(a+h) = f(a) + f'(a)h + h\varepsilon(h)$, où $\varepsilon : \mathbb{R} \to \mathbb{R}$ est une fonction *continue et nulle* en 0 (i.e. $\lim_{h\to 0} \varepsilon(h) = 0 = \varepsilon(0)$).

Dans ce cas, connaissant $f(a)$ on peut obtenir une assez bonne approximation de $f(x)$ au voisinage de a, i.e. de $f(a+h)$ pour h assez petit, en remplaçant $f(a+h)$ par la fonction « linéaire » $h \mapsto f(a)+f'(a)h$. Stricto sensu, c'est une application affine, mais comme la valeur en 0 est donnée par $f(a)$, « ce qui compte » est l'application linéaire $h \mapsto f(a+h) - f(a) = f'(a)h$. En résumé, on considère que $f : \mathbb{R} \to \mathbb{R}$ est une « bonne » fonction au voisinage d'un point a si elle possède une « *bonne approximation* » par une application linéaire $L_a : h \mapsto L_a(h)$, la notion de « bonne approximation » ayant le sens précis que la différence

$$f(a+h) - f(a) - L_a(h)$$

tend vers 0 plus vite que $|h|$, i.e. que $f(a+h) = f(a)+L_a(h)+|h|\varepsilon(h)$, où $\varepsilon : \mathbb{R} \to \mathbb{R}$ est une fonction *continue et nulle* en 0. C'est sous cette forme que la notion de fonction dérivable va s'étendre aux fonctions $\mathbb{R}^n \to \mathbb{R}^p$.

Notations 7.1. — Dans la suite, on munit $\mathbb{R}^n$ et $\mathbb{R}^p$ de la norme $\|\cdot\|_\infty$ (on pourrait aussi choisir la norme euclidienne $\|\cdot\|_2$).

On note $\mathscr{L}(\mathbb{R}^n,\mathbb{R}^p)$ l'espace vectoriel (de dimension np) des applications linéaires $\mathbb{R}^n \to \mathbb{R}^p$. On rappelle qu'une telle application est lipschitzienne donc continue. D'autre part, $\mathscr{L}(\mathbb{R}^n,\mathbb{R}^p)$ s'identifie à $M_{p,n}(\mathbb{R})$, l'espace de matrices à p lignes et n colonnes.

En particulier, si $n = 1$ alors $\mathscr{L}(\mathbb{R},\mathbb{R}^p) = \mathbb{R}^p$ car une application linéaire $f : \mathbb{R} \to \mathbb{R}^p$ est déterminée par la donnée du vecteur $u = f(1)$, i.e. on a $f(t) = tu$ pour tout $t \in \mathbb{R}$. En termes

[1] v. du 2/12/17 : correction de coquilles dans la démo de 9.2 et la déf. 10.1, signalées par Johan Leydet, Tran Trung Nghiem et Nassim Bourarach, merci à eux !

matriciels, on a $\mathscr{L}(\mathbb{R}, \mathbb{R}^p) = M_{p,1}(\mathbb{R})$, ce qui nous conduira dans la suite à représenter un élément $(y_1, \dots, y_p)$ de $\mathbb{R}^p$ par le vecteur **colonne** $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix}$.

Définition 7.2 **(Limites en** 0**).** — Soit $\varphi : \mathbb{R}^n \to \mathbb{R}^p$ une fonction définie sur une boule ouverte $B(0, R)$, sauf en 0, et soit $b \in \mathbb{R}^p$. On écrit

$$\lim_{\substack{h \to 0 \\ h \neq 0}} \varphi(h) = b$$

si pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que pour tout $h \neq 0$ vérifiant $\|h\| < \delta$ on ait $\|\varphi(h) - b\| < \varepsilon$. Dans ce cas, si l'on prolonge φ en 0 en posant $\varphi(0) = b$, on obtient une fonction qui est définie sur $B(0, R)$ et continue en 0. Si de plus $b = 0$, on dira que la fonction ainsi prolongée est « continue et nulle en 0 ».

D'autre part, si l'on note $\varphi_1, \dots, \varphi_p$ les composantes de φ et $b_1, \dots, b_p$ celles de b, alors comme $\|\varphi(h) - b\| = \max_{i=1,\dots,p} |\varphi_i(h) - b_i|$, on voit que la condition précédente équivaut à dire que, pour tout $i = 1, \dots, p$,

$$\lim_{\substack{h \to 0 \\ h \neq 0}} \varphi_i(h) = b_i.$$

(Q) ***Définition 7.3*** **(Fonctions dérivables** $\mathbb{R} \to \mathbb{R}^p$**).** — Soient $a \in \mathbb{R}$ et $f : \mathbb{R} \to \mathbb{R}^p$ une fonction définie sur un intervalle ouvert I contenant a. On « rappelle » que f est **dérivable** en a si la limite

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

existe dans $\mathbb{R}^p$; dans ce cas elle est notée $f'(a)$ et l'on a $f(a+h) = f(a) + hf'(a) + h\varepsilon(h)$, où $\varepsilon : \mathbb{R} \to \mathbb{R}^p$ est une fonction *continue et nulle* en 0 (i.e. $\lim_{h \to 0} \varepsilon(h) = 0 = \varepsilon(0)$).

Notant $f_1, \dots, f_p$ les composantes de f et $\ell_1, \dots, \ell_p$ celles de ℓ, on voit que la condition précédente équivaut à dire que, pour tout $i = 1, \dots, p$,

$$\lim_{\substack{h \to 0 \\ h \neq 0}} \frac{f_i(a+h) - f_i(a)}{h} = \ell_i$$

i.e. que f_i est dérivable en a de dérivée $f_i'(a) = \ell_i$. On obtient donc que f est dérivable en a ssi les f_i le sont, et dans ce cas le vecteur dérivé $f'(a)$ est

$$f'(a) = \begin{pmatrix} f_1'(a) \\ \vdots \\ f_p'(a) \end{pmatrix}.$$

Notant $L_a(h)$ l'application linéaire $\mathbb{R} \to \mathbb{R}^p$, $h \mapsto hf'(a)$, on obtient donc que

$$f(a+h) - f(a) - L_a(h) = h\varepsilon(h),$$

où $\varepsilon : \mathbb{R} \to \mathbb{R}^p$ est une fonction *continue et nulle* en 0. C'est sous cette forme que la notion de fonction dérivable s'étend aux fonctions $\mathbb{R}^n \to \mathbb{R}^p$, avec $n > 1$.

Commençons par le lemme suivant.

Lemme 7.4. — *Soit* $L : \mathbb{R}^n \to \mathbb{R}^p$ *une application linéaire. Si* $\lim_{\substack{h \to 0 \\ h \neq 0}} \dfrac{L(h)}{\|h\|} = 0$ *alors* $L = 0$.

Démonstration. — [2] Fixons $x \neq 0$ dans $\mathbb{R}^n$. Lorsque le réel $t > 0$ tend vers 0, le vecteur tx tend vers 0 et donc, d'après l'hypothèse, le vecteur $L(tx)/\|tx\|$ tend vers 0. Or ce vecteur est le vecteur constant $L(x)/\|x\|$, d'où $L(x) = 0$. □

Terminologie 7.5. — Soient U un ouvert de $\mathbb{R}^n$, f une application $U \to \mathbb{R}^p$ et $a \in U$. Comme U est ouvert, il existe $r > 0$ tel que $B(a,r) \subset U$ et donc la fonction $h \mapsto f(a+h) - f(a)$ est définie pour $\|h\| < r$. On utilisera ceci de façon implicite dans la suite quand on dira que certaines fonctions de h sont définies pour h « assez petit ».

(Q) ***Définition 7.6.*** — Soient U un ouvert de $\mathbb{R}^n$ et f une application $U \to \mathbb{R}^p$.

(i) On dit que f est **différentiable** en un point a de U s'il existe une application linéaire $L_a \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^p)$ telle que

$$(*) \qquad \lim_{\substack{h\to 0\\ h\neq 0}} \frac{f(a+h) - f(a) - L_a(h)}{\|h\|} = 0.$$

D'après le lemme précédent, L_a est uniquement déterminée si elle existe. Dans ce cas, L_a est notée $df(a)$ et est appelée *différentielle* ou *application linéaire tangente* de f en a.

(ii) En posant $\varepsilon(h) = \dfrac{f(a+h) - f(a) - L_a(h)}{\|h\|}$ pour $h \neq 0$ et $\varepsilon(0) = 0$, $(*)$ équivaut à dire que pour h assez petit on a :

$$(\dagger) \qquad f(a+h) = f(a) + df(a)(h) + \|h\|\,\varepsilon(h),$$

avec ε continue et nulle en 0. On peut récrire cette égalité avec la notation $o()$:

$$(\ddagger) \qquad f(a+h) = f(a) + df(a)(h) + o(\|h\|),$$

où $o(\|h\|)$ désigne une fonction $\phi(h)$ telle que $\lim_{\substack{h\to 0\\ h\neq 0}} \dfrac{\phi(h)}{\|h\|} = 0$. Avec l'une ou l'autre notation, ceci montre que si f est différentiable en a, elle admet en a un « développement de Taylor » à l'ordre 1, dont le terme linéaire est $df(a)(h)$.

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

(Q) ***Exemples 7.7.*** — (1) Soit $f : \mathbb{R}^n \to \mathbb{R}^p$ linéaire. Alors f est différentiable sur $\mathbb{R}^n$ et pour tout $a \in \mathbb{R}^n$ on a $df(a) = f$. En effet, pour tout $h \in \mathbb{R}^n$ on a $f(a+h) - f(a) = f(h)$.

(2) L'application $Q : \mathbb{R}^n \to \mathbb{R}$, $x \mapsto x \cdot x = (\|x\|_2)^2$ est différentiable sur $\mathbb{R}^n$. En effet, pour tout $a, h \in \mathbb{R}^n$ on a

$$Q(a+h) = Q(a) + a\cdot h + h\cdot a + Q(h) = Q(a) + 2a\cdot h + (\|h\|_2)^2$$

donc, notant $L(a) : \mathbb{R}^n \to \mathbb{R}$ la forme linéaire $h \mapsto 2a\cdot h$, on a $Q(a+h) = Q(a) + L(a)(h) + o(\|h\|_2)$, ce qui prouve que Q est différentiable en a, de différentielle $dQ(a) = L(a)$.

(3) L'application de multiplication $m : \mathbb{R}^2 \to \mathbb{R}$, $(x_1, x_2) \mapsto x_1x_2$ est différentiable sur $\mathbb{R}^2$. En effet, pour $x = (x_1, x_2)$ et $h = (h_1, h_2)$ dans $\mathbb{R}^2$, on a

$$(x_1 + h_1)(x_2 + h_2) = x_1x_2 + (x_2h_1 + x_1h_2) + h_1h_2$$

et $|h_1h_2| \leq \|h\|_\infty^2$, donc $dm(x)$ est la forme linéaire $\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \mapsto x_2h_1 + x_1h_2$, i.e. $dm(x)$ est donnée par la matrice ligne (x_2, x_1).

[2] Indiquée par Laurent Koelblen.

(4) Si $n = 1$, i.e. si I est un intervalle ouvert de $\mathbb{R}$ et $f = (f_1, \dots, f_p)$ [3] une application de I dans $\mathbb{R}^p$, alors f est différentiable en un point a de I ssi f est dérivable en a (i.e. chaque f_i l'est) et dans ce cas pour tout $h \in \mathbb{R}$ on a :

$$df(a)(h) = hf'(a) = h \begin{pmatrix} f_1'(a) \\ \vdots \\ f_p'(a) \end{pmatrix}.$$

En effet, on a vu en 7.3 que si f est dérivable en a elle y est différentiable et $df(a)$ est comme indiqué.

Réciproquement, supposons f différentiable en a. Comme tout $L \in \mathscr{L}(\mathbb{R}, \mathbb{R}^p)$ est de la forme $h \mapsto hu$ pour un certain $u \in \mathbb{R}^p$, ceci signifie qu'il existe $v \in \mathbb{R}^p$ tel que pour h assez petit on ait $f(a+h) - f(a) - hv = |h|\,\varepsilon(h)$ avec ε continue et nulle en 0, d'où

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

pour $h \neq 0$, et donc

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

Ceci montre que f est dérivable en a et $v = f'(a)$, et donc que $df(a)(h) = hv = hf'(a)$.

Remarque 7.8. — Si f est différentiable en a elle est continue en a, car $df(a)$ est continue (étant linéaire) et ε est continue en 0.

L'exemple (4) ci-dessus se généralise comme suit :

(Q) ***Proposition 7.9***. — *Soient U un ouvert de $\mathbb{R}^n$, $f = (f_1, \dots, f_p)$ [3] une application $U \to \mathbb{R}^p$ et $a \in U$. Alors f est différentiable en a ssi chaque f_i l'est, et dans ce cas on a*

$$df(a)(h) = \begin{pmatrix} df_1(a)(h) \\ \vdots \\ df_p(a)(h) \end{pmatrix}$$

pour tout $h \in \mathbb{R}^n$.

Démonstration. — Se donner une application linéaire $L : \mathbb{R}^n \to \mathbb{R}^p$ est « la même chose » que se donner p formes linéaires $L_i : \mathbb{R}^n \to \mathbb{R}$, i.e. on a

$$L(h) = \begin{pmatrix} L_1(h) \\ \vdots \\ L_p(h) \end{pmatrix}$$

pour tout $h \in \mathbb{R}^n$. (Du point de vue matriciel, on a $L \in M_{p,n}(\mathbb{R})$ et les L_i correspondent aux p lignes de cette matrice.) De même, la fonction $\varepsilon : \mathbb{R}^n \to \mathbb{R}^p$ qui apparaît dans la définition 7.6 s'écrit

$$\varepsilon(h) = \begin{pmatrix} \varepsilon_1(h) \\ \vdots \\ \varepsilon_p(h) \end{pmatrix}$$

[3] On continue à écrire $f = (f_1, \dots, f_p)$ pour des raisons typographiques, mais on pense à $f(x)$ comme au vecteur *colonne* dont les coordonnées sont les $f_i(x)$.

et la condition que ε soit continue et nulle en 0 équivaut à dire que chaque ε_i l'est. On voit donc que f est différentiable en a ssi il existe des formes linéaires $L_1, \dots, L_p$ et des fonctions $\varepsilon_1, \dots, \varepsilon_p : \mathbb{R}^n \to \mathbb{R}$ continues et nulles en 0 telles que

$$f_i(a+h) = f_i(a) + L_i(a)(h) + \|h\|\, \varepsilon_i(h),$$

ce qui équivaut à dire que chaque f_i est différentiable en a et $L_i = df_i(a)$, et dans ce cas pour tout $h \in \mathbb{R}^n$ on a bien

$$df(a)(h) = \begin{pmatrix} L_1(h) \\ \vdots \\ L_p(h) \end{pmatrix} = \begin{pmatrix} df_1(a)(h) \\ \vdots \\ df_p(a)(h) \end{pmatrix}.$$

□

Avant de démontrer le théorème sur la composée d'applications différentiables, introduisons les dérivées partielles et la matrice jacobienne, qui donneront un aspect plus concret à la notion de différentielle.

(Q) ***Définition 7.10* (Dérivée selon la direction v).** — Soit U un ouvert de $\mathbb{R}^n$, f une application $U \to \mathbb{R}$ et $a \in U$. Soit $v \in \mathbb{R}^n$ non nul. On dit que f admet en a une *dérivée partielle dans la direction v* si la fonction $\mathbb{R} \to \mathbb{R}$, $t \mapsto f(a+tv)$ est dérivable en $t = 0$, i.e. si la limite

$$\lim_{\substack{t \to 0 \\ t \neq 0}} \frac{f(a+tv) - f(a)}{t}$$

existe, auquel cas elle est notée $\dfrac{\partial f}{\partial v}(a)$.

Remarque 7.11. — Intuitivement, la dérivée directionnelle $\dfrac{\partial f}{\partial v}(a)$ mesure les variation de f lorsqu'on se déplace au voisinage de a dans la direction v à la vitesse $\|v\|$. **Attention :** la terminologie est légèrement abusive, car cette dérivée dépend de v lui-même, et pas seulement de la direction $\mathbb{R}v$. En effet, si on remplace v par un multiple non nul $w = \lambda v$, alors

$$\frac{\partial f}{\partial w}(a) = \lim_{\substack{t \to 0 \\ t \neq 0}} \frac{f(a+\lambda t v) - f(a)}{t} = \lambda \lim_{\substack{\lambda t \to 0 \\ t \neq 0}} \frac{f(a+\lambda t v) - f(a)}{\lambda t} = \lambda \frac{\partial f}{\partial v}(a).$$

Ceci explique la remarque « intuitive » plus haut (en prenant pour « unité de vitesse » celle donnée par le vecteur unitaire $u = \dfrac{1}{\|v\|} v$).

Conservant les notations précédentes, on a en particulier :

(Q) ***Définition 7.12* (Dérivées partielles).** — Soit $(e_1, \dots, e_n)$ la base canonique de $\mathbb{R}^n$. On note, si elle existe, $\dfrac{\partial f}{\partial x_i}(a)$ ou simplement $\partial_i f(a)$ la dérivée partielle de f en a dans la direction e_i, i.e.

$$\partial_i f(a) = \frac{\partial f}{\partial x_i}(a) = \lim_{\substack{t \to 0 \\ t \neq 0}} \frac{f(a+te_i) - f(a)}{t} = \lim_{\substack{t \to 0 \\ t \neq 0}} \frac{f(a_1, \dots, a_{i-1}, a_i + t, a_{i+1}, \dots, a_n) - f(a)}{t}$$

et l'on dit que c'est la dérivée partielle de f en a selon la i-ème variable.

