Calcul différentiel

Peter Haïssinsky, Université d'Aix-Marseille

2019–2020

Contents

Introduction

L'objet du calcul différentiel est de comparer localement des fonctions compliquées à des fonctions plus simples afin d'en tirer des informations globales sur la fonction d'origine. Par exemple, si $f : \mathrm{I} \to \mathbb{R}$ est une application numérique, définie et dérivable sur un intervalle $\mathrm{I} =]a, b[$, alors cela signifie que pour $x_0 \in \mathrm{I}$, la limite suivante existe

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} .$$

Autrement dit, au voisinage de x_0, on peut écrire

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + (x - x_0)\varepsilon(x)$$

où $\varepsilon : \mathrm{I} \to \mathbb{R}$ est une fonction qui tend vers 0 vers x_0. Cela signifie que le comportement de f proche de x_0 est similaire à l'application affine $x \mapsto f(x_0) + f'(x_0)(x - x_0)$.

A partir de ces données locales (ressembler à une application affine au voisinage de chacun des points), on peut ensuite en déduire des informations globales: le prototype est le théorème des accroissements finis qui permet de contrôler $|f(x) - f(y)|$ par les valeurs de f' prises dans l'intervalle borné par x et y. D'autres applications comprennent l'établissement du tableau de variations d'une fonction et le calcul des extrema.

L'objet de ce cours est de développer le calcul différentiel dans le contexte des fonctions $f : \mathbb{R}^p \to \mathbb{R}^q$, pour $p, q \geq 1$ quelconques. Une partie du cours consiste à étendre ce qui est connu des fonctions numériques et une autre à analyser les nouveautés provenant du contexte de la grande dimension. La plupart des résultats restent valides en dimension infinie, mais cela requiert la maîtrise de quelques points techniques qui ne ferait qu'obscurcir une théorie déjà suffisamment technique.

Dans un premier temps, on étudiera les notions de différentiabilité des fonctions, en donnant des méthodes de calculs des exemples et en établissant quelques propriétés élémentaires. Ensuite, on établira un théorème fondamental permettant le passage du local au global: le théorème de la moyenne (analogue du théorème des accroissements finis). On pourra en déduire des propriétés plus fines des applications différentiables.

Dans un second temps, on étudiera la différentiation d'ordre supérieur, avec un attachement particulier à la différentielle seconde. Cela nous conduira naturellement à la notion de développements limités.

Le chapitre suivant portera sur des théorèmes centraux de la théorie: les théorèmes d'inversion locale, des fonctions implicites et du rang constant. Ces résultats sont de nature nouvelle et nous permettront dans le chapitre suivant d'étudier les sous-variétés de $\mathbb{R}^n$, qui correspondent à des généralisations des courbes dans le plan.

Le bon contexte pour étudier les fonctions de plusieurs variables est celui des espaces vectoriels réels normés. L'introduction se conclut donc par un rappel sur cette notion, ainsi que sur la continuité qui est un préambule au calcul différentiel.

Chapitre I. — contexte et rappels

1 Espaces vectoriels normés

1.1 Généralités

Définition 1.1. — *Un espace vectoriel (réel) normé (evn)* $(\mathrm{E}, \|\cdot\|)$ *est un espace vectoriel* E *muni d'une norme, c'est-à-dire d'une application* $\|\cdot\| : \mathrm{E} \to \mathbb{R}_+$ *qui vérifie les propriétés suivantes:*

(N1) pour tout vecteur $x \in \mathrm{E}$, $\|x\| = 0$ *si et seulement si* $x = 0$;

(N2) pour tout vecteur $x \in \mathrm{E}$ *et tout scalaire* $\lambda \in \mathbb{R}$, *on a* $\|\lambda x\| = |\lambda|\|x\|$ *(homogénéité)*

(N3) pour tous vecteurs $x, y \in \mathrm{E}$, *on a* $\|x + y\| \leq \|x\| + \|y\|$ *(inégalité triangulaire)*

On déduit de l'inégalité triangulaire la minoration souvent pratique, appelé parfois inégalité triangulaire inverse

$$|\|x\| - \|y\|| \leq \|x - y\| .$$

Exemples 1.2. — *On travaille sur* $\mathrm{E} = \mathbb{R}^n$, $n \geq 1$. *Si* $x \in \mathbb{R}^n$, *on note ses coordonnées* $x = (x_j)_{1 \leq j \leq n}$ *où* $x_j \in \mathbb{R}$, $j = 1, \ldots, n$. *Les normes les plus courantes sont les normes suivantes.*

1. $\|x\|_1 = \sum_{1 \leq j \leq n} |x_j|$;

2. $\|x\|_2 = \sqrt{\sum_{1 \leq j \leq n} x_j^2}$;

3. $\|x\|_\infty = \max\{|x_j|, 1 \leq j \leq n\}$.

Une autre source d'exemples provient de produits scalaires.

1.2 Espaces euclidiens

Soit E un espace vectoriel réel.

Définition 1.3. — *Une forme bilinéaire sur un espace vectoriel est une application* $b : \mathrm{E} \times \mathrm{E} \to \mathbb{R}$ *telle que*

- *pour tout* $y \in \mathrm{E}$, $b_1 : x \mapsto b(x, y)$ *est linéaire;*
- *pour tout* $x \in \mathrm{E}$, $b_2 : y \mapsto b(x, y)$ *est linéaire;*

On dit qu'elle est symétrique si $b(x, y) = b(y, x)$ *pour tous* $x, y \in \mathrm{E}$. *De plus, on dit que* b *est*

- *positive si* $b(x, x) \geq 0$ *pour tout* $x \in \mathrm{E}$,
- *définie si* $b(x, x) = 0$ *seulement si* $x = 0$.

On note $q(x) = b(x, x)$ *et on dit que* q *est la forme quadratique associée à* b.

Lorsque b est symétrique, on peut la retrouver à partir de q avec la formule suivante

$$b(x, y) = \frac{1}{2}(q(x + y) - q(x) - q(y)) = \frac{1}{4}(q(x + y) - q(x - y)) .$$

Définition 1.4. — *Un produit scalaire est une forme bilinéaire définie positive. Un espace euclidien est un espace vectoriel de dimension finie sur* $\mathbb{R}$ *muni d'un produit scalaire. On note le produit scalaire* $b(x, y) = \langle x, y \rangle$.

Exemple 1.5. — *Sur* $\mathbb{R}^n$*, on pose*

$$\langle \mathrm{X}, \mathrm{Y}\rangle = {}^t\mathrm{X}\mathrm{Y} = \sum_{j=1}^{n} x_j y_j \,.$$

On vérifie que c'est bien une forme bilinéaire symétrique. De plus, on a $\langle \mathrm{X}, \mathrm{X}\rangle = \sum x_j^2$ *donc elle est positive, et* $\langle \mathrm{X}, \mathrm{X}\rangle = 0$ *seulement si* $x_j = 0$ *pour tout* j*, donc elle est bien définie positive.*

Théorème 1.6. — *Un espace euclidien est naturellement un espace vectoriel normé avec* $\|x\| = \sqrt{q(x)}$*.*

Pour montrer ce résultat, on commence par montrer l'inégalité de Cauchy-Schwarz.

Inégalité de Cauchy-Schwarz. — *Soit* (E, b) *un espace vectoriel euclidien de forme quadratique* q*. On a, pour tout* $x, y \in \mathrm{E}$*,*

$$b(x,y)^2 \leq q(x)q(y)\,,$$

avec égalité si et seulement si x *et* y *sont liés.*

Démonstration de l'inégalité de Cauchy-Schwarz. — Pour $t \in \mathbb{R}$, on a $q(xt + y) \geq 0$. Or, en utilisant la bilinéarité de b et la symétrie, on montre que $q(tx + y) = q(x)t^2 + 2tb(x,y) + q(y)$. Comme ce polynôme quadratique (en t) ne prend que des valeurs positives, il ne peut s'annuler au plus qu'une fois, donc son discriminant doit être négatif :

$$b(x,y)^2 - q(x) \cdot q(y) \leq 0\,.$$

Le cas d'égalité a lieu si on n'a qu'une seule racine, qui est donc, en supposant $x \neq 0$, $t = -b(x,y)/q(x)$, auquel cas on a $q(tx + y) = 0$ soit $tx + y = 0$. Donc x et y sont liés. Sinon, $x = 0$ et x et y sont automatiquement liés (avec égalité). ■

Démonstration du théorème 1.6. — On pose, pour $x \in \mathrm{E}$, $\|x\| = \sqrt{q(x)}$. Pour montrer que $\|\cdot\|$ est une norme, il suffit de vérifier l'inégalité triangulaire. Il suffit de traiter le cas $x + y \neq 0$ de sorte que $\|x + y\| \neq 0$.

$$\begin{aligned}(\|x\| + \|y\|) - \|x+y\| &= \frac{q(x) + q(y) + 2\sqrt{q(x)q(y)} - q(x+y)}{\|x+y\| + \|x\| + \|y\|} \\ &\geq \frac{2\sqrt{q(x)q(y)} - 2b(x,y)}{\|x+y\| + \|x\| + \|y\|} \geq 0\,.\end{aligned}$$

■

1.3 Topologie d'un evn

Rappelons qu'une distance sur un ensemble E est une application $d : \mathrm{E} \times \mathrm{E} \to \mathbb{R}_+$ qui vérifie les propriétés suivantes:

(D1) pour tous $x, y \in \mathrm{E}$, $d(x,y) = 0$ si et seulement si $x = y$;

(D2) pour tous $x, y \in \mathrm{E}$, on a $d(x,y) = d(y,x)$ (symétrie)

(D3) pour tous $x, y, z \in \mathrm{E}$, on a $d(x,y) \leq d(x,z) + d(z,y)$ (inégalité triangulaire)

Comme pour les normes, on déduit de l'inégalité triangulaire l'inégalité triangulaire inverse

$$|d(x,z) - d(z,y)| \leq d(x,y)\,.$$

Si $(\mathrm{E}, \|\cdot\|)$ est un evn, alors $d(x,y) = \|x - y\|$ définit une distance sur E.

Si $x \in \mathrm{E}$ et $r \geq 0$,, on appelle *boule ouverte de centre* x *et de rayon* r (resp. *boule fermée de centre* x *et de rayon* r) l'ensemble

$$\mathrm{B}(x,r) = \{y \in \mathrm{E},\ \|x-y\| < r\} \quad \Big(\text{resp.}\quad \overline{\mathrm{B}(x,r)} = \{y \in \mathrm{E},\ \|x-y\| \leq r\}\Big)\,.$$

Un *ouvert* U de E est une partie telle que, tout point $x \in \mathrm{U}$ est le centre d'une boule ouverte de rayon $r > 0$ contenue dans U. En particulier, toute boule ouverte est un ouvert. Un *fermé* F de E est une partie de E dont le complémentaire est ouvert. Enfin, un *voisinage* d'un point $x \in \mathrm{E}$ est une partie V de E qui contient une boule ouverte non vide centrée en x. Notons qu'un voisinage n'est pas forcément ouvert. Si E est de dimension finie, un sous-ensemble X de E est *compact* s'il est fermé et borné. Une autre caractérisation affirme que toute suite à valeurs dans X admet une sous-suite convergente dans X.

2 Rappels sur la continuité

Soient $(\mathrm{E}, \|\cdot\|_{\mathrm{E}})$ et $(\mathrm{F}, \|\cdot\|_{\mathrm{F}})$ deux espaces normés. Soient U un ouvert de E, $f : \mathrm{U} \to \mathrm{F}$ une application et prenons $x_0 \in \mathrm{U}$ et $y_0 \in \mathrm{F}$. On dit que $f(x)$ tend vers y_0 quand x tend vers x_0 si $\|f(x) - y_0\|_{\mathrm{F}}$ tend vers 0 quand $\|x - x_0\|$ tend vers 0. Autrement dit, pour toute marge d'erreur $\varepsilon > 0$, il existe une précision $\delta > 0$ telle que si $x \in \mathrm{U}$ et $\|x - x_0\|_{\mathrm{E}} \leq \delta$ alors $\|f(x) - y_0\|_{\mathrm{F}} \leq \varepsilon$. On appelle y_0 la limite de f en x_0 et on écrit

$$\lim_{x \to x_0} f(x) = y_0 \,.$$

Notons que la limite, si elle existe, est forcément unique.

On rappelle le critère de Cauchy qui permet de montrer la convergence d'une suite sans connaître la limite *a priori* dans un espace métrique complet —ce qui est le cas des espaces vectoriels normés de dimension finie: *une suite $(x_n)_n$ est convergente si, et seulement si, elle est de Cauchy*: pour tout $\varepsilon > 0$, il existe un rang $n_0 \geq 1$ tel que, si $p, q \geq n_0$ alors $\|x_p - x_q\| < \varepsilon$.

2.1 Généralités

Définition 2.1 (continuité). — *Une fonction $f : \mathrm{U} \to \mathrm{F}$ définie sur un ouvert d'un evn* E *à valeurs dans un evn* F *est continue en un point $x_0 \in \mathrm{E}$ si $\lim_{x_0} f = f(x_0)$. La fonction f est continue sur* U *si elle est continue en chaque point.*

En général, on montre que la fonction f est continue en un point x_0 en montrant que $\lim_{x_0} f(x_n) = f(x_0)$ quelle que soit la suite $(x_n)_n$ qui tend vers x_0.

Exemple.— Soit E un evn; l'application $f : \mathrm{E} \to \mathbb{R}$ définie par $f(x) = \|x\|$ est continue. En effet, l'inégalité triangulaire inverse affirme $|\,\|x\| - \|y\|\,| \leq \|x - y\|$ pour tous $x, y \in \mathrm{E}$. Par conséquent, si $\varepsilon > 0$ est fixé et si $\|x - y\| \leq \varepsilon$, alors $|f(x) - f(y)| \leq \varepsilon$.