Notation 7.13. — Si f admet des dérivées partielles $\dfrac{\partial f}{\partial x_i}(a)$ en tout point $a \in U$, on obtient ainsi pour tout $i = 1, \dots, n$ une application :

$$\frac{\partial f}{\partial x_i} : U \to \mathbb{R}, \qquad a \mapsto \frac{\partial f}{\partial x_i}(a).$$

Avec la notation ci-dessus, il n'y a pas d'ambiguïté, mais souvent on écrit $a = (x_1, \ldots, x_n)$, d'où l'application

$$\frac{\partial f}{\partial x_i} : U \to \mathbb{R}, \qquad (x_1, \ldots, x_n) \mapsto \frac{\partial f}{\partial x_i}(x_1, \ldots, x_n).$$

On prendra garde que dans cette écriture, le x_i dans ∂x_i est un symbole pour désigner la dérivation selon le vecteur e_i, tandis que $(x_1, \ldots, x_n)$ est une « variable » qui décrit l'ouvert U.

Le calcul de $\dfrac{\partial f}{\partial x_i}(a)$ consiste donc à ne dériver l'expression de f que par rapport à la variable x_i.

Exemples 7.14. — (1) Si $f : \mathbb{R}^3 \to \mathbb{R}$ est définie par $f(x, y, z) = -2x \cos y$, on a :

$$\frac{\partial f}{\partial x}(x, y, z) = -2 \cos y, \quad \frac{\partial f}{\partial y}(x, y, z) = 2x \sin y, \quad \frac{\partial f}{\partial z}(x, y, z) = 0.$$

(2) Soit $f : \mathbb{R}_+^* \times \mathbb{R} \to \mathbb{R}$ définie par $f(x, y) = \operatorname{Arctan}(y/x)$. Comme la dérivée de $\operatorname{Arctan}(u)$ est $\dfrac{1}{1+u^2}$, on a :

$$(\dagger) \qquad \frac{\partial f}{\partial x}(x, y) = \frac{1}{1+(y/x)^2}\,\frac{-y}{x^2} = \frac{-y}{x^2+y^2}, \qquad \frac{\partial f}{\partial y}(x, y) = \frac{1}{1+(y/x)^2}\,\frac{1}{x} = \frac{x}{x^2+y^2}.$$

(Q) ***Proposition 7.15.*** — *Soit U un ouvert de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$. On suppose f différentiable en un point $a \in U$.*

(i) *Alors f admet des dérivées partielles en a*[4]*: pour tout $v \in \mathbb{R}^n - \{0\}$, on a $\dfrac{\partial f}{\partial v}(a) = df(a)(v)$.*

(ii) *Pour tout $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$ on a :*

$$df(a)(h) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(a) h_j.$$

Démonstration. — Par hypothèse, il existe une fonction $\varepsilon : \mathbb{R}^n \to \mathbb{R}$ continue et nulle en 0 telle qu'on ait

$$f(a+h) - f(a) - df(a)(h) = \|h\|\, \varepsilon(h)$$

pour tout $h \in \mathbb{R}^n$ de norme assez petite. Fixons $v \in \mathbb{R}^n - \{0\}$ et appliquons ce qui précède à $h = tv$, où t parcourt un petit intervalle ouvert $]-r, r[$. Alors, pour $t \neq 0$ on obtient

$$\frac{f(a+tv) - f(a)}{t} - df(a)(v) = \|v\|\, \varepsilon(tv) = \varphi(t)$$

et $t \mapsto \varphi(t)$ est continue et nulle en 0. Ceci montre que la dérivée partielle $\dfrac{\partial f}{\partial v}(a)$ existe et vaut $df(a)(v)$.

En particulier, pour $v = e_j$ on obtient $df(a)(e_j) = \dfrac{\partial f}{\partial e_j}(a) = \dfrac{\partial f}{\partial x_j}(a)$. Enfin, comme $df(a)$ est linéaire, pour $h = (h_1, \ldots, h_n) = \sum_j h_j e_j$ on obtient :

$$df(a)(h) = \sum_j h_j df(a)(e_j) = \sum_j \frac{\partial f}{\partial x_j}(a)\, h_j.$$

□

Corollaire 7.16. — *Soient $U \neq \varnothing$ un ouvert connexe de $\mathbb{R}^n$ et $f : U \to \mathbb{R}^p$ différentiable. Si $df(a) = 0$ pour tout $a \in U$, alors f est constante sur U.*

---

[4] On verra dans la section suivante que la réciproque est fausse en général, mais que si f admet sur U des dérivées partielles qui sont *continues*, alors f est différentiable sur U.

Démonstration. — Soit $a \in U$ et $c = f(a)$; notons $U_c = \{x \in U \mid f(x) = c\}$. Montrons que U_c est un *ouvert*. Soit $x \in U_c$; il existe $r > 0$ tel que la boule ouverte $B = B(x, r)$ soit contenue dans U. Pour tout $y \in B$, le segment $[x, y]$ est contenu dans B, donc dans U. L'application $\gamma : \mathbb{R} \to \mathbb{R}^n$, $t \mapsto x + t(y - x)$ est dérivable et $\gamma([0, 1]) = [x, y]$. Comme U est ouvert, $\gamma^{-1}(U)$ est un intervalle ouvert I contenant $[0, 1]$. D'après le point (iii) du théorème précédent, $f \circ \gamma : I \to \mathbb{R}^p$ est dérivable, de dérivée nulle, donc constante. Il en résulte $f(y) = f(x)$. Ceci prouve que B est contenue dans U_c, donc U_c est ouvert.

Pour tout réel μ, le même raisonnement montre que $U_\mu = \{x \in U \mid f(x) = \mu\}$ est un ouvert de U, donc $\Omega = \bigcup_{\mu \neq c} U_\mu$ est un ouvert de U, disjoint de U_c et tel que $U = U_c \sqcup \Omega$. Comme U est supposé connexe et que U_c est non vide (car il contient a), on en déduit que $U = U_c$ (et $\Omega = \varnothing$), i.e. f est constante sur U, de valeur c. □

Bien entendu, il est nécessaire de supposer U connexe. Sinon on peut prendre $U = \mathbb{R} - \{0\}$ et $f(x) = 1$ si $x > 0$, $f(x) = -1$ si $x < 0$.

On a vu plus haut (7.15) que si $f : \mathbb{R}^n \to \mathbb{R}$ est différentiable en a, alors $df(a)$ est la **forme linéaire** $\mathbb{R}^n \to \mathbb{R}$ donnée par

$$df(a)(h) = \frac{\partial f}{\partial x_1}(a) h_1 + \cdots + \frac{\partial f}{\partial x_n}(a) h_n$$

i.e. $df(a)$ est donnée par la matrice **ligne** :

$$(\star) \qquad \Big(\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a)\Big).$$

***Définition 7.17* (Matrice jacobienne).** — Soit maintenant $f = (f_1, \ldots, f_p)$ une fonction $\mathbb{R}^n \to \mathbb{R}^p$. D'après la proposition 7.9, f est différentiable en un point a ssi chaque f_i l'est, et dans ce cas pour tout $h \in \mathbb{R}^n$ on a :

$$df(a)(h) = \begin{pmatrix} df_1(a)(h) \\ \vdots \\ df_p(a)(h) \end{pmatrix}.$$

Tenant compte de l'expression pour chaque $df_i(a)(h)$ donnée en $(\star)$ plus haut, on obtient que la
(Q) matrice de $df(a)$ est la matrice

$$Df(a) = \Big(\frac{\partial f_i}{\partial x_j}\Big)_{\substack{1 \le i \le p \\ 1 \le j \le n}} = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(a) & \cdots & \dfrac{\partial f_1}{\partial x_n}(a) \\ \vdots & & \vdots \\ \dfrac{\partial f_p}{\partial x_1}(a) & \cdots & \dfrac{\partial f_p}{\partial x_n}(a) \end{pmatrix}$$

dont la i-ème ligne est donnée par $df_i(a)$. Cette matrice est appelée **matrice jacobienne** de f en a. Lorsque $p = n$, $Df(a)$ est une matrice carrée et on note $J_f(a)$ son déterminant, qu'on appelle le (déterminant) **jacobien** de f en a.

Revenant au cas n, p arbitraires, rappelons que pour toute application linéaire $u : \mathbb{R}^n \to \mathbb{R}^p$, de matrice $A \in M_{p,n}(\mathbb{R})$, l'image par u d'un vecteur $x = (x_1, \ldots, x_n)$ s'obtient en appliquant A au vecteur **colonne** $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, i.e. l'élément $u(x)$ de $\mathbb{R}^p$ est donné par le vecteur colonne $AX \in \mathbb{R}^p$.

(Q) ***Théorème 7.18.*** — *Soient U, V des ouverts de $\mathbb{R}^n$ et $\mathbb{R}^p$ et $f : U \to V$ et $g : V \to \mathbb{R}^q$ des applications.*

(i) *Si f est différentiable en a et g en $f(a)$, alors $g \circ f$ est différentiable en a et l'on a :*

$$(\star) \qquad d(g \circ f)(a) = dg(f(a)) \circ df(a).$$

(ii) *Si f est différentiable sur U et g sur V, alors $g \circ f$ est différentiable sur U.*

(iii) *En particulier, si $n = 1$ et $U = I$ est un intervalle ouvert de $\mathbb{R}$, l'application $g \circ f : I \to \mathbb{R}^p$ est dérivable et pour tout $t \in I$ on a :*

$$(*) \qquad (g \circ f)'(t) = dg(f(t))(f'(t)).$$

Démonstration. — (i) Posons $b = f(a)$. Par hypothèse, il existe des fonctions $\eta : \mathbb{R}^n \to \mathbb{R}^p$ et $\mu : \mathbb{R}^p \to \mathbb{R}^q$, continues et nulles en 0, telles que pour $h \in \mathbb{R}^n$ et $h' \in \mathbb{R}^p$ assez petits, on ait :

$$f(a+h) = b + df(a)(h) + \|h\|\,\eta(h)$$
$$g(b+h') = g(b) + dg(b)(h') + \|h'\|\,\mu(h').$$

Pour h assez petit, posons

$$k(h) = f(a+h) - f(a) = df(a)(h) + \|h\|\,\eta(h).$$

Alors pour $h \neq 0$ on a

$$\frac{g(f(a+h)) - g(b) - dg(b)(df(a)(h))}{\|h\|} = dg(b)(\eta(h)) + \frac{\|k(h)\|}{\|h\|}\mu(k(h)).$$

Montrons que le membre de droite tend vers 0 quand $h \neq 0$ tend vers 0. Pour le premier terme c'est clair, car $dg(b)$ est continue (car linéaire) et η est continue et nulle en 0.

Notons $\psi(h)$ le second terme. Comme η est continue et nulle en 0, il existe $\delta_0 > 0$ tel que $\|\eta(h)\| < 1$ si $\|h\| < \delta_0$. Comme $df(a)$ est L-lipschitzienne, posant $C = L + 1$, on obtient que pour tout h tel que $\|h\| < \delta_0$, on a

$$\|k(h)\| \leq C\,\|h\| \qquad \text{et donc} \qquad \|\psi(h)\| \leq C\,\|\mu(k(h))\|.$$

Comme $\mu \circ k$ est continue et nulle en 0, il en résulte que $\lim_{\substack{h \to 0 \\ h \neq 0}} \psi(h) = 0$. Ceci prouve (i) et (ii).

Déduisons-en le cas particulier (iii). D'après 7.7 (4), une application $\phi : I \to \mathbb{R}^q$ est différentiable en un point t ssi elle est dérivable en t et dans ce cas $d\phi(t)$ est l'application linéaire $h \mapsto h\phi'(t)$.

Ici, on sait d'après (i), que $g \circ f$ est différentiable en t, de différentielle $dg(f(t)) \circ df(t)$. Or $df(t)$ est l'application linéaire $\mathbb{R} \to \mathbb{R}^n$, $h \mapsto hf'(t)$ et donc $d(g \circ f)(t)$ est l'application linéaire $\mathbb{R} \to \mathbb{R}^n$, $h \mapsto hdg(f(t))(f'(t))$. Il en résulte que $(g \circ f)'(t) = dg(f(t))(f'(t))$. □

Remarque 7.19. — La définition de la différentiabilité et le théorème précédent illustrent un principe général en mathématiques : il a fallu travailler un peu pour établir la définition (i.e. montrer que $df(a)$ est unique si elle existe) puis pour démontrer le théorème, mais ce travail ayant été fait une fois pour toutes, on dispose d'une notion qui est facile à manipuler, comme le montre la jolie formule $d(g \circ f)(a) = dg(f(a)) \circ df(a)$, dont on verra plus bas la traduction en termes de produit de matrices.

***Remarque 7.20* (Traduction matricielle).** — Le théorème précédent s'écrit en termes matriciels comme suit. Considérons des ouverts $U \subset \mathbb{R}^n$ et $V \subset \mathbb{R}^p$ et des applications différentiables $f : U \to V$ et $g : V \to \mathbb{R}^q$. Pour tout $a \in U$, soit $A = Df(a) \in M_{p,n}(\mathbb{R})$ la matrice jacobienne de f en a et $B = Dg(f(a)) \in M_{q,p}(\mathbb{R})$ celle de g en $b = f(a)$. Alors la matrice jacobienne de $g \circ f$ en a est

(Q)

$$D(g \circ f)(a) = BA.$$

Remarque 7.21 **(Attention !)**. — Contrairement aux fonctions d'une seule variable, où l'on peut écrire $(g \circ f)'(a) = g'(f(a))f'(a) = f'(a)g'(f(a))$ (puisque le produit dans $\mathbb{R}$ est commutatif), l'ordre d'apparition des différentielles dans la formule $d(g \circ f)(a) = dg(f(a)) \circ f(a)$ est extrêmement important. En effet, $df(a)$ va de $\mathbb{R}^n \to \mathbb{R}^p$ et $dg(f(a))$ va de $\mathbb{R}^p$ dans $\mathbb{R}^q$, donc on ne peut même pas les composer dans le « mauvais sens » si $n \neq q$. Et même si $n = p = q$, la composition dans le « mauvais sens » ne donne pas le bon résultat, puisque la multiplication dans $M_n(\mathbb{R})$ n'est pas commutative.

Remarque 7.22. — Écrivons la différentielle d'une composée $\mathbb{R}^n \xrightarrow{f} \mathbb{R}^p \xrightarrow{g} \mathbb{R}^q$ en termes de dérivées partielles. Posons $A = Df(a)$ et $B = Dg(f(a))$, alors $D(g \circ f) = BA$. Donc pour tout $j = 1, \dots, n$ et $i = 1, \dots, q$, on a

$$(BA)_{ij} = \sum_{k=1}^{p} B_{ik} A_{kj}.$$

Si l'on note $(u_1, \dots, u_p)$ les coordonnées sur $\mathbb{R}^p$, alors on a

$$B_{ik} = \frac{\partial g_i}{\partial u_k}(f(a)) \qquad \text{et} \qquad A_{kj} = \frac{\partial f_k}{\partial x_j}(a)$$

(Q) et donc l'égalité précédente donne :

$$(\dagger) \qquad \frac{\partial (g \circ f)_i}{\partial x_j}(a) = \sum_{k=1}^{p} \frac{\partial g_i}{\partial u_k}(f(a)) \frac{\partial f_k}{\partial x_j}(a).$$

Le calcul consiste donc à : dériver g_i par rapport à la variable u_k et évaluer le résultat en $f(a)$, puis multiplier par la dérivée de f_k par rapport à la variable x_j évaluée en a, puis sommer par rapport à k.

Proposition 7.23. — *Soient U un ouvert de $\mathbb{R}^n$, f, g deux applications différentiables $U \to \mathbb{R}^p$,*
(Q) et $a \in U$.

(i) *Pour tout $\lambda, \mu \in \mathbb{R}$, $\lambda f + \mu g$ est différentiable en a et $d(\lambda f + \mu g)(a) = \lambda df(a) + \mu dg(a)$.*
(ii) *Si $p = 1$, alors fg est différentiable en a et $d(fg)(a) = f(a)dg(a) + g(a)df(a)$.*

Démonstration. — (i) est laissé en exercice ; prouvons (ii). D'après la proposition 7.9, l'application $F : \mathbb{R}^n \to \mathbb{R}^2$, $x \mapsto (f(x), g(x))$ est différentiable en a, et pour tout $h \in \mathbb{R}^n$ on a

$$dF(a)(h) = \big(df(a)(h), dg(a)(h)\big).$$

D'autre part, d'après l'exemple 7.7 (3), l'application $m : \mathbb{R}^2 \to \mathbb{R}$, $(x_1, x_2) \mapsto x_1 x_2$, est différentiable sur $\mathbb{R}^2$ et sa différentielle en $F(a) = (f(a), g(a))$ est la forme linéaire $(h_1, h_2) \mapsto g(a)h_1 + f(a)h_2$.

D'après le théorème 7.18, l'application $fg = m \circ F$ est donc différentiable en a, de différentielle $d(fg)(a) = dm(F(a)) \circ dF(a)$, i.e. pour tout $h \in \mathbb{R}^n$ on a

$$d(fg)(a)(h) = dm(F(a))\big(df(a)(h), dg(a)(h)\big) = g(a)df(a)(h) + f(a)dg(a)(h)$$

i.e. $d(fg)(a)$ est la forme linéaire $g(a)df(a) + f(a)dg(a)$. □

Après ces généralités sur les applications différentiables, donnons un critère concret et utile de différentiabilité, qui sera vérifié par la plupart des fonctions considérées dans ce cours.