Propriété 2.2 (stabilité). — *Soient $f, g : \mathrm{U} \to \mathrm{F}$ deux fonctions définies d'un ouvert d'un evn* E *à valeurs dans un evn* F *et soit $x_0 \in \mathrm{U}$. On suppose que f et g sont continues en x_0. On a les propriétés suivantes.*

1. *Pour tous $\lambda, \mu \in \mathbb{R}$, l'application $\lambda f + \mu g$ est continue en x_0.*
2. *Si $\mathrm{F} = \mathbb{R}$, alors fg est continue en x_0, ainsi que f/g si $g(x_0) \neq 0$.*
3. *Si $h : \mathrm{V} \to \mathrm{G}$ est une application définie sur un ouvert* V *de* F *contenant $f(x_0)$, continue en ce point, et à valeurs dans un evn* G*, alors $h \circ f$ est continue au point x_0.*

La notion de continuité en plusieurs variables peut être compliquée à étudier. Nous illustrons ce propos par quelques exemples.

Exemples de fonctions continues. — Une fonction polynomiale $f : \mathbb{R}^n \to \mathbb{R}$ est une combinaison linéaire de monômes, c'est-à-dire de termes de la forme $x_{j_1}^{k_1} \ldots x_{j_p}^{k_p}$ où $0 \leq p \leq n$, $1 \leq j_1 < j_2 < \ldots < j_p \leq n$ et $k_j \in \mathbb{N}$. Les projections $\pi_k : (x_j) \mapsto x_k$ sont continues, donc les monômes aussi et leurs combinaisons linéaires de même. Donc les fonctions polynomiales sont continues.

Considérons l'application $f : \mathbb{R}^2 \to \mathbb{R}$ définie par $f(x, y) = x^2 y$ si $y \geq 0$ et $f(x, y) = 0$ si $y < 0$. Alors f est continue sur $\mathbb{R}^2$. En effet, si (y_n) tend vers 0 par valeurs positives et (x_n) tend vers x, alors

$f(x_n, y_n)$ tend vers 0.

Exemples de fonctions non continues. — On s'intéresse aux comportements au voisinage de l'origine des fonctions suivantes. Notons $g : \mathbb{R}^2 \to \mathbb{R}$ définie par

$$g(x,y) = \frac{xy}{x^2+y^2} \text{ si } (x,y) \neq (0,0) \quad \text{et} \quad g(0,0) = 0\,.$$

Cette application est continue le long des axes $\{x = 0\}$ et $\{y = 0\}$ puisque $g(x,0) = g(0,y) = g(0,0) = 0$. Cependant, pour $\alpha \neq 0$, on a

$$g(x, \alpha x) = \frac{\alpha}{\alpha^2+1} \neq 0$$

donc g n'est pas continue en 0.

Notons $h : \mathbb{R}^2 \to \mathbb{R}$ définie par

$$h(x,y) = \frac{x^2 y}{(x^4+y)^2} \text{ si } (x,y) \neq (0,0) \quad \text{et} \quad h(0,0) = 0\,.$$

Cette application est continue le long des droites passant par l'origine $y = \alpha x$, $\alpha \in \mathbb{R}$ et $x = 0$ car

$$h(x, \alpha x) = \frac{\alpha x^3}{(x^4+\alpha x)^2} = \frac{\alpha x}{(x^3+\alpha)^2}$$

si $x \neq 0$ mais

$$h(x, x^4) = \frac{1}{4x^2}$$

qui est divergent. Donc h n'est pas continue à l'origine.

Il faut donc retenir que la continuité est une propriété globale: il ne suffit pas de considérer uniquement les axes, ou les droites, etc. Il faut un contrôle simultané dans toutes les directions. Une condition nécessaire pour la continuité d'une fonction $f : \mathbb{R}^2 \to \mathbb{R}$ en un point (x_0, y_0) est que, quelle que soit la fonction continue $\varphi : \mathbb{R} \to \mathbb{R}$ telle que $\varphi(x_0) = y_0$, la fonction $x \mapsto f(x, \varphi(x))$ est continue en x_0. Mais cette condition n'est pas suffisante.

Théorème 2.3. — *L'image réciproque d'un ouvert (resp. fermé) est ouverte (resp. fermée) par une fonction continue.*

La propriété suivante est très utile.

Proposition 2.4. — *Une application continue $f : \mathrm{K} \to \mathbb{R}$ définie sur un ensemble compact atteint ses bornes.*

2.2 Normes équivalentes, applications linéaires en dimension finie

Définition 2.5. — *Soit E un espace vectoriel réel. Deux normes $\|\cdot\|$ et $\|\cdot\|'$ sont équivalentes s'il existe une constante $\mathrm{C} \geq 1$ telle que, pour tout $x \in \mathrm{E}$,*

$$\frac{1}{\mathrm{C}}\|x\| \leq \|x\|' \leq \mathrm{C}\|x\|\,.$$

Cela implique que les limites pour une norme ou une autre coïncident.

Proposition 2.6. — *En dimension finie, toutes les normes sont équivalentes.*

Démonstration. — On se fixe une base de E afin de l'identifier avec $\mathbb{R}^n$, où $n = \dim \mathrm{E}$. Notons (e_j) la base canonique de $\mathbb{R}^n$. Nous allons montrer que toute norme est équivalente à la norme $\|\cdot\|_\infty$. Notons donc N une norme sur $\mathbb{R}^n$.

Montrons tout d'abord que $\mathrm{N} : (\mathbb{R}^n, \|\cdot\|_\infty) \to \mathbb{R}$ est continue. Pour cela, on observe

$$\mathrm{N}(x) = \mathrm{N}\left(\sum_{j=1}^n x_j \mathrm{e}_j\right) \leq \sum_{j=1}^n |x_j| \mathrm{N}(\mathrm{e}_j) \leq \left(\sum_{j=1}^n \mathrm{N}(\mathrm{e}_j)\right) \cdot \|x\|_\infty\,.$$

Par conséquent, si on note $\mathrm{M} = \sum_{j=1}^{n} \mathrm{N}(\mathrm{e}_j)$, alors on a $\mathrm{N}(x) \leq \mathrm{M}\|x\|_\infty$. Du coup, l'inégalité triangulaire inverse nous dit

$$|\mathrm{N}(x) - \mathrm{N}(y)| \leq \mathrm{N}(x-y) \leq \mathrm{M}\|x-y\|_\infty$$

ce qui nous montre la continuité de N.

Soit $\mathrm{S} = \{x \in \mathbb{R}^n,\ \|x\|_\infty = 1\}$. D'après le lemme 2.7 ci-dessous, c'est un compact de $(\mathbb{R}^n, \|\cdot\|_\infty)$. La fonction $\mathrm{N} : \mathrm{S} \to \mathbb{R}$ atteint donc ses bornes: il existe $a, b \in \mathrm{S}$ tels que, pour tout $x \in \mathrm{S}$, on a $\mathrm{N}(a) \leq \mathrm{N}(x) \leq \mathrm{N}(b)$. Du coup, si $x \in \mathbb{R}^n$, alors, en utilisant l'homogénéité de la norme N et le fait que $x/\|x\|_\infty \in \mathrm{S}$, on obtient

$$\mathrm{N}(a)\|x\|_\infty \leq \mathrm{N}(x) \leq \mathrm{N}(b)\|x\|_\infty\,.$$

■

Lemme 2.7. — *L'ensemble* $\mathrm{S} = \{x \in \mathbb{R}^n,\ \|x\|_\infty = 1\}$ *est un compact de* $(\mathbb{R}^n, \|\cdot\|_\infty)$.

Démonstration. — L'ensemble S est l'ensemble des points de $[0,1]^n$ qui ont au moins une coordonnée égale à ± 1, donc est borné. Par ailleurs, l'application $f : x \in \mathbb{R}^n \mapsto \|x\|_\infty$ est continue car on a $|\,\|x\|_\infty - \|y\|_\infty\,| \leq \|x-y\|_\infty$ pour tous $x, y \in \mathbb{R}^n$ par l'inégalité triangulaire inverse; par conséquent $\mathrm{S} = f^{-1}(\{1\})$ est fermé. Nous avons montré que S est fermé et borné dans un espace vectoriel normé de dimension finie, donc S est compact. ■

Passons maintenant aux applications linéaires. Si E et F sont deux espaces vectoriels, on notera $\mathcal{L}(\mathrm{E},\mathrm{F})$ l'espace des applications linéaires de E dans F, et plus simplement $\mathcal{L}(\mathrm{E}) = \mathcal{L}(\mathrm{E},\mathrm{E})$.

Proposition 2.8. — *Soit* $\mathrm{L} : \mathrm{E} \to \mathrm{F}$ *une application linéaire entre evn. Les propositions suivantes sont équivalentes.*

1. *L'application* L *est continue sur* E.
2. *L'application est* L *est continue en* 0.
3. *Il existe* $\mathrm{M} > 0$ *telle que* $\|\mathrm{L}(x)\|_\mathrm{F} \leq \mathrm{M}\|x\|_\mathrm{E}$ *pour tout* $x \in \mathrm{E}$.

Démonstration. — Si L est continue sur E, alors L est continue en 0. Si L est continue en 0, alors, il existe $\delta > 0$ telle que si $\|x\|_\mathrm{E} \leq \delta$ alors $\|\mathrm{L}(x)\|_\mathrm{F} \leq 1$. Du coup, en notant $\mathrm{M} = 1/\delta$, on constate que, si $x \in \mathrm{E} \setminus \{0\}$, alors $\|x\delta/\|x\|_\mathrm{E}\|_\mathrm{E} \leq \delta$ donc $\|\mathrm{L}(x)\|_\mathrm{F} \leq \mathrm{M}\|x\|_\mathrm{E}$. Enfin, si $\|\mathrm{L}(x)\|_\mathrm{F} \leq \mathrm{M}\|x\|_\mathrm{E}$ pour tout $x \in \mathrm{E}$, alors

$$\|\mathrm{L}(x) - \mathrm{L}(y)\|_\mathrm{F} = \|\mathrm{L}(x-y)\|_\mathrm{F} \leq \mathrm{M}\|x-y\|_\mathrm{E}$$

donc L est continue sur tout E. ■

Corollaire 2.9. — *Si* E *est de dimension finie, alors toute application linéaire* $\mathrm{L} : \mathrm{E} \to \mathrm{F}$ *est continue.*

Démonstration. — Comme toutes les normes sont équivalentes sur E, on se fixe une base de E pour l'identifier à $\mathbb{R}^n$, $n = \dim \mathrm{E}$ et on travaille avec la norme $\|\cdot\|_\infty$. Notons $\mathrm{M} = \sum_{j=1}^{n} \|\mathrm{L}(\mathrm{e}_j)\|_\mathrm{F}$. Si $x = \sum x_j \mathrm{e}_j$ alors on a par linéarité et homogénéité

$$\|\mathrm{L}(x)\|_\mathrm{F} \leq \sum |x_j| \|\mathrm{L}(\mathrm{e}_j)\|_\mathrm{F} \leq \mathrm{M}\|x\|_\infty\,.$$

■

Proposition 2.10. — *Soient* E, F *des evn avec* E *de dimension finie. Pour* $\mathrm{L} \in \mathcal{L}(\mathrm{E},\mathrm{F})$*, on note* $\|\mathrm{L}\| = \sup_{\|x\|_\mathrm{E}=1} \|\mathrm{L}(x)\|_\mathrm{F}$*. Alors* $(\mathcal{L}(\mathrm{E},\mathrm{F}), \|\cdot\|)$ *est un evn et* $\|\cdot\|$ *s'appelle la norme d'opérateur. On a de plus les propriétés suivantes.*

1. *Pour tout* $x \in \mathrm{E}$*, on a* $\|\mathrm{L}(x)\|_\mathrm{F} \leq \|\mathrm{L}\| \cdot \|x\|_\mathrm{E}$ *et s'il existe* $\mathrm{M} \geq 0$ *telle que* $\|\mathrm{L}(x)\|_\mathrm{F} \leq \mathrm{M}\|x\|_\mathrm{E}$ *pour tout* $x \in \mathrm{E}$*, alors* $\|\mathrm{L}\| \leq \mathrm{M}$.
2. *Si* E, F, G *sont des evn de dimension finie alors, pour toutes* $\mathrm{L} \in \mathcal{L}(\mathrm{E},\mathrm{F})$ *et* $\mathrm{M} \in \mathcal{L}(\mathrm{F},\mathrm{G})$*, on a* $\|\mathrm{M} \circ \mathrm{L}\| \leq \|\mathrm{M}\| \times \|\mathrm{L}\|$.

Démonstration. — Comme L est continue, il existe $\mathrm{M}_0 \geq 0$ tel que $\|\mathrm{L}(x)\|_{\mathrm{F}} \leq \mathrm{M}_0\|x\|_{\mathrm{E}}$ pour tout $x \in \mathrm{E}$, donc $\|\mathrm{L}\| = \sup\{\|\mathrm{L}(x)\|_{\mathrm{F}}, x \in \mathrm{S}\}$ est bien définie (et majorée par M_0), où $\mathrm{S} = \{x \in \mathrm{E},\ \|x\|_{\mathrm{E}} = 1\}$.

Notons que, pour tout $x \in \mathrm{E}\backslash\{0\}$, on a $x/\|x\|_{\mathrm{E}} \in \mathrm{S}$, donc $\|\mathrm{L}(x)\|_{\mathrm{F}} \leq \|\mathrm{L}\| \cdot \|x\|_{\mathrm{E}}$. Si $\|\mathrm{L}(x)\|_{\mathrm{F}} \leq \mathrm{M}\|x\|_{\mathrm{E}}$ pour tout $x \in \mathrm{E}$, alors, pour $x \in \mathrm{S}$, on a donc $\|\mathrm{L}(x)\|_{\mathrm{F}} \leq \mathrm{M}$, ce qui implique $\|\mathrm{L}\| \leq \mathrm{M}$.

Vérifions que $\|\cdot\|$ est bien une norme.