Définition 7.24 **(Fonctions de classe C^1)**. — Soient $U \subset \mathbb{R}^n$ un ouvert et $f : U \to \mathbb{R}$ une application. On dit que f est de classe C^1 sur U si f admet des dérivées partielles $\partial_j f$ pour
(Q) $j = 1, \dots, n$ et si celles-ci sont **continues** sur U.

De même, pour $f = (f_1, \dots, f_p) : U \to \mathbb{R}^p$ on dit que f est de classe C^1 si chaque f_i l'est : ceci équivaut à dire que l'application

$$\Phi : U \to M_{p,n}(\mathbb{R}), \qquad a \mapsto \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(a) & \cdots & \dfrac{\partial f_1}{\partial x_n}(a) \\ \vdots & & \vdots \\ \dfrac{\partial f_p}{\partial x_1}(a) & \cdots & \dfrac{\partial f_p}{\partial x_n}(a) \end{pmatrix}$$

est **continue**.

On notera $\mathscr{C}^1(U, \mathbb{R}^p)$ l'ensemble des fonctions de classe C^1 sur U à valeurs dans $\mathbb{R}^p$.

(Q) ***Théorème 7.25.*** — *Soient $U \subset \mathbb{R}^n$ un ouvert et $f \in \mathscr{C}^1(U, \mathbb{R}^p)$. Alors f est différentiable sur U et l'application*

$$Df : U \to M_{p,n}(\mathbb{R}), \qquad a \mapsto Df(a)$$

est une application continue.

Démonstration. — D'abord, en considérant les composantes $f_1, \dots, f_p$ de f, il suffit de traiter le le cas $p = 1$. Faisons alors la démonstration pour $p = 1$ et $n = 3$, ce qui est suffisant pour bien comprendre l'idée. On utilise la norme $\|\cdot\|_\infty$. Soit $a \in U$. Quitte à faire le changement de variable $x' = a + x$, on peut supposer $a = 0$, ce qui va permettre d'alléger l'écriture. Fixons $\varepsilon > 0$.

Soit $r > 0$ tel que $B(a, r) \subset U$. Pour $j = 1, 2, 3$, comme $\partial_j f$ est continue en $a = 0$, il existe $\delta_j \in \,]0, r[$ tel que $|\partial_j f(x) - \partial_j f(0)| < \varepsilon/3$ si $\|x\| < \delta_j$. Posons $\delta = \min_{i=1,2,3} \delta_i$, alors pour tout $x \in B(0, \delta)$ on a

$$(*) \qquad |\partial_j f(x) - \partial_j f(0)| < \frac{\varepsilon}{3}.$$

Soit $h = (h_1, h_2, h_3) \in B(0, \delta)$, alors $(h_1, h_2, 0)$ et $(h_1, 0, 0)$ sont aussi dans $B(0, \delta)$. Comme f est dérivable par rapport à la variable x_3 au point $(h_1, h_2, 0)$ alors, d'après le théorème des accroissements finis en une variable, il existe un réel $\theta_3 \in [0, 1]$ (dépendant de h_1, h_2 et h_3) tel que [5]

$$f(h_1, h_2, h_3) - f(h_1, h_2, 0) = \partial_3 f(h_1, h_2, \theta_3 h_3) h_3.$$

De même, il existe $\theta_2 \in [0, 1]$ (dépendant de h_1 et h_2) tel que

$$f(h_1, h_2, 0) - f(h_1, 0, 0) = \partial_2 f(h_1, \theta_2 h_2, 0) h_2$$

et $\theta_1 \in [0, 1]$ (dépendant de h_1) tel que

$$f(h_1, 0, 0) - f(0, 0, 0) = \partial_1 f(\theta_1 h_1, 0, 0) h_1.$$

En sommant ces trois égalités, soustrayant $L(h) = \sum_{i=1}^3 \partial_i f(0, 0, 0) h_i$ et utilisant $(*)$, on obtient

$$\big| f(h) - f(0) - L(h) \big| \le \frac{\varepsilon}{3} \sum_{i=1}^{3} |h_i| \le \varepsilon \, \|h\|_\infty.$$

Ceci prouve que f est différentiable en $a = 0$, de différentielle la forme linéaire

$$L : (h_1, h_2, h_3) \mapsto \sum_{j=1}^{3} \frac{\partial f}{\partial x_j}(a) h_j.$$

---

[5] C'est la formule $f(b) - f(a) = (b - a) f'(c)$ pour un $c \in [a, b]$ que l'on écrit $c = a + \theta(b - a)$ avec $\theta \in [0, 1]$.

De plus, l'application $df : U \to \mathscr{L}(\mathbb{R}^3, \mathbb{R}) = M_{1,3}(\mathbb{R})$ (matrices à une ligne et 3 colonnes) est donnée par

$$a \mapsto \big(\partial_1 f(a), \partial_2 f(a), \partial_3 f(a)\big)$$

donc est continue. De même, pour p et n arbitraires, l'application $Df : U \to M_{p,n}(\mathbb{R})$ associe à a la matrice $\left(\dfrac{\partial f_i}{\partial x_j}(a)\right)$ dont la composante d'indice (i, j) est $\dfrac{\partial f_i}{\partial x_j}$ qui est continue sur U, donc Df est bien continue. Ceci achève la preuve du théorème. □

***Remarque 7.26* (Attention !).** — Il faut se garder de croire que l'application $U \to M_{p,n}(\mathbb{R})$, $a \mapsto Df(a)$ est linéaire : en effet, chaque coefficient d'indice (i, j) de la matrice $Df(a)$ est donné par $\dfrac{\partial f_i}{\partial x_j}(a)$, qui en général n'est pas une fonction linéaire de a et peut être arbitrairement compliquée. Par exemple, pour $U = \mathbb{R}$ et $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^n$, c'est l'application $U \to M_{1,1}(\mathbb{R})$ qui à tout $a \in U$ associe la matrice (na^{n-1}).

Ou bien, si l'on préfère, soit $f : \mathbb{R}^2 \to \mathbb{R}^2$, $(x, y) \mapsto (x^3 + y^4, x^5 + y^6)$. Alors f est de classe C^1 et pour tout $(x, y) \in \mathbb{R}^2$, on a

$$Df(x, y) = \begin{pmatrix} 3x^2 & 4y^3 \\ 5x^4 & 6y^5 \end{pmatrix}.$$

Remarques 7.27. — 1) Une fonction dont toutes les dérivées partielles existent n'est pas nécessairement différentiable. Par exemple, soit $f : \mathbb{R}^2 \to \mathbb{R}$ définie, pour tout $(x, y) \in \mathbb{R}^2$, par

$$f(x, y) = \begin{cases} \dfrac{xy(x + y)}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0), \\ 0 & \text{si } (x, y) = (0, 0). \end{cases}$$

Alors, f est de classe C^1 sur $U = \mathbb{R}^2 - \{(0, 0)\}$. De plus, pour tout (x, y) on a $\max(|x|, |y|) \leq \sqrt{x^2 + y^2}$ donc $|f(x, y)| \leq 2\sqrt{x^2 + y^2}$ donc f est continue en $0 = (0, 0)$.(6) De plus, pour tout vecteur non nul $v = \binom{a}{b}$ et tout $t \neq 0$, on a

$$\frac{f(0 + tv) - f(0)}{t} = \frac{t^3 ab(a + b)}{t^3(a^2 + b^2)} = \frac{ab(a + b)}{a^2 + b^2},$$

donc f est dérivable en 0 dans la direction v. En particulier, pour $v = \binom{1}{0} = e_1$ (resp. $\binom{0}{1} = e_2$) on obtient $(\partial f/\partial x_1)(0) = 0$ (resp. $(\partial f/\partial x_2)(0) = 0$). Donc f admet en 0 des dérivées dans toutes les directions. Mais f n'est pas différentiable en 0. En effet, si elle l'était alors d'après la proposition 7.15 on aurait $df(0) = 0$ et donc le rapport

$$\frac{f(x, y) - f(0)}{\|(x, y)\|_2} = \frac{xy(x + y)}{(x^2 + y^2)^{3/2}}$$

devrait tendre vers 0 quand (x, y) tend vers 0. Mais ceci n'est pas le cas : si ce rapport est bien nul sur les droites ke_1 et ke_2 (et aussi sur la droite d'équation $x + y = 0$), sur chaque droite $y = \mu x$ on a pour $x \neq 0$:

$$\frac{xy(x + y)}{(x^2 + y^2)^{3/2}} = \frac{x^3 \mu(1 + \mu)}{|x|^3(1 + \mu^2)^{3/2}} = \begin{cases} \dfrac{\mu(1 + \mu)}{(1 + \mu^2)^{3/2}} & \text{si } x > 0, \\ \dfrac{-\mu(1 + \mu)}{(1 + \mu^2)^{3/2}} & \text{si } x < 0. \end{cases}$$

2) D'autre part, la fonction $\mathbb{R} \to \mathbb{R}$ définie par $f(0) = 0$ et $f(x) = x^2 \sin(1/x)$ est dérivable sur $\mathbb{R}$ mais sa dérivée n'est pas continue en 0. On a donc, lorsque U est un ouvert de $\mathbb{R}^n$ avec $n \geq 2$, des inclusions strictes :

$$\mathscr{C}^1(U, \mathbb{R}) \subset \{\text{fonctions différentiables sur } U\} \subset \{\text{fonctions } f : U \to \mathbb{R} \text{ admettant des dérivées partielles}\}$$

(6) Pour $(x, y) \neq (0, 0)$ on peut aussi passer en coordonnées polaires : $f(r, \theta) = r\cos(\theta)\sin(\theta)(\cos(\theta) + \sin(\theta)) \leq 2r$.

Exercice 7.28. — Soit U l'ouvert $\mathbb{R}^2 - \{(x,0) \mid x \leq 0\}$. Soient $r : \mathbb{R}^2 - \{(0,0)\} \to \mathbb{R}$ et $g : U \to \mathbb{R}$ les fonctions définies par $r(x,y) = \sqrt{x^2+y^2}$ et $g(x,y) = 2\operatorname{Arctan}\left(\dfrac{y}{x+r(x,y)}\right)$. On rappelle que la dérivée de $\operatorname{Arctan}(t)$ est $\operatorname{Arctan}'(t) = 1/(1+t^2)$. Calculer les dérivées partielles $\partial_x r$ et $\partial_y r$ sur $\mathbb{R}^2 - \{(0,0)\}$ et les dérivées partielles $\partial_x g$ et $\partial_y g$ sur U. Montrer que l'application $F : U \to \mathbb{R}^2$, $(x,y) \mapsto (r(x,y), g(x,y))$ est de classe C^1.

Réponse partielle : $\dfrac{\partial g}{\partial x}(x,y) = \dfrac{-y}{x^2+y^2}$ et $\dfrac{\partial g}{\partial y}(x,y) = \dfrac{x}{x^2+y^2}$.

Signalons au passage la proposition suivante. Comme les dérivées partielles s'obtiennent en dérivant par rapport à une seule variable, elles jouissent des mêmes propriétés que celles connues pour les fonctions d'une seule variable ; la démonstration est donc omise.

Proposition 7.29. — *Soient $f, g : \mathbb{R}^n \to \mathbb{R}$ deux fonctions admettant en a une dérivée partielle par rapport à la variable x_i.*

(i) *Pour tout $\lambda, \mu \in \mathbb{R}$, $\lambda f + \mu g$ admet en a une dérivée partielle par rapport à la variable x_i et*

$$\frac{\partial(\lambda f + \mu g)}{\partial x_i}(a) = \lambda \frac{\partial f}{\partial x_i}(a) + \mu \frac{\partial g}{\partial x_i}(a).$$

(ii) *fg admet une dérivée partielle en a par rapport à la variable x_i et*

$$\frac{\partial(fg)}{\partial x_i}(a) = f(a)\frac{\partial g}{\partial x_i}(a) + g(a)\frac{\partial f}{\partial x_i}(a).$$

(iii) *Si $f(a) \neq 0$ alors $1/f$ admet une dérivée partielle en a par rapport à la variable x_i et*

$$\frac{\partial(1/f)}{\partial x_i}(a) = \frac{-1}{f(a)^2}\frac{\partial f}{\partial x_i}(a).$$

8. Difféomorphismes. Exemple des coordonnées polaires, cylindriques et sphériques

Définition 8.1. — Soient U un ouvert de $\mathbb{R}^n$ et $f : U \to \mathbb{R}^n$ une application de classe C^1. On dit que f est un C^1-**difféomorphisme** de U sur son image si les trois conditions suivantes sont vérifiées :

(Q)

a) f est injective.
b) $V = f(U)$ est un ouvert de $\mathbb{R}^n$.
c) La bijection réciproque $f^{-1} : V \to U$ est de classe C^1.

Alors les égalités $f^{-1}(f(x)) = x$ pour tout $x \in U$ et $f(f^{-1}(y)) = y$ pour tout $y \in V$ entraînent, d'après la formule de différentiation des fonctions composées (Th. 7.18), les égalités :

$$d(f^{-1})(f(x)) \circ df(x) = \mathrm{id}, \qquad df(f^{-1}(y)) \circ d(f^{-1})(y) = \mathrm{id}$$

pour tout $x \in U$, $y \in V$. Ceci montre que les conditions (a,b,c) ci-dessus entraînent :

d) Chaque $df(x)$ est **inversible** et la différentielle de f^{-1} en $f(x)$ est l'application linéaire $df(x)^{-1}$. En termes matriciels, ceci signifie que la matrice jacobienne $Df^{-1}(f(x))$ est la matrice inverse de la matrice $Df(x)$.

Donc, la condition d) que chaque $df(x)$ soit inversible est une condition *nécessaire* pour que f soit un difféomorphisme. Elle n'est pas suffisante, car elle n'entraîne pas que f soit injective : on le verra plus bas lors de l'introduction des coordonnées polaires. Donnons ici un autre exemple, d'ailleurs lié aux coordonnées polaires. Soit $f : \mathbb{R}^2 \to \mathbb{R}^2$ l'application définie par $f(x, y) = (e^x \cos y, e^x \sin y)$. La matrice jacobienne

$$Df(x,y) = \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix}$$

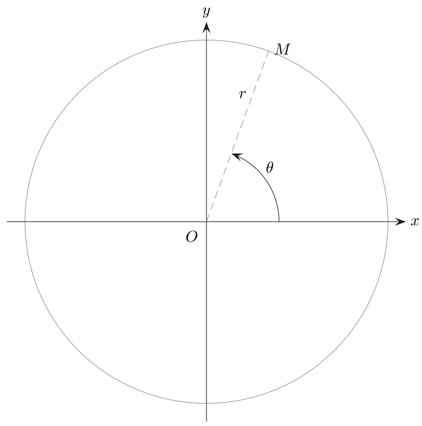
a pour déterminant e^{2x} donc est inversible pour tout $(x, y) \in \mathbb{R}^2$. Mais f n'est pas injective car $f(x, y+2\pi) = f(x, y)$. On peut montrer (cf. les coordonnées polaires plus bas) que $f(\mathbb{R}^2)$ est l'ouvert $\mathbb{R}^2 - \{(0, 0)\}$. [7] Mais, d'après un théorème que l'on verra plus loin (le théorème d'inversion locale), on a le résultat suivant, que nous admettrons pour le moment.

***Corollaire 8.2* (du th. d'inversion locale).** — *Soient U un ouvert de $\mathbb{R}^n$ et $f : U \to \mathbb{R}^n$ une application* injective *de classe C^1 telle que $df(x)$ soit inversible pour tout $x \in U$. Alors $V = f(U)$ est un ouvert de $\mathbb{R}^n$ et f est un C^1-difféomorphisme de U sur V, i.e. l'application inverse $f^{-1} : V \to U$ est de classe C^1.*

***Définition 8.3* (Coordonnées polaires).** — L'application $f : \mathbb{R}_+^* \times \mathbb{R} \to \mathbb{R}^2$, $(r, \theta) \mapsto$
(Q) $(r \cos\theta, r \sin\theta)$ est de classe C^1 : sa matrice jacobienne en tout point (r, θ) est

$$Df(r,\theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}.$$

Le déterminant jacobien est égal à $r > 0$, donc $Df(r, \theta)$ est inversible. Tout point $(x, y) \neq (0, 0)$ s'écrit $(r \cos\theta, r \sin\theta)$ où r est déterminé par $r = \sqrt{x^2 + y^2}$ et θ est unique modulo 2π. Donc, pour que f soit injective il faut la restreindre à $\mathbb{R}_+^* \times I$, où I est un intervalle ouvert $]\alpha, \alpha + 2\pi[$ de longueur 2π ; on obtient alors que f induit une bijection de $\mathbb{R}_+^* \times I$ sur l'ouvert $V_\alpha = \mathbb{R}^2$ privé de la demi-droite $D_\alpha = \{r(\cos\alpha, \sin\alpha) \mid r \in \mathbb{R}_+\}$. Un choix usuel est de prendre $\alpha = -\pi$. On obtient ainsi un C^1-difféomorphisme entre $U = \mathbb{R}_+^* \times\ ]-\pi, \pi[$ et l'ouvert $V = \mathbb{R}^2$ privé de la demi-droite fermée formée des réels ≤ 0. On obtient ainsi les « coordonnées polaires » (r, θ) sur V, i.e. pour tout point $M = (x, y)$ de V, ses coordonnées polaires sont l'unique couple $(r, \theta) \in U$ tel que $x = r \cos\theta$ et $y = r \sin\theta$. [8]



Exemple 8.4. — Étudions le difféomorphisme inverse $\phi = V \to U$, $(x, y) \mapsto (r, \theta)$. Il est clair que $r = r(x, y) = \sqrt{x^2 + y^2}$. Pour exprimer θ, si l'on se place dans le demi-plan d'équation $x > 0$, on peut écrire que $\theta = \mathrm{Arctan}(y/x)$. On a des formules analogues dans chacun des demi-plans $x < 0$,

---

[7] **Exercice**. Via l'identification $\mathbb{C} = \mathbb{R}^2$, à quoi correspond l'application f ?