Si $\|\mathrm{L}\| = 0$ alors, pour tout $x \in \mathrm{E}$, on a $\|\mathrm{L}(x)\| \leq 0 \times \|x\|_{\mathrm{E}} = 0$, soit $\mathrm{L}(x) = 0$. Ceci signifie $\mathrm{L} = 0$ et montre (N1). Si $\lambda \in \mathbb{R}$, alors $\|\lambda\mathrm{L}(x)\|_{\mathrm{F}} = |\lambda| \cdot \|\mathrm{L}(x)\|_{\mathrm{F}}$ donc

$$\|\lambda\mathrm{L}\| = \sup_{x\in\mathrm{S}} \|\lambda\mathrm{L}(x)\|_{\mathrm{F}} = |\lambda| \sup_{x\in\mathrm{S}} \|\mathrm{L}(x)\|_{\mathrm{F}} = |\lambda| \cdot \|\mathrm{L}\|$$

impliquant (N2). Enfin, si $\mathrm{L}_1, \mathrm{L}_2 \in \mathcal{L}(\mathrm{E}, \mathrm{F})$, on a

$$\|(\mathrm{L}_1 + \mathrm{L}_2)(x)\|_{\mathrm{F}} \leq \|\mathrm{L}_1(x)\|_{\mathrm{F}} + \|\mathrm{L}_2(x)\|_{\mathrm{F}} \leq (\|\mathrm{L}_1\| + \|\mathrm{L}_2\|)\|x\|_{\mathrm{E}}\,.$$

Du coup, $\|\mathrm{L}_1 + \mathrm{L}_2\| \leq \|\mathrm{L}_1\| + \|\mathrm{L}_2\|$ et (N3) est vérifiée.

On montre le dernier point:

$$\|(\mathrm{M} \circ \mathrm{L})(x)\|_{\mathrm{G}} \leq \|\mathrm{M}\| \cdot \|\mathrm{L}(x)\|_{\mathrm{F}} \leq \|\mathrm{M}\| \cdot \|\mathrm{L}\| \cdot \|x\|_{\mathrm{E}}$$

par conséquent $\|\mathrm{M} \circ \mathrm{L}\| \leq \|\mathrm{M}\| \cdot \|\mathrm{L}\|$. ∎

Corollaire 2.11. — *Soient* E, F *des evn avec* E *de dimension finie et* $\mathrm{L} \in \mathcal{L}(\mathrm{E}, \mathrm{F})$. *Si*

$$\lim_{x(\neq 0)\to 0} \|\mathrm{L}(x)\|_{\mathrm{F}}/\|x\|_{\mathrm{E}} = 0$$

alors $\mathrm{L} = 0$.

De manière équivalente, la condition du corollaire peut s'exprimer des deux maniéres suivantes:

$$\|\mathrm{L}(x)\|_{\mathrm{F}} = o(\|x\|_{\mathrm{E}}) \Longrightarrow \mathrm{L} = 0$$

ou: si, pour tout $\varepsilon > 0$, il existe $\delta > 0$, tel que $\|x\|_{\mathrm{E}}| \leq \delta \Rightarrow \|\mathrm{L}(x)\|_{\mathrm{F}} \leq \varepsilon\|x\|_{\mathrm{E}}$ alors $\mathrm{L} = 0$.

Démonstration. — Supposons que $\lim_{x(\neq 0)\to 0} \|\mathrm{L}(x)\|_{\mathrm{F}}/\|x\|_{\mathrm{E}} = 0$. Prenons $\varepsilon > 0$; il existe $\delta > 0$ telle que si $\|x\|_{\mathrm{E}} \leq \delta$ alors $\|\mathrm{L}(x)\|_{\mathrm{F}} \leq \varepsilon\|x\|_{\mathrm{E}}$. Du coup, si $x \in \mathrm{E}$ est de norme 1, alors $\|\delta x\|_{\mathrm{E}} \leq \delta$ donc $\delta\|\mathrm{L}(x)\|_{\mathrm{F}} = \|\mathrm{L}(\delta x)\|_{\mathrm{F}} \leq \|\delta x\|_{\mathrm{E}} \leq \delta\varepsilon$ et donc $\|\mathrm{L}(x)\|_{\mathrm{F}} \leq \varepsilon$. Ceci implique $\|\mathrm{L}\| \leq \varepsilon$. Comme ε est arbitraire, $\|\mathrm{L}\| = 0$, donc $\mathrm{L} = 0$. ∎

Remarque 2.12. — *Faisons quelques observations qui découlent des résultats précédents quand on travaille dans* $\mathbb{R}^n$.

1. *En dimension finie, toutes les normes sont équivalentes. Donc on peut travailler avec celle qui est la plus commode. Quand on travaille avec des applications linéaires (p.ex. des endomorphismes), la norme d'opérateur est souvent plus pratique.*

2. *Les projections* $\pi_k : \mathbb{R}^n \to \mathbb{R}$, $x = (x_j)_j \mapsto x_k$ *sont continues. Du coup, si* $f : \mathbb{R}^p \to \mathbb{R}^q$ *est continue alors les composantes* $f_k = \pi_k \circ f$ *sont continues aussi. En fait, la réciproque est aussi vraie: si toutes les fonctions composantes* $(f_j)_j$ *sont continues en* x, *alors* f *est continue en* x *aussi. En effet, soit* $x \in \mathbb{R}^p$ *et fixons-nous* $\varepsilon > 0$. *Pour chaque* $j \in \{1, \ldots, p\}$, *il existe* $\delta_j > 0$ *telle que, si* $\|x - y\| \leq \delta_j$ *alors* $|f_j(x) - f_j(y)| \leq \varepsilon$. *Notons* $\delta = \min\{\delta_j\}$ *de sorte que* $\delta > 0$ *et si* $\|x - y\| \leq \delta$ *alors* $\|x - y\| \leq \delta_j$ *pour chaque* j, *donc* $\|f(x) - f(y)\|_\infty \leq \varepsilon$. *Ceci implique la continuité.*

Du coup, l'étude de la continuité d'une fonction se ramène à celle de fonctions à valeurs réelles.

3. *Un polynôme est une combinaison linéaire de produits de projections donc continue. En particulier,* $\det : \mathcal{L}(\mathrm{E}) \to \mathbb{R}$ *est continue. Comme* $\mathrm{GL}(\mathrm{E}) = \{\mathrm{L} \in \mathcal{L}(\mathrm{E}),\ \det(\mathrm{E}) \neq 0\}$*, on en déduit que* $\mathrm{GL}(\mathrm{E})$ *est un ouvert de* $\mathcal{L}(\mathrm{E})$*.*

4. *Le corollaire 2.11 dit qu'une application linéaire qui tend trop vite vers zéro en zéro est forcément l'application nulle.*

On énonce des propriétés similaires pour les applications multinéaires; leurs démonstrations sont des adaptations proches des précédentes (et sont donc omises).