[8] Les figures qui suivent sont dues à Laurent Koelblen.

$y > 0$ ou $y < 0$. Mais on peut obtenir une formule uniforme en utilisant l'astuce suivante. Comme $1 + \cos\theta = 2\cos^2(\theta/2)$ et $\sin\theta = 2\sin(\theta/2)\cos(\theta/2)$, on a pour $-\pi < \theta < \pi$:

$$\tan(\theta/2) = \frac{\sin\theta}{1+\cos\theta} = \frac{y}{r+x} = \frac{y}{x+\sqrt{x^2+y^2}}$$

d'où $\theta = 2\,\mathrm{Arctan}\left(\dfrac{y}{x+\sqrt{x^2+y^2}}\right) = g(x,y)$.

Au point $(x,y) = f(r,\theta)$, on sait que

$$D\phi(x,y) = Df(r,\theta)^{-1} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}^{-1} = \frac{1}{r}\begin{pmatrix} r\cos\theta & r\sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

On en déduit, sans calcul supplémentaire, que :

$$\frac{\partial r}{\partial x}(x,y) = \cos\theta = \frac{x}{\sqrt{x^2+y^2}}, \qquad \frac{\partial r}{\partial y}(x,y) = \sin\theta = \frac{y}{\sqrt{x^2+y^2}}$$

et

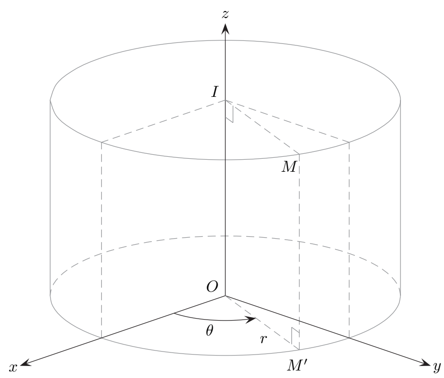
$$\frac{\partial g}{\partial x}(x,y) = \frac{-\sin\theta}{r} = \frac{-y}{x^2+y^2}, \qquad \frac{\partial g}{\partial y}(x,y) = \frac{\cos\theta}{r} = \frac{x}{x^2+y^2}.$$

On retrouve ainsi les résultats d'un exercice précédent.

(Q) ***Définition 8.5* (Coordonnées cylindriques).** — L'application $f : \mathbb{R}_+^* \times \mathbb{R}^2 \to \mathbb{R}^3$, $(r,\theta,z) \mapsto (r\cos\theta, r\sin\theta, z)$ est de classe C^1 : sa matrice jacobienne en tout point (r,θ,z) est

$$Df(r,\theta,z) = \begin{pmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Le déterminant jacobien est égal à $r > 0$, donc $Df(r,\theta,z)$ est inversible. Tout $M = (x,y,z)$ de $\mathbb{R}^3$ avec $(x,y) \neq (0,0)$ (i.e. $M \notin Oz$) s'écrit $(r\cos\theta, r\sin\theta, z)$ où r est déterminé par $r = \sqrt{x^2+y^2}$ et θ est unique modulo 2π. Comme précédemment, pour que f soit injective il faut la restreindre à $\mathbb{R}_+^* \times I \times \mathbb{R}$, où I est un intervalle ouvert $]\alpha, \alpha+2\pi[$ de longueur 2π ; on obtient alors que f induit une bijection de $\mathbb{R}_+^* \times I \times \mathbb{R}$ sur l'ouvert $V_\alpha = \mathbb{R}^3$ privé du demi-plan $H_\alpha = \{(r\cos\alpha, r\sin\alpha, z) \mid r \in \mathbb{R}_+,\ z \in \mathbb{R}\}$. Un choix usuel est de prendre $\alpha = -\pi$. On obtient ainsi un C^1-difféomorphisme entre $U = \mathbb{R}_+^* \times\,]-\pi,\pi[\, \times \mathbb{R}$ et l'ouvert $V = \mathbb{R}^3$ privé du demi-plan fermé $H = \{(x,0,z) \mid x \leq 0\}$. On obtient ainsi les « coordonnées cylindriques » (r,θ,z) sur V, i.e. pour tout point $M = (x,y,z)$ de V, ses coordonnées cylindriques sont l'unique triplet $(r,\theta,z) \in U$ tel que $x = r\cos\theta$, $y = r\sin\theta$ et $z = z$.



***Définition 8.6* (Coordonnées sphériques).** — Dans $\mathbb{R}^3$, on introduit les coordonnées sphériques comme suit. Posons $O = (0,0,0)$. Pour tout $M = (x,y,z) \neq O$ on pose $r = OM = \sqrt{x^2+y^2+z^2}$ et l'on note θ l'unique élément de $[0,\pi]$ tel que $z = r\cos\theta$. Si M n'appartient pas à la droite Oz, i.e. si $0 < \theta < \pi$, alors notant M' le projeté orthogonal de M sur le plan Oxy et $\rho = OM' = r\sin\theta > 0$, on a $M' = \rho(\cos\varphi, \sin\varphi, 0)$, avec φ unique modulo 2π. On dit alors que (r,θ,φ) sont les « coordonnées sphériques » de M.

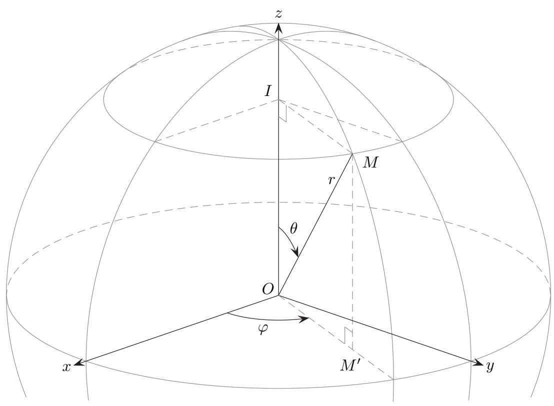
(Q) Autrement dit, l'application $f : \mathbb{R}_+^* \times \,]0,\pi[\, \times \mathbb{R} \to \mathbb{R}^3$,

$$(r,\theta,\varphi) \mapsto (r\sin\theta\cos\varphi, r\sin\theta\sin\varphi, r\cos\theta)$$

est de classe C^1, sa matrice jacobienne en un point (r,θ,φ) est

$$Df(r,\theta,\varphi) = \begin{pmatrix} \sin\theta\cos\varphi & r\cos\theta\cos\varphi & -r\sin\theta\sin\varphi \\ \sin\theta\sin\varphi & r\cos\theta\sin\varphi & r\sin\theta\cos\varphi \\ \cos\theta & -r\sin\theta & 0 \end{pmatrix}$$

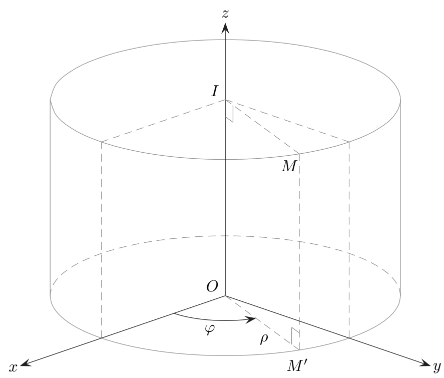
et son déterminant jacobien vaut $r^2\sin\theta$, qui est > 0. D'après ce qui précède, pour tout intervalle ouvert $I = \,]\alpha, \alpha+2\pi[$ de longueur 2π, f induit une bijection de $\mathbb{R}_+^* \times \,]0,\pi[\, \times I$ sur l'ouvert $V_\alpha = \mathbb{R}^3$ privé du demi-plan $H_\alpha = \{(r\cos\alpha, r\sin\alpha, z) \mid r \in \mathbb{R}_+,\ z \in \mathbb{R}\}$. Le choix usuel est de prendre $\alpha = -\pi$. On obtient ainsi un C^1-difféomorphisme entre $U = \mathbb{R}_+^* \times \,]0,\pi[\, \times \,]-\pi,\pi[$ et l'ouvert $V = \mathbb{R}^3$ privé du demi-plan fermé $H = \{(x,0,z) \mid x \leq 0\}$. On obtient ainsi les « coordonnées sphériques » (r,θ,φ) sur V, i.e. pour tout point $M = (x,y,z)$ de V, ses coordonnées sphériques sont l'unique triplet $(r,\theta,\varphi) \in U$ tel que $x = r\sin\theta\cos\varphi$, $y = r\sin\theta\sin\varphi$ et $z = r\cos\theta$.



Remarque 8.7. — Le difféomorphisme inverse est donné par $r = \sqrt{x^2+y^2+z^2}$, $\theta = \mathrm{Arccos}(z/r)$ et φ est donné par la même formule que pour les coordonnées polaires dans le plan Oxy.

Remarque. — Les coordonnées sphériques diffèrent des coordonnées « géographiques », où $\theta \in [-\pi/2, \pi/2]$ est défini par $z = r\sin\theta$, i.e. dans ces coordonnées on mesure θ (variant entre -90 et $+90$ degrés) à partir de l'équateur, tandis que dans les coordonnées sphériques on mesure θ (en radians) à partir du pôle Nord.

Remarque 8.8. — Pour que les coordonnées polaires dans $\mathbb{R}^2$ soient les restrictions à $\mathbb{R}^2$ des coordonnées sphériques dans $\mathbb{R}^3$ (et que les notations des coordonnées cylindriques et sphériques soient compatibles), il serait plus astucieux d'utiliser la notation (ρ,φ) (resp. (ρ,φ,z)) pour les coordonnées polaires (resp. cylindriques). C'est d'ailleurs la convention adoptée en physique :



Terminons cette section avec la définition du gradient de f puis la notion de point critique.

***Définition 8.9* (gradient).** — Soient U un ouvert de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$ une application différentiable. On a vu que pour tout $a \in U$, $df(a)$ est une *forme linéaire* sur $\mathbb{R}^n$. Comme il est psychologiquement plus facile (et plus visuel) de travailler avec des vecteurs que des formes linéaires, on fait ce qui suit.

(1) On munit $\mathbb{R}^n$ du produit scalaire standard $x \cdot y = \sum_i x_i y_i$ et l'on munit désormais, sauf mention du contraire, $\mathbb{R}^n$ de la norme euclidienne $\|\cdot\|_2$, qu'on notera simplement $\|\cdot\|$.

(2) Le produit scalaire induit un isomorphisme θ entre $\mathbb{R}^n$ et son dual, donné par $\theta(x)(h) = x \cdot h$ pour tout $x, h \in \mathbb{R}^n$. Par conséquent, pour toute forme linéaire $\phi : \mathbb{R}^n \to \mathbb{R}$, il existe un unique vecteur $u = u_\phi$ tel que $\phi(h) = u \cdot h$ pour tout $h \in \mathbb{R}^n$.

(3) Le vecteur colonne de $\mathbb{R}^n$ correspondant à la forme linéaire $df(a)$ est noté $\nabla f(a)$ et appelé (Q) le **gradient** de f en a ; on a donc :

$$\nabla f(a) = \begin{pmatrix} \partial_1 f(a) \\ \vdots \\ \partial_n f(a) \end{pmatrix} \qquad \text{et} \qquad df(a)(h) = \nabla f(a) \cdot h = \sum_{i=1}^{n} \partial_i f(a)\, h_i$$

pour tout $h = (h_1, \dots, h_n)$.

***Définitions 8.10* (Extrema locaux).** — Soient U un ouvert de $\mathbb{R}^n$, f une application $U \to \mathbb{R}$ et $a \in U$.

(i) On dit que f admet en a un minimum (resp. maximum) *local* s'il existe $r > 0$ tel que $B(a,r) \subset U$ et $f(a) \leq f(x)$ (resp. $f(a) \geq f(x)$) pour tout $x \in B(a,r)$.

(ii) On dit que f admet en a un minimum (resp. maximum) global si $f(a) \leq f(x)$ (resp. $f(a) \geq f(x)$) pour tout $x \in U$.

(iii) On utilisera le mot « extremum »[9] (local ou global) pour désigner sans distinction un maximum ou un minimum (local ou global).

(iv) Il est clair qu'un extremum global est *a fortiori* un extremum local. Mais il se peut que f admette des extrema locaux mais aucun extremum global. Par exemple, si $P(x)$ est un polynôme de degré 3 ayant trois racines réelles, par exemple $P(x) = x(x^2-1)$, alors P admet un maximum local et un mininum local (en les points où le polynôme dérivé $P'(x)$ s'annule) mais aucun extremum global.

[9] Le pluriel est *extrema*.

Définition 8.11. — Soient U un ouvert de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$ une application différentiable. On dit que $a \in U$ est un **point critique** de f si $df(a) = 0$, ce qui équivaut à dire que le gradient $\nabla f(a)$ est nul.

(Q) ***Proposition 8.12.*** — *Soient U un ouvert de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$ une application différentiable. Si f admet un extremum local en $a \in U$, alors a est un point critique de f (i.e. $df(a) = 0$).*

Démonstration. — Supposons par exemple que f ait un minimal local en a. Il existe $r > 0$ tel que $f(a + h) \geq f(a)$ pour tout $h \in \mathbb{R}^n$ vérifiant $\|h\| < 2r$. Fixons un tel h, posons $I =]-1, 1[$ et considérons la fonction $g : I \to \mathbb{R}$ définie par $g(t) = f(a + th)$. Alors g est dérivable sur I et

$$g'(t) = df(a + th)(h)$$

pour tout $t \in I$; en particulier $g'(0) = df(a)(h)$. D'autre part, comme g a un minimum local en 0, on a $g'(0) = 0$. Rappelons la démonstration : pour tout $t \in I$ on a $g(t) - g(0) \geq 0$ donc

$$\lim_{t\to 0^-} \frac{g(t) - g(0)}{t} \leq 0 \leq \lim_{t\to 0^+} \frac{g(t) - g(0)}{t}$$

et donc $g'(0)$, qui est la valeur commune des deux limites, est nul. On a donc $df(a)(h) = 0$ pour tout h vérifiant $\|h\| < 2r$. Comme on l'a déjà vu, ceci entraîne que $L = df(a)$ est nulle : en effet, pour $h \neq 0$ arbitraire, posons $\lambda = \|h\|/r$ et $h' = (1/\lambda)h$, alors $\|h'\| = r < 2r$ donc $L(h') = 0$, et comme $h = \lambda h'$ on a $L(h) = \lambda L(h') = 0$. □

Être un point critique est donc une condition nécessaire pour être un extremum. Elle n'est cependant pas suffisante au regard des exemples suivants.

Exemples 8.13. — (1) Soit $f : \mathbb{R}^2 \to \mathbb{R}$ la fonction définie par $f(x, y) = x^2 - y^2$ pour tout $(x, y) \in \mathbb{R}^2$. Elle est différentiable puisque c'est un polynôme. Pour tout $(x, y) \in \mathbb{R}^2$, $df(x, y)$ est la forme linéaire donnée par la matrice ligne :

$$\left(\frac{\partial f}{\partial x}(x, y), \frac{\partial f}{\partial y}(x, y)\right) = (2x, -2y)$$

qui s'annule en $(0, 0)$ (et en ce point uniquement). Donc $(0, 0)$ est un point critique de f. Mais, $f(0, 0) = 0$ et pour tout $\varepsilon \neq 0$ on a

$$f(\varepsilon, 0) = \varepsilon^2 > f(0, 0) > -\varepsilon^2 = f(0, \varepsilon)$$

donc f n'a pas en a de maximum ou minimum local.

(2) Un autre exemple est donné par $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^3$. La dérivée $f'(x) = 3x^2$ s'annule en 0 mais f n'a pas en 0 un extremum local car $f(x) > f(0) > f(-x)$ pour tout $x > 0$.

Pour avoir plus d'information afin de décider si un point critique a de f est un extremum local, on supposera dans la section 10 que f admet des dérivées partielles d'ordre 2 et l'on considérera la matrice hessienne de f en a.

9. Inégalité des accroissements finis

Dans cette section on va démontrer l'important théorème des accroissements finis, qui utilise la notion de convexité. Avant d'énoncer le théorème, il est utile d'introduire la définition suivante.

***Définition 9.1* (Norme d'opérateur).** — On munit $E = \mathbb{R}^n$ (resp. $F = \mathbb{R}^p$) d'une norme $\|\cdot\|_E$ (resp. $\|\cdot\|_F$). On définit alors sur $\mathscr{L}(E,F)$ la norme suivante, appelée la norme d'opérateur[10] (ou norme matricielle) *subordonnée* aux normes choisies sur E et F. Pour tout $\phi \in \mathscr{L}(E,F)$, on note

$$|\!|\!|\phi|\!|\!| = \max_{\|x\|_E=1} \|\phi(x)\|_F = \max_{x\neq 0} \frac{\|\phi(x)\|_F}{\|x\|_E}$$

la constante de Lipschitz de ϕ. On vérifie facilement que c'est bien une norme.

Donnons quelques exemples, d'abord lorsque $F = \mathbb{R}$ (muni de la valeur absolue). Dans ce cas, $\mathscr{L}(E,\mathbb{R})$ est le dual E^* de $E = \mathbb{R}^n$, i.e. l'espace des matrices lignes $L = (a_1, \dots, a_n)$

(1) Si E est muni de la norme $\|\cdot\|_\infty$, alors

$$|\!|\!|L|\!|\!| = \max_{|x_i|=1} \left|\sum_{i=1}^n a_i x_i\right| \leq \sum_{i=1}^n |a_i|$$

et l'égalité est obtenue si $x_i = \varepsilon_i$, où $\varepsilon_i = \pm 1$ est le signe de a_i (on prend $+1$ si $a_i = 0$). Donc dans ce cas la norme d'opérateur sur E^* est la norme $\|L\|_1 = \sum_{i=1}^n |a_i|$.

(2) Si E est muni de la norme euclidienne $\|\cdot\|_2$, l'inégalité de Cauchy-Schwarz donne :

$$|\!|\!|L|\!|\!| = \max_{\|x\|_2=1} \left|\sum_{i=1}^n a_i x_i\right| \leq \|L\|_2$$

et l'on a égalité si $L = 0$ ou si $x = \dfrac{1}{\|L\|_2}\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$. Donc dans ce cas la norme d'opérateur sur E^* est la norme euclidienne $\|\cdot\|_2$. [11]

(3) Si l'on munit $\mathbb{R}^n$ et $\mathbb{R}^p$ de la norme $\|\cdot\|_\infty$, alors la norme d'opérateur d'une matrice $A = (a_{ij})_{\substack{1\leq i\leq p\\ 1\leq j\leq n}}$ est le max des normes $\|\cdot\|_1$ des lignes, i.e. $|\!|\!|A|\!|\!| = \max_{i=1,\dots,n} \sum_{j=1}^n |a_{ij}|$.

(4) Si l'on munit $\mathbb{R}^n$ et $\mathbb{R}^p$ de la norme euclidienne, alors pour $A = (a_{ij})_{\substack{1\leq i\leq p\\ 1\leq j\leq n}}$ et $x \in \mathbb{R}^n$, l'inégalité de Cauchy-Schwarz donne :

$$\|Ax\|_2 = \left(\sum_{i=1}^p (L_i \cdot x)^2\right)^{1/2} \leq \|x\|_2 \left(\sum_{i,j} a_{ij}^2\right)^{1/2}$$

donc la norme d'opérateur de A est majorée par la norme de Frobenius $\left(\sum_{i,j} a_{ij}^2\right)^{1/2}$, mais cette inégalité est en général stricte. Par exemple, si $p = n$ et si A est la matrice identité I_n, sa norme d'opérateur est 1 tandis que sa norme de Frobenius est $\sqrt{n}$.

***Théorème 9.2* (Inégalité des accroissements finis).** — *Soit $U \subset \mathbb{R}^n$ un ouvert* <u>convexe</u> *et $f : U \to \mathbb{R}^p$ de classe C^1. Soient $a, b \in U$. Comme l'application $U \to M_{p,n}(\mathbb{R})$, $z \mapsto df(z)$ est continue, l'application $U \to \mathbb{R}$, $z \mapsto |\!|\!|df(z)|\!|\!|$ l'est aussi, donc elle est bornée sur le compact $[a,b]$. Posant $M = \max_{z\in[a,b]} |\!|\!|df(z)|\!|\!|$, on a alors :*

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

---

[10] En analyse, une application linéaire entre deux evn E et F est souvent appelée un « opérateur linéaire », d'où le nom « norme d'opérateur ».

[11] De façon générale, on peut montrer que si $E = \mathbb{R}^n$ est muni de la norme $\|\cdot\|_p$ avec $p \in [1,+\infty]$, alors la norme d'opérateur sur E^* est la norme $\|\cdot\|_q$ où $\dfrac{1}{q} + \dfrac{1}{p} = 1$.

Démonstration. — La démonstration pour une norme quelconque est assez technique. Contentons-nous de le prouver pour la norme $\|\cdot\|_\infty$ ou $\|\cdot\|_2$ sur $\mathbb{R}^p$. Posons $f = (f_1, \ldots, f_p)$.

Le cas de la norme $\|\cdot\|_\infty$ est facile : il suffit de montrer que

$$|f_i(b) - f_i(a)| \leq M \, \|b - a\|$$

pour tout i. On définit les applications $u : [0,1] \to [a,b]$ et $g_i : [0,1] \to \mathbb{R}$ par $u(t) = a + t(b-a)$ et $g_i(t) = f_i(u(t))$. Alors g_i est continue sur $[0,1]$ et dérivable sur $]0,1[$, de dérivée en t

$$g_i'(t) = df_i(u(t))(u'(t)) = df_i(u(t))(b-a).$$

D'après le théorème des accroissements finis en dimension 1, il existe $t_i \in]0,1[$ tel que

$$f_i(b) - f_i(a) = g_i(1) - g_i(0) = g_i'(t_i)(1-0) = g_i'(t_i)$$

et d'après la définition de $M = \max_{z\in[a,b]} \max_{i=1,\ldots,p} \|\!|df_i(z)|\!\|$, on a

$$|g_i'(t_i)| = \|df_i(u(t_i))(b-a)\| \leq M \, \|b-a\|$$

d'où le résultat.

Considérons maintenant la norme euclidienne $\|\cdot\|_2$. Si $f(b) = f(a)$ il n'y a rien à montrer, donc on peut supposer $f(b) \neq f(a)$. Posons $v = f(b) - f(a)$ et définissons les applications $u : [0,1] \to [a,b]$ et $g : [0,1] \to \mathbb{R}$ par $u(t) = a + t(b-a)$ et

$$g(t) = v \cdot f(u(t)) = \sum_{i=1}^{p} v_i f_i(u(t)),$$

i.e. g est la composée de $f \circ u$ et de l'application linéaire $x \mapsto v \cdot x$. Alors g est continue sur $[0,1]$ et dérivable sur $]0,1[$, de dérivée en t

$$g'(t) = v \cdot (f \circ u)'(t) = v \cdot df(u(t))(u'(t)) = v \cdot df(u(t))(b-a).$$

D'après le théorème des accroissements finis en dimension 1, il existe $t_0 \in]0,1[$ tel que

$$v \cdot f(b) - v \cdot f(a) = g(1) - g(0) = g'(t_0)(1-0) = g'(t_0),$$

soit $\|v\|^2 = v \cdot df(u(t_0))(b-a)$. Par conséquent, d'après l'inégalité de Cauchy-Schwarz et la définition de M, on a :

$$\|v\|^2 \leq \|v\| \, \|df(u(t_0))(b-a)\| \leq \|v\| \, M \, \|b-a\|$$

et en divisant par $\|v\| > 0$ on obtient

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

□

10. Dérivées partielles d'ordre deux, matrice hessienne, application aux extrema

Soient U un ouvert de $\mathbb{R}^n$, $f : U \to \mathbb{R}$ différentiable et a un point critique de f (i.e. $df(a) = 0$). Une condition *suffisante* pour que f ait en a un extremum local sera donnée en termes de la matrice des dérivées partielles d'ordre 2 de f en a, que nous allons maintenant introduire. (La notion de dérivées partielles d'ordre k existe aussi pour $k \geq 3$, mais nous nous limiterons ici au cas $k = 2$.)

***Définition 10.1* (Matrice hessienne et applications de classe C^2)**
Soit $U \subset \mathbb{R}^n$ un ouvert et $f : U \to \mathbb{R}$ de classe C^1.

(i) Soit $a \in U$. On dit que f admet en a des dérivées partielles d'ordre 2 si les fonctions $\dfrac{\partial f}{\partial x_1}, \dots, \dfrac{\partial f}{\partial x_n}$ admettent en a des dérivées partielles. On note alors pour tout $i, j = 1, \dots, n$,

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(a) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(a)$$

la j-ème dérivée partielle de la i-ème dérivée partielle de f en a. Lorsque $j = i$ on la note $\dfrac{\partial^2 f}{\partial x_i^2}(a)$.

(ii) On dit que f est *deux fois différentiable en a* ou que f *admet en a une différentielle seconde* $d^2 f(a)$ si l'application $g = df : U \to M_{1,n}(\mathbb{R}) \simeq \mathbb{R}^n$ est différentiable en a. Dans ce cas, sa
(Q) différentielle est l'application $dg = d^2 f : U \to M_n(\mathbb{R})$ donnée par la matrice suivante :

$$Dg(a) = D^2 f(a) = \begin{pmatrix} \dfrac{\partial g_1}{\partial x_1} & \cdots & \dfrac{\partial g_1}{\partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial g_n}{\partial x_1} & \cdots & \dfrac{\partial g_n}{\partial x_n} \end{pmatrix}(a) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f}{\partial x_n \partial x_1} \\ \vdots & & \vdots \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{pmatrix}(a)$$

appelée **matrice hessienne** de f au point a.

(iii) On dit que f est de classe C^2 sur U si les dérivées partielles $\dfrac{\partial^2 f}{\partial x_j \partial x_i}$ sont définies et continues en tout point de U. Dans ce cas, chaque application $\dfrac{\partial f}{\partial x_i}$ est de classe C^1 sur U, donc *différentiable* sur U, d'après le théorème 7.25.

(Q) ***Théorème 10.2* (de Schwarz).** — *Soit $U \subset \mathbb{R}^n$ un ouvert, $f : U \to \mathbb{R}$ une application de classe C^1 et $a \in U$.*

(i) *On suppose que f est deux fois différentiable en a. Alors pour tout $i, j = 1, \dots, n$, on a :*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a).$$

Par conséquent, la matrice hessienne $D^2 f(a)$ est **symétrique**.

(ii) *Ceci est le cas, en particulier, si les dérivées partielles $\dfrac{\partial^2 f}{\partial x_j \partial x_i}$ sont définies sur U et continues au point a.*

Démonstration. — Quitte à faire un changement de variables $x_i' = a_i + x_i$, on peut supposer que $a = 0$. Ensuite, fixant deux indices $i < j$ et considérant la fonction $(x, y) \mapsto f(0, \dots, x, \dots, y, \dots, 0)$ où x (resp. y) est à la i-ème (resp. j-ème) place, on se ramène au cas où $n = 2$, i.e. f est une fonction des deux variables x, y. Posons alors

$$g(x, y) = f(x, y) - f(x, 0) - f(0, y) + f(0, 0).$$

Alors $g(0, 0) = 0$ et g est, comme f, de classe C^1 sur un ouvert U de $\mathbb{R}^2$ contenant $0 = (0, 0)$ et ses dérivées partielles sont (en notant $f_x = \partial_x f$, etc.) :

$$g_x(x, y) = f_x(x, y) - f_x(x, 0), \qquad g_y(x, y) = f_y(x, y) - f_y(0, y).$$

En particulier, $g_x(0, 0) = 0 = g_y(0, 0)$. Notons $A = f_{xy}(0, 0)$ et $B = f_{yx}(0, 0)$. Fixons $\varepsilon > 0$.

Par hypothèse, les applications f_x et f_y, de U vers $\mathbb{R}$ sont différentiables en a donc il en est de même de g_x et g_y. Considérons d'abord g_x. Ses dérivées partielles sont :

$$g_{xx}(x,y) = f_{xx}(x,y) - f_{xx}(x,0), \qquad g_{yx}(x,y) = f_{yx}(x,y).$$

En particulier on a $g_{xx}(0,0) = 0$ et $g_{yx}(0,0) = f_{yx}(0,0) = B$. Donc pour tout $\begin{pmatrix} h \\ k \end{pmatrix} \in \mathbb{R}^2$ on a

$$dg_x(0,0)\begin{pmatrix} h \\ k \end{pmatrix} = (0,B)\begin{pmatrix} h \\ k \end{pmatrix} = Bk.$$

Par définition de la différentiabilité, il existe $\delta > 0$ tel que pour $\|(x,y)\|_\infty < \delta$ on ait $(x,y) \in U$ et

$$|g_x(x,y) - By| = \left| g_x(x,y) - g_x(0,0) - dg_x(0,0)\begin{pmatrix} x \\ y \end{pmatrix} \right| \leq \varepsilon \, \|(x,y)\|.$$

D'autre part, la fonction $h : t \mapsto g(t,y) - tBy$ est dérivable sur l'intervalle $]-\delta,\delta[$, de dérivée $h'(t) = g_x(t,y) - By$. Donc, d'après le théorème des accroissements finis en une variable, pour tout $(x,y) \in B_\infty(0,\delta)$ on a

$$|g(x,y) - Bxy| = |h(x) - h(0)| \leq \varepsilon \, |x| \, \|(x,y)\|_\infty. \tag{1}$$

On obtient de même que $g_{yy}(0,0) = 0$ et $g_{xy}(0,0) = f_{xy}(0,0) = A$ et qu'il existe $\delta' > 0$ tel que pour $\|(x,y)\|_\infty < \delta'$ on ait $(x,y) \in U$ et

$$|g(x,y) - Axy| \leq \varepsilon \, |y| \, \|(x,y)\|_\infty. \tag{2}$$

En utilisant l'inégalité triangulaire, on déduit de (1) et (2) que si $\|(x,y)\|_\infty < \min(\delta,\delta')$ on a

$$|xy(B-A)| \leq \varepsilon \, (|x| + |y|) \, \|(x,y)\|_\infty$$

et en prenant $x = y \neq 0$ et en simplifiant par x^2, on obtient

$$|B - A| \leq 2\varepsilon.$$

Comme ε était arbitraire, ceci montre que $A = B$. □

Remarque 10.3. — Lorsqu'on suppose que f est de classe C^2, on peut donner une démonstration un peu plus simple, comme suit. On se ramène comme plus haut au cas $n = 2$. Comme U est ouvert, il contient un carré ouvert $I \times I$, où I est un intervalle ouvert contenant 0. Fixons $t \neq 0$ dans I et considérons la fonction $h : x \mapsto f(x,t) - f(x,0)$. D'après le théorème des accroissements finis en une variable, pour tout $x \in I$ il existe un réel $\theta_1 \in [0,1]$ tel que

$$g(x,t) = h(x) - h(0) = xh'(\theta_1 x) = x\big(f_x(\theta_1 x, t) - f_x(\theta_1 x, 0)\big).$$

Appliquons ceci à $x = t$ et considérons maintenant la fonction $\psi : s \mapsto f_x(\theta_1 t, s)$. On a $\psi'(s) = f_{yx}(\theta_1 t, s)$. D'après le théorème des accroissements finis, à nouveau, il existe $\theta_2 \in [0,1]$ tel que

$$g(t,t) = t\big(\psi(t) - \psi(0)\big) = t^2\psi'(\theta_2 t) = t^2 f_{yx}(\theta_1 t, \theta_2 t).$$

Soit maintenant $\varepsilon > 0$. Comme f_{yx} est continue en 0, il existe $\delta > 0$ tel que, pour tout $h \in B_\infty(0,\delta)$, on ait :

$$|f_{yx}(h) - f_{yx}(0)| < \varepsilon. \tag{$*$}$$

Si $|t| < \delta$ alors le point $h = (\theta_1 t, \theta_2 t)$ apparaissant dans $(*)$ appartient à $B_\infty(0,\delta)$, puisque $\theta_1, \theta_2 \in [0,1]$, donc on a

$$\left| \frac{g(t,t)}{t^2} - f_{yx}(0) \right| < \varepsilon$$

pour tout $t \neq 0$ vérifiant $|t| < \delta$. Ceci prouve que

$$\lim_{\substack{t \to 0 \\ t \neq 0}} \frac{g(t,t)}{t^2} = f_{yx}(0).$$

Comme les deux variables x, y jouent des rôles symétriques dans la définition de $g(x,y)$, on obtient de même que la limite est aussi égale à $f_{yx}(0)$, d'où l'égalité voulue.

Remarques 10.4. — On peut montrer que les fonctions $f, g : \mathbb{R}^2 \to \mathbb{R}$ définie par $f(0,0) = 0 = g(0,0)$ et, pour $(x,y) \neq (0,0)$,

$$f(x,y) = \frac{x^2y^2}{x^2+y^2}, \qquad g(x,y) = \frac{xy(x^2-y^2)}{x^2+y^2}$$

sont de classe C^1 sur $\mathbb{R}^2$, de classe C^2 sur $\mathbb{R}^2 - \{(0,0)\}$, admettent en $(0,0)$ des dérivées partielles secondes, mais ne sont pas deux fois différentiables en $(0,0)$. Pour f, on a $f_{xy}(0,0) = 0 = f_{yx}(0,0)$ tandis que pour g on a $g_{xy}(0,0) = 1 \neq g_{yx}(0,0) = -1$.

Avant de prouver la formule de Taylor à l'ordre 2 pour une application de classe C^2, commençons par le lemme suivant. Pour tout $x, y \in \mathbb{R}^n$ on note $x \cdot y = \sum_{i=1}^n x_i y_i$.