Définition 2.13 (application multilinéaire). — *Soient un entier* $k > 0$ *et des espaces vectoriels réels* $(\mathrm{E}_1, \ldots, \mathrm{E}_k, \mathrm{F})$*. Une application* $f : \mathrm{E}_1 \times \ldots \times \mathrm{E}_k \to \mathrm{F}$ *est dite multilinéaire (ou plus précisément : k-linéaire) si elle est linéaire en chaque variable, c'est-à-dire si, pour des vecteurs* $x_1, ..., x_k, x'_i$ *et des scalaires* λ, μ*, on a*

$$f(x_1, \ldots, x_{i-1}, \lambda x_i + \mu x'_i, x_{i+1}, \ldots, x_k) = \lambda f(x_1, \ldots, x_i, \ldots, x_k) + \mu f(x_1, \ldots, x'_i, \ldots x_k).$$

On note $\mathcal{L}(\mathrm{E}_1, \ldots, \mathrm{E}_k; \mathrm{F})$ *l'espace des applications k-linéaires* $f : \mathrm{E}_1 \times \ldots \times \mathrm{E}_k \to \mathrm{F}$*.*

Proposition 2.14. — *Soient un entier* $k > 0$ *et des evn de dimension finie* $(\mathrm{E}_1, \ldots, \mathrm{E}_k, \mathrm{F})$*. Une application multilinéaire* $f : \mathrm{E}_1 \times \ldots \times \mathrm{E}_k \to \mathrm{F}$ *est continue. Plus précisément, si on note* $\|f\| = \sup\{\|f(x_1, \ldots, x_k)\|, \|x_1\|_{\mathrm{E}_1} = \ldots = \|x_k\|_{\mathrm{E}_k} = 1\}$*, alors, pour tout* $x = (x_1, \ldots, x_k)$*, on a*

$$\|f(x_1, \ldots, x_k)\|_{\mathrm{F}} \leq \|f\| \|x_1\|_{\mathrm{E}_1} \cdots \|x_k\|_{\mathrm{E}_k}.$$

L'espace $(\mathcal{L}(\mathrm{E}_1, \ldots, \mathrm{E}_n; \mathrm{F}), \|.\|)$ *est un espace vectoriel normé, de dimension finie si tous les espaces* $\mathrm{E}_1, \ldots, \mathrm{E}_k$ *et* F *sont de dimension finie.*

Chapitre II. — APPLICATIONS DIFFERENTIABLES

Dorénavant, tous nos espaces vectoriels sont de dimension finie. On note $(\mathrm{E}, ||.||_{\mathrm{E}})$ et $(\mathrm{F}, ||.||_{\mathrm{F}})$ des espaces vectoriels normés sur $\mathbb{R}$, U un ouvert non vide de E et $f : \mathrm{U} \to \mathrm{F}$ une application. Par un choix de bases pour E et F, on se ramène à $f : \mathbb{R}^p \to \mathbb{R}^q$.

3 Définitions, exemples et premières propriétés

Définition 3.1 (différentiabilité). — *Une application* $f : \mathrm{U} \to \mathrm{F}$ *est différentiable au point* $a \in \mathrm{U}$ *s'il existe une application linéaire* $\mathrm{L} \in \mathcal{L}(\mathrm{E}, \mathrm{F})$ *et une application* $\varepsilon : \mathrm{U} \to \mathrm{F}$ *telles que*

$$f(x) = f(a) + \mathrm{L}(x - a) + \|x - a\|_{\mathrm{E}}\varepsilon(x)$$

et où $\lim_{x\to a} \varepsilon(x) = 0$.

Autrement dit, si on se fixe $r > 0$ de sorte que $\mathrm{B}(a, r) \subset \mathrm{U}$ et h avec $\|h\|_{\mathrm{E}} \leq r$, alors

$$f(a + h) = f(a) + \mathrm{L}(h) + \|h\|_{\mathrm{E}}\rho(h)$$

où $\rho : \mathrm{B}(0, r) \to \mathrm{F}$ vérifie $\lim_{h\to 0} \rho(h) = 0$.

Une autre formulation consiste à écrire que, pour toute marge d'erreur $\varepsilon > 0$, il existe un degré de précision $\delta > 0$ telle que, si $\|h\|_{\mathrm{E}} \leq \delta$ alors $\|f(a + h) - f(a) - \mathrm{L}(h)\|_{\mathrm{F}} \leq \varepsilon\|h\|_{\mathrm{E}}$.

Lemme 3.2. — *Si* f *est différentiable en* a *alors l'application linéaire* L *est unique et s'appelle la différentielle de* f *au point* a, *ou la dérivée. On la note* $\mathrm{L} = \mathrm{D}_a f$ *ou* $f'(a)$.

Démonstration. — Supposons que l'on ait deux applications linéaires L_1 et L_2 de sorte qu'il existe $\varepsilon_1, \varepsilon_2$ telles que $\lim_0 \varepsilon_1 = \lim_0 \varepsilon_2 = 0$ et

$$\begin{cases} f(a + h) = f(a) + \mathrm{L}_1(h) + \|h\|_{\mathrm{E}}\varepsilon_1(h) \\ f(a + h) = f(a) + \mathrm{L}_2(h) + \|h\|_{\mathrm{E}}\varepsilon_2(h) \end{cases}$$

En faisant la différence, on obtient

$$\|(\mathrm{L}_1 - \mathrm{L}_2)(h)\|_{\mathrm{F}} = \|h\|_{\mathrm{E}} \times \|\varepsilon_1(h) - \varepsilon_2(h)\|_{\mathrm{F}}$$

pour tout h. D'après le corollaire 2.11, cela implique $\mathrm{L}_1 = \mathrm{L}_2$. ■

Définition 3.3. — *On dit que* $f : \mathrm{U} \to \mathrm{F}$ *est différentiable sur* U *si* f *est différentiable en tout point de* U. *On obtient alors* $\mathrm{D}f : \mathrm{U} \to \mathcal{L}(\mathrm{E}, \mathrm{F})$. *On dit que* f *est continûment différentiable (ou* $\mathcal{C}^1$*) au point* $a \in \mathrm{U}$ *si* f *est différentiable au voisinage de* a *et* $\mathrm{D}f$ *est continue au point* a. *On dit que* f *est de classe* $\mathcal{C}^1$ *dans* U *si* f *est continûment différentiable en chaque point de* U.

3.1 Exemples

1. *Si* f *est constante, alors* f *est différentiable, même* $\mathcal{C}^1$, *et* $\mathrm{D}_a f = 0$.

En effet, on a $f(a + h) = f(a) + 0 + 0$.

2. *Si* f *est la restriction d'une application linéaire, alors* f *est de classe* $\mathcal{C}^1$ *et* $\mathrm{D}_a f = f$ *pour tout* $a \in \mathrm{U}$.

En effet, on a $f(a + h) = f(a) + f(h) + 0$ avec f linéaire. Le corollaire 2.9 montre que $\mathrm{D}f = f$ est continue.

3. *Si* $f : \mathrm{E}_1 \times \mathrm{E}_2 \to \mathrm{F}$ *est bilinéaire alors* f *est différentiable et* $\mathrm{D}_{(a_1,a_2)}f(h, k) = f(a_1, k) + f(h, a_2)$.

On munit $\mathrm{E} = \mathrm{E}_1 \times \mathrm{E}_2$ de la norme $\|(a, b)\|_{\mathrm{E}} = \max\{\|a\|_{\mathrm{E}_1}, \|b\|_{\mathrm{E}_2}\}$. On écrit

$$f(a_1 + h, a_2 + k) = f(a_1, a_2) + f(h, a_2) + f(a_1, k) + f(h, k)\,.$$

On vérifie que $\mathrm{L} : (h, k) \mapsto f(h, a_2) + f(a_1, k)$ est bien linéaire et $\|f(h, k)\| \leq \|f\|\|h\|_{\mathrm{E}_1}\|k\|_{\mathrm{E}_2} \leq \|(h, k)\|^2_{\mathrm{E}} \times \|f\|$ d'après la proposition 2.14.

4. *Si* $(\mathrm{E}, \|\cdot\|)$ *est un espace euclidien, alors* $\mathrm{N} : x \mapsto \|x\|^2$ *est différentiable et* $\mathrm{D}_x\mathrm{N}(h) = 2\langle x, h\rangle$.

Comme ci-dessus, on a $\mathrm{N}(x+h) = \mathrm{N}(x) + 2\langle x, h\rangle + \|h\|^2$.

5. *Soit* E *un evn; l'application* $f : u \in \mathrm{GL}(\mathrm{E}) \mapsto u^{-1}$ *est différentiable et* $\mathrm{D}_u f(h) = -u^{-1} \circ h \circ u^{-1}$ *pour tout* $u \in \mathrm{GL}(\mathrm{E})$.

On munit $\mathcal{L}(\mathrm{E})$ de la norme d'opérateur et on rappelle que $\mathrm{GL}(\mathrm{E})$ est un ouvert de $\mathcal{L}(\mathrm{E})$. Commençons par étudier la différentiabilité de f en l'identité Id. On étudie donc $(\mathrm{Id}+h)^{-1}$, $h \in \mathcal{L}(\mathrm{E})$. En analogie avec la formule $(1+x)^{-1} = \sum_{n\geq 0}(-1)^n x^n$, on s'intéresse à la suite d'applications linéaires $u_n = \sum_{0\leq k\leq n}(-1)^k h^k$ où $h^k = h \circ \ldots \circ h$ k fois. Si $\|h\| < 1$, alors cette série est normalement convergente puisque

$$\sum_{0\leq k\leq n} \|(-1)^k h^k\| \leq \sum_{0\leq k\leq n} \|h\|^k \leq \frac{1}{1-\|h\|}\,.$$

Donc (u_n) est convergente car $\mathcal{L}(\mathrm{E})$ est de dimension finie et complet. Vérifions maintenant que $(1+h)^{-1} = \sum_{k\geq 0}(-1)^k h^k$: pour cela, il suffit de montrer que $\lim(\mathrm{Id}+h)\circ u_n = \mathrm{Id}$. Or, on a

$$(\mathrm{Id}+h)\circ u_n = \sum_{0\leq k\leq n}(-1)^k h^k + \sum_{0\leq k\leq n}(-1)^k h^{k+1} = \mathrm{Id} + (-1)^n h^{n+1}\,.$$

Comme $\|(-1)^n h^{n+1}\| \leq \|h\|^{n+1}$ et $\|h\| < 1$, on a bien montré que $(\mathrm{Id}+h)$ est inversible et on a calculé $(\mathrm{Id}+h)^{-1}$. L'application $h \mapsto -h$ est linéaire et $\sum_{k\geq 2}(-1)^k h^k = \|h\|\varepsilon(h)$ avec $\lim_0 \varepsilon = 0$ puisque $\varepsilon(h) = (h^2/\|h\|)\sum_{k\geq 0}(-1)^k h^k = (h^2/\|h\|)(\mathrm{Id}+h)^{-1}$.

Du coup, on vient d'établir $\mathrm{D}_{\mathrm{Id}} f(h) = -h$.

Si $u \in \mathrm{GL}(\mathrm{E})$, alors il existe $\delta > 0$ telle que $\mathrm{B}(u,\delta) \subset \mathrm{GL}(\mathrm{E})$. Prenons $h \in \mathcal{L}(\mathrm{E})$, avec $\|h\| < \delta$. On a $(u+h)^{-1} = (\mathrm{Id}+u^{-1}h)^{-1}u^{-1}$ donc

$$\begin{aligned}(u+h)^{-1} &= \left(\mathrm{Id} - u^{-1}h + \|h\|\frac{(u^{-1}h)^2}{\|h\|}(\mathrm{Id}+u^{-1}h)^{-1}\right)u^{-1} \\ &= u^{-1} - u^{-1}hu^{-1} + \|h\|\frac{(u^{-1}h)^2}{\|h\|}(\mathrm{Id}+u^{-1}h)^{-1}u^{-1}\end{aligned}$$

donc on en déduit que f est différentiable au point u et $\mathrm{D}_u f(h) = -u^{-1}hu^{-1}$.

3.2 Remarques

1. Si $p = q = 1$, on retrouve bien la définition standard: $f(a+h) = f(a) + f'(a)h + h\varepsilon(h)$. Plus généralement, si $f : \mathbb{R} \to \mathrm{F}$ est différentiable au point a, alors la dérivée est de la forme $h \in \mathbb{R} \mapsto hv$ où $v \in \mathrm{F}$ est un vecteur. Du coup, on peut identifier $\mathrm{D}_a f$ avec v.

2. La différentiabilité d'une application et la différentielle ne dépend pas des normes choisies car elles sont toutes équivalentes: si $\|\cdot\|_{\mathrm{E}}$ et $\|\cdot\|'_{\mathrm{E}}$ (resp. $\|\cdot\|_{\mathrm{F}}$ et $\|\cdot\|'_{\mathrm{F}}$) sont des normes sur E (resp. sur F), alors il existe $\mathrm{C} \geq 1$ telle que $(1/\mathrm{C})\|\cdot\|_{\mathrm{E}} \leq \|\cdot\|'_{\mathrm{E}} \leq \mathrm{C}\|\cdot\|_{\mathrm{E}}$ et $(1/\mathrm{C})\|\cdot\|_{\mathrm{F}} \leq \|\cdot\|'_{\mathrm{F}} \leq \mathrm{C}\|\cdot\|_{\mathrm{F}}$. Donc, si $f : (\mathrm{U}, \|\cdot\|_{\mathrm{E}}) \to (\mathrm{F}, \|\cdot\|_{\mathrm{F}})$ est différentiable alors $f : (\mathrm{U}, \|\cdot\|'_{\mathrm{E}}) \to (\mathrm{F}, \|\cdot\|'_{\mathrm{F}})$ est aussi différentiable. En effet, on écrit $f(a+h) = f(a) + \mathrm{D}_a f(h) + \|h\|_{\mathrm{E}}\varepsilon_h(h)$ avec $\lim_{h\to 0}\|\varepsilon(h)\|_{\mathrm{F}} = 0$ et on doit vérifier

$$\lim_{h\to 0}\left\|\frac{\|h\|_{\mathrm{E}}}{\|h\|'_{\mathrm{E}}}\varepsilon(h)\right\|'_{\mathrm{F}} = 0,\ \text{qui découle de}\ \left\|\frac{\|h\|_{\mathrm{E}}}{\|h\|'_{\mathrm{E}}}\varepsilon(h)\right\|'_{\mathrm{F}} \leq \mathrm{C}\|\varepsilon(h)\|'_{\mathrm{F}} \leq \mathrm{C}^2\|\varepsilon(h)\|_{\mathrm{F}}.$$