Lemme 10.5. — *Soient $A \in M_n(\mathbb{R})$ symétrique et $Q : \mathbb{R}^n \to \mathbb{R}$ la forme quadratique définie par*

$$Q(x) = x \cdot Ax = \sum_{i=1}^n x_i \Big(\sum_{j=1}^n A_{ij} x_j \Big) = \sum_{i,j=1}^n A_{ij} x_i x_j = \sum_{i=1}^n A_{ii} x_i^2 + 2 \sum_{i<j} A_{ij} x_i x_j.$$

Alors Q est différentiable en tout $x \in \mathbb{R}^n$ et pour tout $h \in \mathbb{R}^n$ on a

$$dQ(x)(h) = 2 \sum_{i,j=1}^n A_{ij} x_j h_i = 2 \sum_{i=1}^n \Big(\sum_{j=1}^n A_{ij} x_j \Big) h_i = 2Ax \cdot h.$$

Démonstration. — Fixons $x \in \mathbb{R}^n$. Tenant compte de l'égalité $A_{ij} = A_{ji}$, on a pour tout $h \in \mathbb{R}^n$:

$$\begin{aligned} Q(x+h) = \sum_{i,j=1}^n A_{ij}(x_i + h_i)(x_j + h_j) &= Q(x) + 2 \sum_{i,j=1}^n A_{ij} x_j h_i + Q(h) \\ &= Q(x) + 2 \sum_{i=1}^n \Big(\sum_{j=1}^n A_{ij} x_j \Big) h_i + Q(h) = Q(x) + 2Ax \cdot h + Q(h). \end{aligned}$$

De plus, on a $|Q(h)| \leq \sum_{i,j=1}^n |A_{ij}| \, (\|h\|_\infty)^2$ et l'application $h \mapsto 2Ax \cdot h$ est linéaire. Ceci prouve que Q est différentiable en x et $dQ(x)(h) = 2Ax \cdot h$. □

(Q) ***Théorème 10.6*** **(Formule de Taylor à l'ordre** 2**).** — *Soit $U \subset \mathbb{R}^n$ un ouvert et $f : U \to \mathbb{R}$ de classe C^2 sur U. Pour tout $a \in U$, on a pour $h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$ assez petit,*

$$\begin{aligned} f(a+h) &= f(a) + \sum_{j=1}^n \frac{\partial f}{\partial x_j}(a) h_j + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(a) h_i h_j + o(\|h\|^2) \\ &= f(a) + df(a)(h) + \frac{1}{2} Q(h) + o(\|h\|^2), \end{aligned}$$

où $Q(h) = h \cdot D^2 f(a) h$ est la forme quadratique associée à la matrice symétrique $D^2 f(a)$.

Démonstration. — Fixons a et notons $A = D^2 f(a)$ et $\alpha_{ij} = \dfrac{\partial^2 f}{\partial x_j \partial x_i}(a)$. Soit $r > 0$ tel que U contienne $a + h$ pour tout h dans la boule ouverte $B = B(0, r)$. Pour tout $h \in B$, posons

$$\phi(h) = f(a+h) - df(a)(h) - \frac{1}{2} Q(h).$$

L'application $h \mapsto f(a+h)$ est la composée de f et de la translation $h \mapsto a + h$, dont la différentielle est l'application identique de $\mathbb{R}^n$. L'application $df(a) : h \mapsto df(a)(h)$ est linéaire, donc sa différentielle en tout point est $df(a)$. Enfin, le lemme précédent donne la différentielle $dQ(h)$. Par conséquent, ϕ est différentiable sur B et pour tout $h \in B$, $k \in \mathbb{R}^n$ on a :

$$d\phi(h)(k) = df(a+h)(k) - df(a)(k) - \sum_{i,j=1}^{n} \alpha_{ij} h_j k_i = \sum_{i=1}^{n} \beta_i(h) k_i$$

où $\beta_i(h) = \partial_i f(a+h) - \partial_i f(a) - \sum_{j=1}^{n} \partial_j(\partial_i f)(a) h_j = \partial_i f(a+h) - \partial_i f(a) - d(\partial_i f)(a)(h)$.

Soit $\varepsilon > 0$. Chaque $\partial_i f$ étant différentiable en a, il existe $\delta \in\,]0, r]$ tel que pour $\|h\| < \delta$ on ait $|\beta_i(h)| \leq \dfrac{\varepsilon}{n}\|h\|$ et donc, d'après la définition 9.1 (1),

$$(*) \qquad \|\!|d\phi(h)|\!\| = \sum_{i=1}^{n} |\beta_i(h)| \leq \varepsilon \, \|h\|.$$

Fixons $h \in B(0, \delta)$. Alors pour tout h' du segment $[0, h]$ (i.e. $h' = th$ avec $t \in [0, 1]$) on a $\|h'\| \leq \|h\| < \delta$, donc h' vérifie $(*)$ d'où :

$$\max_{h' \in [0,h]} \|\!|d\phi(h')|\!\| \leq \varepsilon \, \|h'\| \leq \varepsilon \, \|h\|.$$

Posant $C = \varepsilon \, \|h\|$, on déduit de l'inégalité des accroissements finis (Th. 9.2) que :

$$|\phi(h) - \phi(0)| \leq C \, \|h\| = \varepsilon \, \|h\|^2.$$

Comme $\phi(0) = f(a)$, ceci montre que pour tout $h \in B(0, \delta)$ on a

$$|f(a+h) - f(a) - df(a)(h) - \frac{1}{2}Q(h)| \leq \varepsilon \, \|h\|^2$$

ce qui prouve que la différence $f(a+h) - f(a) - df(a)(h) - \dfrac{1}{2}Q(h)$ est bien $o(\|h\|^2)$. $\square$

Remarque 10.7. — Donnons aussi une démonstration plus simple d'un résultat un peu plus faible, à savoir le cas où h varie dans une direction fixée. Fixons $h \in B(0, r)$ et posons $I = \,]-1, 1[$. Alors la fonction $g : I \to \mathbb{R}$ définie par $g(t) = f(a + th)$ est dérivable, de dérivée

$$g'(t) = df(a+th)(h) = \sum_{j=1}^{n} \partial_j f(a+th) h_j.$$

Comme f est de classe C^2, chaque $\partial_j f$ est de classe C^1 donc différentiable, donc $g'(t)$ est dérivable (et même de classe C^1) de dérivée :

$$g''(t) = \sum_{j=1}^{n} h_j d(\partial_j f)(a+th) = \sum_{j=1}^{n} h_j \sum_{i=1}^{n} \partial_i \partial_j f(a+th) h_i = \sum_{i,j=1}^{n} \partial_i \partial_j f(a+th) h_i h_j.$$

Comme $g : I \to \mathbb{R}$ est deux fois dérivable (et même de classe C^2), la formule de Taylor donne

$$f(a+th) = g(t) = g(0) + tg'(0) + \frac{t^2}{2} g''(0) + o(t^2) = f(a) + t\, df(a)(h) + \frac{t^2}{2} Q(h) + o(t^2)$$

au voisinage de $t = 0$.

(Q) ***Théorème 10.8.*** — *Soient U un ouvert de $\mathbb{R}^n$, $f : U \to \mathbb{R}$ de classe C^2 et $a \in U$ un point critique de f.*

(i) *Si f admet un minimum local en a, alors les valeurs propres de la matrice hessienne $D^2 f(a)$ sont toutes ≥ 0.*

(ii) *Réciproquement, si ces valeurs propres sont toutes > 0, f admet un minimum local en a.*

(iii) *De même, si f admet un maximum local en a, alors les valeurs propres de la matrice hessienne $D^2 f(a)$ sont toutes ≤ 0. Réciproquement, si ces valeurs propres sont toutes < 0, alors f f admet un maximum local en a.*

Démonstration. — Dans cette démonstration, on munit $\mathbb{R}^n$ de la norme euclidienne $\|\cdot\|_2$. D'après le théorème de Schwarz, la matrice hessienne $S = D^2 f(a) \in M_n(\mathbb{R})$ est **symétrique**. Donc, d'après un théorème d'algèbre linéaire (cf. 11.3), S est *diagonalisable dans une base orthonormée* ; il existe donc des réels $\lambda_1 \leq \cdots \leq \lambda_n$ et une base orthonormée[12] $(v_1, \ldots, v_n)$ de $\mathbb{R}^n$ tels que $Sv_i = \lambda_i v_i$ pour $i = 1, \ldots, n$. Alors, pour tout $v = \sum_{i=1}^n x_i v_i$ on a $Q(v) = v \cdot Sv = \sum_{i=1}^n \lambda_i x_i^2$.

Comme $df(a) = 0$, la formule de Taylor donne pour tout t assez petit :

$$f(a + tv_i) = f(a) + \frac{t^2}{2} v_i \cdot Sv_i + t^2 \eta(t) = f(a) + \frac{t^2}{2}(\lambda_i + 2\eta(t))$$

avec $\lim_{t\to 0} \eta(t) = 0$.

Si f admet un minimum local en a, on a donc $\lambda_i \geq -2\eta(t)$ pour tout t assez petit, d'où par passage à la limite $\lambda_i \geq 0$. Ceci prouve (i).

Supposons maintenant $\lambda_1 > 0$. Pour $h = \sum_{i=1}^n h_i v_i$ assez petit, la formule de Taylor donne

$$f(a + h) = f(a) + \frac{1}{2} h \cdot Sh + \|h\|^2 \eta(h) = f(a) + \frac{1}{2} \sum_{i=1}^n \lambda_i h_i^2 + \|h\|^2 \eta(h)$$

avec $\lim_{t\to 0} \eta(t) = 0$. Comme $\lambda_n \geq \cdots \geq \lambda_1$, on a $\sum_{i=1}^n \lambda_i h_i^2 \geq \lambda_1 \sum_{i=1}^n h_i^2 = \lambda_1 \|h\|^2$ d'où

$$f(a + h) \geq f(a) + \|h\|^2 \Big(\frac{\lambda_1}{2} + \eta(h)\Big).$$

Comme $\lambda_1 > 0$, on obtient que $f(a+h) \geq f(a)$ pour tout h tel que $|\eta(h)| \leq \lambda_1/2$. Ceci prouve que f admet un minimum local en a, d'où (ii). Et bien entendu, le cas (iii) se traite de façon analogue (ou en changeant f en $-f$). $\square$

(Q) ***Remarque 10.9***. — En dimension $n = 2$, il est facile de connaître le signe des valeurs propres λ_1 et λ_2 de $S = D^2 f(a)$. En effet, $\lambda_1 \lambda_2 = \mathrm{dét}(S)$ et $\lambda_1 + \lambda_2 = \mathrm{Tr}(S)$ (la *trace* de S, égale à la somme des coefficients diagonaux de S). Donc, si $\mathrm{dét}(S) > 0$ alors λ_1, λ_2 sont non nulles et de même signe, qui est le signe de $\mathrm{Tr}(S)$; dans ce cas, d'après le théorème précédent f admet en a un extremum local (minimum si $\mathrm{Tr}(S) > 0$ et maximum si $\mathrm{Tr}(S) < 0$). Si $\mathrm{dét}(S) < 0$ alors λ_1 et λ_2 sont non nulles et de signe opposé, on dit alors que f admet en a un *point selle* ; dans ce cas, d'après le théorème précédent f n'admet en a ni minimum ni maximum local. Enfin, si $\mathrm{dét}(S) = 0$ alors l'une au moins des valeurs propres est nulle et dans ce cas le théorème précédent ne permet pas de conclure. En résumé, on a la situation suivante :

(i) Si $\mathrm{dét}(S) > 0$ et $\mathrm{Tr}(S) > 0$ alors $\lambda_1, \lambda_2 > 0$ et f admet en a un minimum local.

(ii) Si $\mathrm{dét}(S) > 0$ et $\mathrm{Tr}(S) < 0$ alors $\lambda_1, \lambda_2 < 0$ et f admet en a un maximum local.

(iii) Si $\mathrm{dét}(S) < 0$ alors a est un *point selle* pour f ; ce n'est ni un minimum ni un maximum local.

(iv) Si $\mathrm{dét}(S) = 0$ on ne peut pas conclure sans une étude plus approfondie.

Terminons cette section avec les compléments suivants sur les applications bilinéaires et la différentielle seconde. Ces compléments ne seront pas traités en cours.

***Définition 10.10* (Applications bilinéaires)**. — Soient E, F, G trois $\mathbb{R}$-ev. Une application $\phi : E \times F \to G, (x, y) \mapsto \phi(x, y)$ est dite **bilinéaire** si elle est linéaire en chaque variable, i.e. si pour tout $x, x' \in E$, $y, y' \in F$ et $\mu, \nu \in \mathbb{R}$, on a

(1) $\phi(\mu x + x', y) = \mu\phi(x, y) + \phi(x', y)$

(2) $\phi(x, \nu y + y') = \nu\phi(x, y) + \phi(x, y')$.

---

[12] C.-à-d., $v_i \cdot v_j = 0$ pour $i \neq j$ et $v_i \cdot v_i = 1$.

Bien sûr, ces conditions impliquent (et sont équivalentes à) la condition

$$\phi(\mu x + x', \nu y + y') = \mu\nu\phi(x, y) + \nu\phi(x', y) + \mu\phi(x, y') + \phi(x', y').$$

Si de plus $G = \mathbb{R}$, on dit que ϕ est une **forme** bilinéaire sur $E \times F$. Dans le cas où $F = E$, on dit simplement que ϕ est une forme bilinéaire « sur E », et l'on dit que ϕ est *symétrique* si $\phi(x, y) = \phi(y, x)$ pour tout $x, y \in E$. Par exemple, si $E = \mathbb{R}^n$, le produit scalaire $E \times E \to \mathbb{R}$, $(x, y) \mapsto x \cdot y = \sum_{i=1}^n x_i y_i$ est une forme bilinéaire symétrique sur $E = \mathbb{R}^n$.

On note $\mathrm{Bil}(E \times F, G)$ l'ensemble des applications bilinéaires $E \times F \to G$. On a des isomorphismes canoniques

$$\mathscr{L}(E, \mathscr{L}(F, G)) = \mathrm{Bil}(E \times F, G) = \mathscr{L}(F, \mathscr{L}(E, G))$$

qui sont donnés comme suit. Si ϕ est une application bilinéaire, alors pour tout $x \in E$, l'application $\phi(x, -) : y \mapsto \phi(x, y)$ appartient à $\mathscr{L}(F, G)$ d'après la linéarité en la 2ème variable. De plus, la linéarité en la 1ère variable entraîne que l'application $E \to \mathscr{L}(F, G)$, $x \mapsto \phi(x, -)$ est linéaire : on obtient ainsi une application

$$(\dagger) \qquad \mathrm{Bil}(E \times F, G) \to \mathscr{L}(E, \mathscr{L}(F, G)), \qquad \phi \mapsto \Big(x \mapsto \phi(x, -)\Big).$$

Réciproquement, soit $x \mapsto u(x)$ une application linéaire $E \to \mathscr{L}(F, G)$, alors l'application $E \times F \to G$, $(x, y) \mapsto u(x)(y)$ est linéaire en chaque variable, donc appartient à $\mathrm{Bil}(E \times F, G)$. On obtient ainsi une application

$$(\ddagger) \qquad \mathscr{L}(E, \mathscr{L}(F, G)) \to \mathrm{Bil}(E \times F, G), \qquad u \mapsto \Big(\phi_u : (x, y) \mapsto u(x)(y)\Big).$$

Ces deux applications sont des bijections réciproques, car pour tout ϕ on a $\phi(x, -)(y) = \phi(x, y)$ et pour tout u on a $\phi_u(x, -) = u(x)$. On obtient de même l'isomorphisme $\mathrm{Bil}(E \times F, G) = \mathscr{L}(F, \mathscr{L}(E, G))$.

Définition 10.11 **(Matrice d'une forme bilinéaire sur $\mathbb{R}^n$).** — Soit ϕ une forme bilinéaire sur $\mathbb{R}^n$ et $\mathscr{B} = (e_1, \dots, e_n)$ une base de $\mathbb{R}^n$ (par exemple la base canonique) ; pour tout $x = \sum_{i=1}^n x_i e_i$ et $y = \sum_{j=1}^n y_j e_j$, la bilinéarité donne

$$\phi(x, y) = \phi\Big(\sum_{i=1}^n x_i e_i, \sum_{j=1}^n y_j e_j\Big) = \sum_{i,j=1}^n a_{ij} x_i y_j,$$

où $a_{ij} = \phi(e_i, e_j)$. On dit que la matrice $A = (a_{ij})_{1 \le i,j \le n}$ est la matrice de ϕ dans la base $\mathscr{B}$. Si l'on représente x, y par des vecteurs colonnes X, Y, on a alors la formule matricielle :

$$\phi(X, Y) = {}^t X A Y = \sum_{i,j=1}^n x_i a_{ij} y_j.$$

Réciproquement, pour toute matrice $A \in M_n(\mathbb{R})$, la formule ci-dessus définit une forme bilinéaire ϕ_A sur $\mathbb{R}^n$. On voit que ϕ_A est symétrique ssi $a_{ij} = a_{ji}$ pour tout i, j, i.e. si la matrice A est symétrique (i.e. $A = {}^t A$).

Si $\mathscr{C}$ est une autre base de $\mathbb{R}^n$ et $P = \mathrm{Mat}_{\mathscr{B}}(\mathscr{C})$ la matrice de passage, les coordonnés X et X' d'un vecteur x dans $\mathscr{B}$ et $\mathscr{C}$ sont reliées par $X = PX'$; on a donc

$$\phi(x, y) = {}^t X A Y = {}^t X' \, {}^t P A P Y'$$

et l'on en déduit que la matrice de ϕ dans la base $\mathscr{C}$ est ${}^t PAP$. [13]

Définition 10.12 **(Dérivée seconde, applications C^2).** — Soit U un ouvert de $E = \mathbb{R}^n$ et $f : U \to \mathbb{R}$ de classe C^1. On munit E de la base canonique $\mathscr{B} = (e_1, \dots, e_n)$ et E^* de la base duale $\mathscr{B}^* = (e_1^*, \dots, e_n^*)$. Alors la différentielle $g = df$ est une application continue $U \to E^*$, donnée par ses composantes $g_i = \partial_i f$ pour $i = 1, \dots, n$. Soit $a \in U$.

[13] En particulier, on voit que la symétrie (ou non) de la matrice est indépendante de la base choisie.

(1) On dit que f admet en a une différentielle seconde si l'application $g = df$ est différentiable en a, i.e. si chaque g_i l'est. Dans ce cas, la matrice jacobienne de g est

$$Dg = \begin{pmatrix} \partial_1 g_1 & \cdots & \partial_n g_1 \\ \vdots & \ddots & \vdots \\ \partial_1 g_n & \cdots & \partial_n g_n \end{pmatrix}(a) = \begin{pmatrix} \partial_1\partial_1 f & \cdots & \partial_n\partial_1 f \\ \vdots & \ddots & \vdots \\ \partial_1\partial_n f & \cdots & \partial_n\partial_n f \end{pmatrix}(a) = D^2 f(a),$$

i.e. dg est l'application $U \to \mathscr{L}(E, E^*)$ donnée par la matrice $A_{ij} = \partial g_i/\partial x_j = \partial^2 f/\partial x_j \partial x_i$ i.e. la matrice hessienne de f en a.

De plus, on peut montrer que si df est différentiable en a le théorème de Schwarz est vérifié, i.e. $\partial_i\partial_j f(a) = \partial_j\partial_i f(a)$ pour tout i, j, donc la matrice $D^2 f(a)$ est *symétrique.*

(2) On dit que f est de classe C^2 si l'application $g = df : U \to M_n(\mathbb{R})$ est de classe C^1, i.e. si les dérivées partielles $\partial g_i/\partial x_j = \partial^2 f/\partial x_j \partial x_i$ existent et sont continues sur U. On retrouve ainsi la définition donnée en 10.1.

11. Extrema liés : cas d'une sphère euclidienne. Conséquences

Notons S la sphère unité de $\mathbb{R}^{n+1}$ muni de la norme euclidienne, i.e.

$$S = \{x = (x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} \mid Q(x) = 1\}$$

où $Q(x) = x \cdot x = x_1^2 + \cdots + x_{n+1}^2$, et soit $f : U \to \mathbb{R}$ une application différentiable, où U est un ouvert de $\mathbb{R}^{n+1}$ contenant S. Notons g la restriction de f à S et soit a un point de S en lequel g admet un extremum local, i.e. il existe $r > 0$ tel que $f(a)$ soit maximum ou minimum parmi les valeurs $f(x)$ pour $x \in B(a, r) \cap S$. (De tels extrema existent, puisque g est continue sur le compact S, donc y atteint un maximum (resp. minimum) global.) Comment trouver un tel a ? Que peut-on dire de $df(a)$? On n'a pas nécessairement $df(a) = 0$: par exemple la fonction Q est constante sur S, mais pour tout $a \in S$ on a $\nabla Q(a) = 2a \neq 0$.

(Q) ***Théorème 11.1* (Extrema sur une sphère).** — *Supposons que g admette en $a \in S$ un extremum local. Alors il existe $\mu \in \mathbb{R}$ tel que $\nabla f(a) = \mu\, a$.*

Démonstration. — En prenant une base orthonormée de $\mathbb{R}^{n+1}$ dont a est le dernier vecteur, on se ramène au cas où $a = (0, \dots, 0, 1)$. Notons q la restriction de Q à $\mathbb{R}^n = \{x \in \mathbb{R}^{n+1} \mid x_{n+1} = 0\}$ et B la boule unité ouverte de $\mathbb{R}^n$, i.e.

$$B = \{x \in \mathbb{R}^n \mid q(x) < 1\}.$$

Alors l'application

$$u : B \to \mathbb{R}^{n+1}, \qquad x \mapsto \big(x, z(x)\big),$$

où $z(x) = \sqrt{1 - q(x)}$, est de classe C^1, vérifie $u(0) = a$ et est une bijection de B sur le demi-hémisphère supérieur

$$S_+ = \{x = (x_1, \dots, x_{n+1}) \in S \mid x_{n+1} > 0\}.$$

Si f admet un extremum local en a alors $f \circ u$ admet un extremum local en 0 et donc

$$0 = d(f \circ u)(0) = df(a) \circ du(0).$$

D'autre part, comme z est la composée de $1 - q : B \to \mathbb{R}_+^*$ et de $\sqrt{\cdot} : \mathbb{R}_+^* \to \mathbb{R}_+^*$, on obtient

$$\frac{\partial z}{\partial x_i}(x) = \frac{-x_i}{\sqrt{1 - q(x)}}.$$

Ces dérivées partielles s'annulent pour $x = 0$, donc $dz(0)$ est nulle. D'autre part, comme $u : B \to \mathbb{R}^{n+1}$ est définie par

$$u(x) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \\ z(x) \end{pmatrix}$$

alors $du(0) : \mathbb{R}^n \to \mathbb{R}^{n+1}$ est donnée par la matrice $Du(0) \in M_{n+1,n}(\mathbb{R})$ ci-dessous

$$\left(\begin{array}{cccc} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \\ \hline 0 & \cdots & 0 & 0 \end{array}\right)$$

(la dernière ligne est donnée par $du(0)$ qui est nulle), i.e. $du(0)(e_i) = e_i$ pour $i = 1, \dots, n$. Donc l'image $\operatorname{Im} du(0)$ de $du(0)$ est l'hyperplan de $\mathbb{R}^{n+1}$ défini par l'équation $x_{n+1} = 0$, i.e. orthogonal au vecteur $e_{n+1} = a$. Comme la condition $df(a) \circ du(0) = 0$ signifie que $df(a)$ est nulle sur $\operatorname{Im} du(0)$, i.e. que $\nabla f(a)$ est orthogonal à $\operatorname{Im} du(0)$, on obtient que $\nabla f(a) = \mu\, a$, pour un certain $\mu \in \mathbb{R}$. □

(Q) ***Corollaire 11.2.*** — *Soit $A \in M_n(\mathbb{R})$ une matrice symétrique réelle. Alors A possède au moins une valeur propre réelle.*

Démonstration. — D'après le lemme 10.5, l'application $f(x) = x \cdot Ax$ est de classe C^1, sa différentielle en tout point x étant $h \mapsto 2Ax \cdot h$. Notons g la restriction de f à la sphère unité S. Comme S est compacte, g atteint son maximum en un point $v \in S$. D'après le théorème précédent, il existe $\mu \in \mathbb{R}$ tel que, pour tout $h \in \mathbb{R}^n$ on ait

$$2Av \cdot h = df(v)(h) = \nabla f(v) \cdot h = \mu v \cdot h$$

d'où $Av = \mu v$, i.e. v est un vecteur propre de A pour la valeur propre μ. □

(Q) ***Théorème 11.3.*** — *On munit $\mathbb{R}^n$ du produit scalaire standard. Soit $A \in M_n(\mathbb{R})$ une matrice symétrique réelle. Alors A est diagonalisable dans une base orthonormée, i.e. il existe une base orthonormée $(v_1, \dots, v_n)$ formée de vecteurs propres de A.*

Démonstration. — On procède par récurrence sur n. C'est OK pour $n = 1$, donc on peut supposer $n \geq 2$ et le résultat démontré pour tout espace euclidien E de dimension $< n$. Soit u l'endomorphisme de $\mathbb{R}^n$ défini par A. Soient $\mathscr{B}$ la base canonique, $\mathscr{C}$ une base orthonormée quelconque de $\mathbb{R}^n$ et P la matrice de passage de $\mathscr{B}$ à $\mathscr{C}$. Comme $\mathscr{B}$ et $\mathscr{C}$ sont orthonormées, P vérifie ${}^tP = P^{-1}$ et donc la matrice

$$A' = P^{-1}AP = {}^tPAP$$

de u dans la base $\mathscr{C}$ est encore symétrique.

D'après le corollaire précédent, u possède au moins un vecteur propre v_1 pour une valeur propre μ_1. Quitte à multiplier v_1 par un scalaire, on peut supposer v_1 de norme 1 (pour la norme euclidienne). Soit E l'orthogonal de v_1, i.e.

$$E = \{x \in \mathbb{R}^n \mid x \cdot v_1 = 0\}.$$

On sait que $\mathbb{R}^n = \mathbb{R}v_1 \oplus E$. Montrons que E est stable par u. Soit $y \in E$ et Y le vecteur colonne associé. Comme ${}^tA = A$, on a

$$u(y) \cdot v_1 = {}^tY\,{}^tAv_1 = {}^tY\,Av_1 = \mu_1\,y \cdot v_1 = 0.$$

Ceci prouve que E est stable par u. Notons $u|_E : E \to E$ la restriction de u à E. Soit $(f_2, \dots, f_n)$ une base orthonormée de E et B la matrice de $u|_E$ dans cette base. Alors $\mathscr{C} = (v_1, f_2, \dots, f_n)$ est une base orthonormée de $\mathbb{R}^n$ et la matrice de u dans $\mathscr{C}$ est

$$\left(\begin{array}{c|c} \mu_1 & 0 \\ \hline 0 & B \end{array}\right)$$

donc B est symétrique. Donc par hypothèse de récurrence, il existe une base orthonormée $(v_2, \dots, v_n)$ de E formée de vecteurs propres de $u|_E$, et alors $(v_1, \dots, v_n)$ est une base orthonormée de $\mathbb{R}^n$ formée de vecteurs propres de u. $\square$

Le théorème 11.1 est un cas particulier du théorème suivant, que nous admettrons. D'abord, une définition.

Définition 11.4. — Soient U un ouvert de $\mathbb{R}^{n+1}$ et $F : U \to \mathbb{R}$ une application de classe C^1. L'ensemble

$$\Sigma = \mathscr{V}(F) = \{x \in \mathbb{R}^{n+1} \mid F(x) = 0\}$$

s'appelle l'hypersurface définie par F. On dit que cette hypersurface est une **sous-variété** de $\mathbb{R}^{n+1}$ de classe C^1 si pour tout $a \in \Sigma$ on a $dF(a) \neq 0$. Dans ce cas, l'hyperplan affine $\mathscr{H}$ défini par l'équation $dF(a)(x - a) = 0$, i.e.

$$\mathscr{H} = \{a + h \in \mathbb{R}^{n+1} \mid dF(a)(h) = 0\}$$

s'appelle **l'hyperplan tangent** à Σ en a et se note $T_a\Sigma$.

***Théorème 11.5* (Extrema liés).** — *On se place sous les hypothèses précédentes. Soit $f : U \to \mathbb{R}$ différentiable et soit g la restriction de f à Σ. Si g admet un extremum local en un point a de Σ alors il existe $\mu \in \mathbb{R}$ tel que $df(a) = \mu\, dF(a)$.*

Remarque 11.6. — Le nom « extrema liés » (ou « extrema sous contraintes ») provient du fait qu'on cherche à maximiser ou minimiser $f(x)$ lorsque x est lié à (i.e. contraint de) rester sur l'hypersurface Σ.

12. Champs de vecteurs et équations différentielles

Définition 12.1. — Soient U un intervalle ouvert de $\mathbb{R}$, $v : U \to \mathbb{R}$ de classe C^1, $u \in U$ et t_0 un réel fixé. On considère **l'équation différentielle** $\boxed{x'(t) = v(x(t))}$ avec la « condition initiale » $x(t_0) = u$, c.-à-d. on cherche une fonction dérivable $x : I \to U$, où I est un intervalle ouvert contenant t_0 et aussi grand que possible, telle que $f(t_0) = u$ et $x'(t) = v(x(t))$ pour tout $t \in I$; on dit alors que x est solution de l'équation différentielle $x' = v(x)$ avec la condition initiale $x(t_0) = u$. On verra plus bas (Th. de Cauchy-Lischitz) que sous les hypothèses précédentes, il existe une solution unique. On la note $x(u, t)$, c.-à-d.pour tout $u \in U$ la fonction $t \mapsto x(u, t)$ est l'unique solution de l'équation différentielle $x'(t) = v(x(t))$ telle que $x(u, t_0) = u$. (Elle est définie sur un intervalle ouvert I_u qui dépend de u.)

Remarquons qu'une telle fonction x est continue (car dérivable) ; comme v est C^1 donc continue alors $x' = v \circ x$ est continue donc x est C^1 ainsi que v, donc $x' = v \circ x$ l'est aussi, donc x est de classe C^2.(14)

Dans les exemples ci-dessous on prendra $t_0 = 0$. (On peut toujours s'y ramener en remplaçant la variable t par $t + t_0$.)

Exemples 12.2. — On prend $U = \mathbb{R}$.

(1) Pour $v(x) = a \neq 0$ on obtient l'équation différentielle $x'(t) = a$. Pour la condition initiale $x(0) = u$, la solution est $x(u,t) = at + u$, qui est définie (et de classe C^∞) sur $\mathbb{R}$.

(2) Pour $v(x) = ax$, où $a \neq 0$, on obtient l'équation différentielle $x'(t) = ax(t)$ avec, disons, la condition initiale $x(0) = u$. On voit que la fonction $x(u,t) = u\exp(at)$ est solution. C'est la seule solution car si $z(t)$ est une autre solution définie sur un intervalle ouvert I contenant 0, la fonction $f(t) = z(t)\exp(-at)$ vérifie pour tout $t \in I$:

$$f'(t) = z'(t)\exp(-at) - az(t)\exp(-at) = az(t)\exp(-at) - az(t)\exp(-at) = 0$$

donc $f(t)$ est constante, de valeur $f(0) = z(0) = u$, d'où $z(t) = u\exp(at)$. **Remarque**. On peut aussi trouver la solution comme suit : si $x(0) = 0$ la fonction nulle est solution ; si $x(0) \neq 0$ alors par continuité x est non nulle, donc garde un signe constant $\varepsilon = \pm 1$, sur un intervalle ouvert I contenant 0, et sur cet intervalle l'équation s'écrit $\dfrac{x'(t)}{x(t)} = a$ et le terme de gauche est la dérivée de la fonction $\log(\varepsilon x(t))$, d'où $\log(\varepsilon x(t)) = at + b$ avec $b = \log(\varepsilon u)$, puis $x(t) = u\exp(at)$.

(3) Pour $v(x) = x^2$, on obtient l'équation différentielle $x'(t) = x(t)^2$, avec la condition initiale $x(0) = u$. Si $u = 0$, la fonction nulle est solution. Supposons que $x(t)$ soit une solution avec $x(0) = u \neq 0$. Par continuité x est non nulle, donc garde un signe constant $\varepsilon = \pm 1$, sur un intervalle ouvert I contenant 0, et sur cet intervalle l'équation s'écrit $\dfrac{-x'(t)}{x(t)^2} = -1$ et le terme de gauche est la dérivée de la fonction $1/x(t)$, d'où

$$\frac{1}{x(t)} = -t + \frac{1}{u} = \frac{1 - ut}{u}$$

et donc $x(u,t) = \dfrac{u}{1 - ut}$, qui est de classe C^1 (et même C^∞) sur le plus grand intervalle ouvert I_u contenant 0 mais ne contenant pas $1/u$, i.e. $I_u =]1/u, +\infty[$ si $u < 0$ et $I_u =]-\infty, 1/u[$ si $u > 0$. Remarquons que si $u < 0$ la condition $t > 1/u$ équivaut à $ut < 1$; de même si $u > 0$ la condition $t < 1/u$ équivaut à $ut < 1$. Donc, posant

$$\Omega = \{(t,u) \in \mathbb{R}^2 \mid tu < 1\}$$

on obtient que la fonction de deux variables

$$x(u,t) = \frac{u}{1 - ut}$$

définie sur l'ouvert Ω de $\mathbb{R}^2$, donne la solution de l'équation différentielle $x' = x^2$ avec la condition initiale $x(u,0) = u$, et est de classe C^1 (et même C^∞) sur Ω. Il en est de même dans les deux exemples précédents, où $x(u,t) = at + u$ et $x(u,t) = u\exp(at)$.

(14) En procédant par récurrence, le même raisonnement montre que si v est de classe C^p, avec $p \geq 1$, alors toute solution x est de classe C^{p+1}.

Remarque 12.3. — Dans les deux premiers exemples, $x(t,u)$ est définie sur $\mathbb{R}^2$ tout entier, i.e. on a $I_u = \mathbb{R}$ pour tout u. Mais l'exemple (3) montre que ceci n'est pas toujours le cas en général.[15]

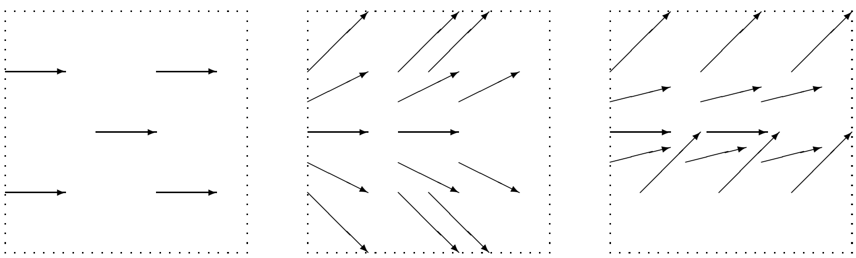
Terminologie. — Une « courbe paramétrée » dans $\mathbb{R}^n$, de classe C^1, n'est rien d'autre qu'une application $\gamma : I \to \mathbb{R}^n$, de classe C^1, où I est un intervalle de $\mathbb{R}$.

Dans ce cas, le *support géométrique* de la courbe γ est le sous-ensemble $\Gamma = \{\gamma(t) \mid t \in I\}$ de $\mathbb{R}^n$. Il contient évidemment beaucoup moins d'information que l'application γ elle-même. Par exemple, toutes les applications $\gamma : \mathbb{R} \to \mathbb{R}$ de classe C^1 telles que $\lim_{-\infty} \gamma = -\infty$ et $\lim_{+\infty} \gamma = +\infty$ ont le même support $\Gamma = \mathbb{R}$.

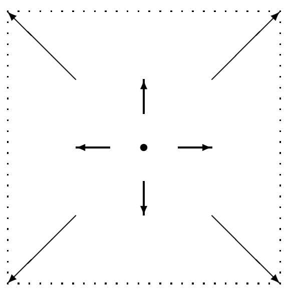
***Définition 12.4* (Champs de vecteurs).** — Soit U un ouvert de $\mathbb{R}^n$. Un *champ de vecteurs* de classe C^1 sur U est une application $v : U \to \mathbb{R}^n$ de classe C^1, qui associe à tout $x \in U$ le vecteur $v(x)$, auquel on pense comme le vecteur $v(x)$ avec son origine placée en x. Tous les champs de vecteurs considérés seront supposés de classe C^1, et l'on écrira simplement « champ de vecteurs ».

Exemples 12.5. — Prenons $U = \mathbb{R}^2$.

(1) Le champ de vecteurs constant $v_0(x,y) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, et les champs de vecteurs $v_1(x,y) = \begin{pmatrix} 1 \\ y \end{pmatrix}$ et $v_2(x,y) = \begin{pmatrix} 1 \\ y^2 \end{pmatrix}$:

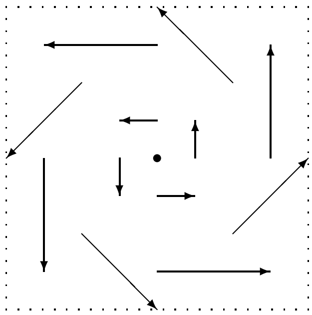


(2) Le champ de vecteurs $v(x_1,x_2) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$.



(3) Le champ de vecteurs $v(x_1,x_2) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}$.

[15] On peut montrer que c'est le cas si $v(x) = ax + b$.



Définition 12.6 **(Équation différentielle associée à un champ de vecteurs)**

Soient U un ouvert de $\mathbb{R}^n$ et v un champ de vecteurs de classe C^1 sur U. Fixant un point $u \in U$ et un réel t_0, on lui associe l'équation différentielle $\boxed{x'(t) = v(x(t))}$ avec la condition initiale $x(t_0) = u$, c.-à-d. on cherche une fonction dérivable $x : I \to U$, où I est un intervalle ouvert contenant t_0 et aussi grand que possible, telle que $f(t_0) = u$ et $x'(t) = v(x(t))$ pour tout $t \in I$. [16] Donc, notant x_i (resp. v_i) les composantes de x (resp. v), on a le système d'équations différentielles

$$\begin{cases} x_1'(t) = v_1(x_1(t), \dots, x_n(t)) \\ \vdots \\ x_n'(t) = v_n(x_1(t), \dots, x_n(t)) \end{cases}$$

et la condition initiale $x_i(t_0) = u_i$ pour $i = 1, \dots, n$. On dit alors que l'application $I \to U$, $t \mapsto x(t)$ est une **courbe intégrale** du champ de vecteurs v, passant par le point u au temps t_0. On verra plus bas (Th. de Cauchy-Lischitz) que sous les hypothèses précédentes, il existe une **unique** courbe intégrale passant par u au temps t_0. On la note $x(u, t)$, c.-à-d.pour tout $u \in U$ la fonction $t \mapsto x(u, t)$ est l'unique solution de l'équation différentielle $x'(t) = v(x(t))$ telle que $x(u, t_0) = u$.

On dit que le point $u_0 \in U$ est un point **stationnaire** du champ de vecteurs v si $v(u_0) = 0$. Dans ce cas, la fonction constante $x : \mathbb{R} \to U$, $t \mapsto u_0$ est une courbe intégrale de v passant par u_0 (et c'est la seule, d'après le théorème de Cauchy-Lipschitz 12.11).

Remarque 12.7. — Les équations différentielles précédentes $x' = v(x)$ sont dites « autonomes » car la variable t n'apparaît pas explicitement au second membre, i.e. le champ de vecteurs v ne dépend que de x et pas de t. Plus généralement, si U est un ouvert de $\mathbb{R}^n$ et I un intervalle de $\mathbb{R}$, on peut considérer une équation différentielle

$$(*) \qquad x'(t) = \begin{pmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{pmatrix} = v(t, x) = \begin{pmatrix} v_1(t, x_1(t), \dots, x_n(t)) \\ \vdots \\ v_n(t, x_1(t), \dots, x_n(t)) \end{pmatrix}$$

où $v : I \times U \to \mathbb{R}^n$ est une application de classe C^1, comme par exemple l'équation $x'(t) = t\,x(t)$.

Mais ceci se ramène à une équation autonome sur l'ouvert $\Omega = I \times U$ de $\mathbb{R}^{n+1}$ en notant $\widetilde{x} = (x_0, x_1, \dots, x_n)$ les points de Ω et en considérant le champ de vecteurs $w : \Omega \to \mathbb{R}^{n+1}$ défini

[16] Comme v est de classe C^1, on voit comme en 12.1 que toute solution x de l'équation différentielle $x'(t) = v(x(t))$ est de classe C^2.

par

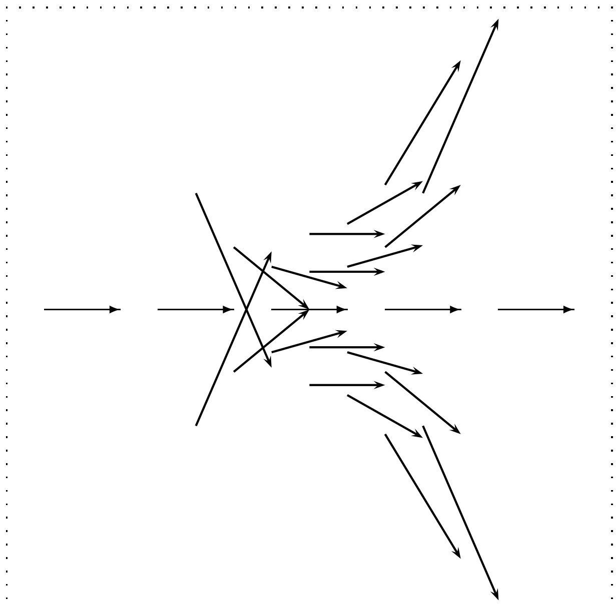
$$w(\widetilde{x}) = w(x_0, x_1, \ldots, x_n) = \begin{pmatrix} 1 \\ v_1(x_0, x_1, \ldots, x_n) \\ \vdots \\ v_n(x_0, x_1, \ldots, x_n) \end{pmatrix}$$

et l'équation différentielle $\widetilde{x}'(t) = w(\widetilde{x}(t))$ avec la condition initiale $\widetilde{x}(t_0) = (t_0, u_1, \ldots, u_n)$. En effet, si $\widetilde{x} : J \to \Omega$ est une solution, où J est un intervalle ouvert contenant t_0 alors pour tout $t \in J$ on a $x_0(t) = t$ et, pour $i = 1, \ldots, n$,

$$x_i'(t) = v_i(x_0(t), x_1(t), \ldots, x_n(t)) = v_i(t, x_1(t), \ldots, x_n(t))$$

donc $x(t) = (x_1(t), \ldots, x_n(t))$ est bien solution de $(*)$ avec la condition initiale $x(t_0) = u$.

Exemple 12.8. — L'équation différentielle $x'(t) = t\,x(t)$ avec la condition initiale $x(0) = u$ conduit au champ de vecteurs $v(t,x) = \begin{pmatrix} 1 \\ tx \end{pmatrix}$ sur $\mathbb{R}^2$, cf. la figure suivante :



Résolvons cette équation différentielle. Si $u = 0$, la fonction nulle est solution. Supposons que $x(t)$ soit une solution avec $x(0) = u \neq 0$. Par continuité x est non nulle, donc garde un signe constant $\varepsilon = \pm 1$, sur un intervalle ouvert I contenant 0, et sur cet intervalle l'équation s'écrit $\dfrac{x'(t)}{x(t)} = t$ et le terme de gauche est la dérivée de la fonction $\log(\varepsilon x(t))$, d'où $\log(\varepsilon x(t)) = t^2/2 + b$ avec $b = \log(\varepsilon u)$, et donc $x(u,t) = u \exp(t^2/2)$. Ayant ainsi trouvé cette solution, on montre l'unicité comme dans l'exemple (2) de 12.2 : soit $z(t)$ une autre solution, alors la fonction $f(t) = z(t)\exp(-t^2/2)$ vérifie

$$f'(t) = z'(t)\exp(-t^2/2) - tz(t)\exp(-t^2/2) = tz(t)\exp(-t^2/2) - tz(t)\exp(-t^2/2) = 0$$

donc $f(t)$ est constante, de valeur $f(0) = z(0) = u$, d'où $z(t) = u\exp(-t^2/2)$. Ceci montre que les courbes intégrales de v sont les applications $x(u,t) = u\exp(t^2/2)$, i.e. $x(u,t)$ est la courbe intégrale de v qui passe par le point $(0,u)$ de $I \times U$ au temps $t = 0$.

Remarque 12.9. — Le même argument que ci-dessus permet de montrer que pour une application continue $a : I \to \mathbb{R}$ et $t_0 \in I$, $u \in \mathbb{R}$, l'équation différentielle $x'(t) = a(t)x(t)$ avec la condition initiale $x(t_0) = u$ admet pour unique solution l'application

$$x : U \times I \to \mathbb{R}, \qquad (u,t) \mapsto u \exp\left(\int_{t_0}^{t} a(s)ds\right).$$

Exercices 12.10. — (1) Pour chacun des exemples de 12.2, dessiner plusieurs courbes intégrales du champ de vecteurs v, correspondant à différentes conditions initiales u.

(2) Déterminer les courbes intégrales des champs de vecteurs des exemples 12.5.

***Remarque 12.11* (Théorème de Cauchy-Lipschitz)**. — Soient U un ouvert de $\mathbb{R}^n$ et $v : U \to \mathbb{R}^n$ un champ de vecteurs de classe C^1. Résumons les résultats observés sur tous les exemples précédents, à savoir :

(i) Pour tout point $p = (t_0, u) \in \mathbb{R} \times U$ il existe une unique courbe intégrale de v passant par p au temps t_0, i.e. il existe une unique application $x : I_u \to U$ de classe C^2 vérifiant $x(t_0) = u$ et $x'(t) = v(x(t))$ pour tout $t \in I_u$, où I_u est un intervalle ouvert contenant t_0 et aussi grand que possible.

(ii) Fixant une fois pour toutes le point t_0 (par exemple prenant $t_0 = 0$), notons $x(u,t)$ la solution telle que $x(u, t_0) = u$. Alors l'application $(u,t) \mapsto x(u,t)$ est définie sur un ouvert Ω de $\mathbb{R} \times U$ (contenant $U \times \{t_0\}$) et est de classe C^1 sur Ω.

Ces résultats sont toujours valables sous l'hypothèse que v est C^1 : ceci est le contenu de l'important théorème de Cauchy-Lipschitz (qui est au programme de L3).

Terminons cette section avec l'exemple suivant.

***Exemple 12.12* (Champ de gradients)**. — Soient U un ouvert de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$ de classe C^2. Alors l'application $U \to \mathbb{R}^n$, $x \mapsto \nabla(f)(x)$ est un champ de vecteurs sur U de classe C^1. Ses courbes intégrales sont appelées les lignes de gradient de f.

13. Une application : l'équation de transport en dimension 1

Commençons avec le problème très simple suivant : on a des particules dans $\mathbb{R}^n$ qui sont transportées par un fluide en mouvement avec un vecteur vitesse constant $\overrightarrow{v} \in \mathbb{R}^n$.(17) Sachant que la situation à l'instant $t = 0$ est décrite par une certaine fonction $f_0(x)$ qui décrit le nombre (ou la densité) de particules à la position $x \in \mathbb{R}^n$, on note $f(x,t)$ la fonction décrivant la situation au point x à l'instant t et l'on veut exprimer, si possible, $f(x,t)$ en fonction de la condition initiale $f(x,0) = f_0(x)$ et du vecteur vitesse $\overrightarrow{v}$ du fluide.

La situation qui se trouvait en $t_0 = 0$ en un point x a été déplacé par le fluide et se trouve à l'instant t au point $x + t\overrightarrow{v}$; on a donc pour tout $x \in \mathbb{R}^n$ et $t \in \mathbb{R}$ l'égalité

$$f(x + t\overrightarrow{v}, t) = f(x, 0)$$

ce qui permet de répondre à la question de départ : pour tout (x,t) on a $\boxed{f(x,t) = f_0(x - t\overrightarrow{v}).}$

Considérons maintenant le cas un peu plus compliqué (et physiquement plus réaliste) où la vitesse du fluide en un point x dépend de x, i.e. est donnée par un champ de vecteurs $v(x)$. Alors,

(17) Ou bien on peut imaginer des plateaux repas posés sur un tapis roulant.

adoptant le langage de la physique, [18] on peut dire : « la situation qui se trouvait au point $x \in \mathbb{R}^n$ à l'instant t se trouve à un instant $t + dt$ (avec dt infiniment petit) au point $x + dt\, v(x)$ », d'où l'égalité

$$f(x, t) = f(x + dt\, v(x), t + dt). \tag{1}$$

Supposons la fonction $f(x, t)$ cherchée de classe C^1. Alors, en négligeant les termes d'ordre > 1, on peut écrire que $f(x + dt\, v(x), t + dt)$ égale :

$$f(x,t) + df(x,t)\begin{pmatrix} v_1(x)dt \\ \vdots \\ v_n(x)dt \\ dt \end{pmatrix} = f(x,t) + \frac{\partial f}{\partial t}(x,t)dt + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x,t)v_i(x)dt \tag{2}$$

et donc en notant $\nabla_x f(x, t)$ le vecteur « gradient partiel » relativement aux coordonnées x_i, i.e.

$$\nabla_x f(x,t) = \begin{pmatrix} \partial f/\partial x_1 \\ \vdots \\ \partial f/\partial x_n \end{pmatrix}(x,t),$$

on voit que (1) et (2) donnent l'équation aux dérivées partielles (EDP) :

$$\frac{\partial f}{\partial t}(x,t) + v(x) \cdot \nabla_x f(x,t) = 0 \tag{3}$$

qu'on appelle *équation de transport.* Nous allons utiliser les résultats de la section précédente pour étudier cette EDP dans le cas où $n = 1$. Dans ce cas, l'équation de transport s'écrit simplement

$$\frac{\partial f}{\partial t}(x,t) + v(x)\frac{\partial f}{\partial x}(x,t) = 0 \tag{$*$}$$

Comme le champ des vitesses est v, la trajectoire de chaque particule est une courbe intégrale de l'équation différentielle $x'(t) = v(x(t))$. Plus précisément, notant $x(u, t)$ la solution de $x'(t) = v(x(t))$ avec la condition initiale $x(0) = u$, qui est définie sur un intervalle ouvert I_u contenant 0 et aussi grand que possible, chaque particule se trouvant à l'instant $t_0 = 0$ à la position $x(0) = u$ se trouvera au temps $t \in I_u$ à la position $x(u, t)$. Étant donné un point arbitraire (x, t) de $\mathbb{R}^2$, la question est de savoir si x égale $x(u, t)$ pour un certain couple (u, t), dans ce cas on aura $f(x, t) = f(u, 0) = f_0(u)$ et il s'agit donc d'exprimer u en fonction de x et t.

Étudions l'EDP $(*)$ pour les valeurs du champ de vitesse $v(x)$ données dans l'exemple 12.2.

(1) Si $v(x) = a$, i.e. si la vitesse est constante, la courbe intégrale $x(u, t)$ telle que $x(0) = u$ est $x(u, t) = at + u$. On voit que l'application $(u, t) \mapsto (at + u, t)$ est un C^1-difféomorphisme de $\mathbb{R}^2$ sur lui-même, dont la réciproque est donnée par $(x, t) \mapsto (u = x - at, t)$. Pour tout (x, t), on a donc $f(x, t) = f_0(x - at)$ comme déjà vu au début de cette section.

(2) Si $v(x) = ax$, la courbe intégrale $x(t)$ telle que $x(0) = u$ est $x(u, t) = ue^{at}$. L'application $(u, t) \mapsto (ue^{at}, t)$ est un C^1-difféomorphisme de $\mathbb{R}^2$ sur lui-même, dont la réciproque est donnée par $(x, t) \mapsto (u = xe^{-at}, t)$. Pour tout (x, t), on a donc $f(x, t) = f_0(xe^{-at})$.

(3) Si $v(x) = x^2$, la courbe intégrale $x(u, t)$ telle que $x(u, 0) = u$ est donné par $x(u, t) = \dfrac{u}{1 - tu}$, pour t vérifiant $tu < 1$. Posons $\Omega = \{(u, t) \in \mathbb{R}^2 \mid tu < 1\}$. Pour tout $(u, t) \in \Omega$, on a

$$t\, x(t, u) = \frac{tu}{1 - tu} = -1 + \frac{1}{1 - tu} > -1$$

[18] On verra plus loin une justification rigoureuse de ce raisonnement.

donc l'application $\phi : (u,t) \mapsto (x(t,u),t)$ envoie Ω dans l'ouvert $V = \{(x,t) \in \mathbb{R}^2 \mid tx > -1\}$. De plus, pour t fixé la fonction homographique

$$u \mapsto x = \frac{u}{1-tu}$$

admet pour réciproque la fonction $u = \dfrac{x}{1+tx}$,[19] qui est bien définie si $tx > -1$. On obtient donc que $\phi : \Omega \to V$ est un C^1-difféomorphisme, donc la réciproque est donnée par

$$\phi^{-1} : (x,t) \mapsto \Big(\frac{x}{1+tx}, t\Big).$$

Pour tout $(x,t) \in V$, posant $u = \dfrac{x}{1+tx}$, on obtient donc

$$f(x,t) = f(x(t,u),t) = f(u,0) = f_0\Big(\frac{x}{1+tx}\Big).$$

14. Formes différentielles de degré 1 et intégrales de chemins

Cette section aurait sa place dans ce chapitre, mais faute de temps elle sera traitée après le chapitre sur les intégrales multiples.

[19] En effet, $x(1-tu) = u$ équivaut à $u(1+tx) = x$.