3. On a l'interprétation géométrique suivante: Si $\mathrm{E} = \mathbb{R}^p$ et $\mathrm{F} = \mathbb{R}^q$, on représente le graphe Γ de f dans $(\mathbb{R}^{p+q}, \|\cdot\|_\infty)$ par l'ensemble des couples $(x, f(x))$. Si f est différentiable au point a, on note $\mathrm{T}_a\Gamma$ l'espace affine décrit par $\{(x, f(a) + \mathrm{D}_a f(x-a)),\ x \in \mathbb{R}^p\}$. Cet espace est tangent au point $(a, f(a))$ à Γ au sens suivant:

$$\lim_{x\to a}\frac{d((x,f(x)),\mathrm{T}_a\Gamma)}{d((x,f(x)),(a,f(a)))} = 0\,.$$

En effet,

$$d((x,f(x)),\mathrm{T}_a\Gamma) \leq d((x,f(x)),(x,f(a){+}\mathrm{D}_af(x{-}a))) \leq \|f(x){-}f(a){-}\mathrm{D}_af(x{-}a)\| \leq \|x{-}a\|\|\varepsilon(x{-}a)\|$$

et $d((x,f(x)),(a,f(a))) = \max\{\|x-a\|,\|f(x)-f(a)\|\} \geq \|x-a\|$, donc

$$\frac{d((x,f(x)),\mathrm{T}_a\Gamma)}{d((x,f(x)),(a,f(a)))} \leq \frac{\|x-a\|\|\varepsilon(x-a)\|}{\|x-a\|} \leq \|\varepsilon(x-a)\|\,.$$

4. La notion de dérivabilité et celle de dérivée au point a, comme la notion de limite au point a dont elles sont issues, sont des notions *locales* (il en est de même de la notion de continuité au point a). *Localité* signifie que si deux fonctions f et g coïncident sur un voisinage de a et que f est différentiable en a, alors g aussi et $\mathrm{D}_af = \mathrm{D}_ag$.

3.3 Propriétés

On donne des outils pour vérifier qu'une transformation est différentiable et pour donner des méthodes générales de calcul. On établit ensuite deux propriétés simples découlant de la différentiabilité.

3.3.1 Propriétés de stabilité

Proposition 3.4. — *Soient $\lambda,\mu \in \mathbb{R}$. Si $f,g : \mathrm{U} \to \mathrm{F}$ sont différentiables au point a, alors $\lambda f + \mu g$ est aussi différentiable au point a et $\mathrm{D}_a(\lambda f+\mu g) = \lambda\mathrm{D}_af+\mu\mathrm{D}_ag$. Si f et g sont de classes $\mathcal{C}^1$, alors $\lambda f+\mu g$ aussi.*

Démonstration. — On écrit

$$\begin{aligned}(\lambda f+\mu g)(a+h) &= \lambda f(a+h)+\mu g(a+h)\\ &= \lambda f(a)+\lambda\mathrm{D}_af(h)+\lambda\|h\|\varepsilon_f(h)+\mu g(a)+\mu\mathrm{D}_ag(h)+\mu\|h\|\varepsilon_g(h)\\ &= (\lambda f+\mu g)(a)+(\lambda\mathrm{D}_af+\mu\mathrm{D}_ag)(h)+\|h\|(\lambda\varepsilon_f(h)+\mu\varepsilon_g(h))\,.\end{aligned}$$

■

Proposition 3.5. — *Soient $\mathrm{E},\mathrm{F},\mathrm{G}$ trois evn, $\mathrm{U}\subset\mathrm{E}$ et $\mathrm{V}\subset\mathrm{F}$ des ouverts. Si $f : \mathrm{U}\to\mathrm{F}$ et $g : \mathrm{V}\to\mathrm{G}$ sont des applications différentiables au point $a\in\mathrm{U}$ et au point $f(a)\in\mathrm{V}$ alors $g\circ f$ est différentiable et $\mathrm{D}_a(g\circ f) = \mathrm{D}_{f(a)}g\circ\mathrm{D}_af$. Si f,g sont de classes $\mathcal{C}^1$ alors $g\circ f$ est aussi de classe $\mathcal{C}^1$ sur $\mathrm{U}\cap f^{-1}(\mathrm{V})$.*

Démonstration. — On écrit

$$f(a+h) = f(a)+(\mathrm{D}_af(h)+\|h\|\varepsilon_f(h))$$

de sorte que

$$\begin{aligned}(g\circ f)(a+h) &= (g\circ f)(a)+\mathrm{D}_{f(a)}g(\mathrm{D}_af(h)+\|h\|\varepsilon_f(h))+\|\mathrm{D}_af(h)+\|h\|\varepsilon_f(h)\|\varepsilon_g(\mathrm{D}_af(h)+\|h\|\varepsilon_f(h))\\ &= (g\circ f)(a)+(\mathrm{D}_{f(a)}g\circ\mathrm{D}_af)(h)+\|h\|\mathrm{D}_{f(a)}g(\varepsilon_f(h))+\|\mathrm{D}_af(h)+\|h\|\varepsilon_f(h)\|\varepsilon_g(\mathrm{D}_af(h)+\|h\|\varepsilon_f(h))\,.\end{aligned}$$

On vérifie que $h\mapsto(\mathrm{D}_{f(a)}g\circ\mathrm{D}_af)(h)$ est dans $\mathcal{L}(\mathrm{E},\mathrm{G})$ et

$$\lim_{h\to0}\frac{\|h\|\mathrm{D}_{f(a)}g(\varepsilon_f(h))+\|\mathrm{D}_af(h)+\|h\|\varepsilon_f(h)\|\varepsilon_g(\mathrm{D}_af(h)+\|h\|\varepsilon_f(h))}{\|h\|} = 0$$

en utilisant le fait que $\|\mathrm{D}_af(h)\|\leq\|\mathrm{D}_af\|\cdot\|h\|$. En effet, on en déduit que

$$\begin{cases}\|\mathrm{D}_{f(a)}g(\varepsilon_f(h))\|\leq\|\mathrm{D}_{f(a)}g\|\cdot\|\varepsilon_f(h)\|\to0 \text{ quand } h\to0\,,\\ \lim_{h\to0}\varepsilon_g(\mathrm{D}_af(h)+\|h\|\varepsilon_f(h))=0\\ \|\mathrm{D}_af(h)+\|h\|\varepsilon_f(h)\|\leq(\|\mathrm{D}_af\|+\|\varepsilon_f(h)\|)\cdot\|h\|\,.\end{cases}$$

■

Proposition 3.6. — *Soient* $\mathrm{E}, \mathrm{F}_1, \ldots, \mathrm{F}_k$ *des evn et* $f : \mathrm{U} \to \mathrm{F}_1 \times \ldots \times \mathrm{F}_k$ *une fonction définie par* $x \mapsto f(x) = (f_1(x), \ldots, f_k(x))$*. Soit* $a \in \mathrm{U}$*; la fonction* f *est différentiable au point* a *si et seulement si toutes les composantes* f_j *sont différentiables au point* a *et*

$$\mathrm{D}_a f(h) = \big(\mathrm{D}_a f_1(h), \ldots, \mathrm{D}_a f_k(h)\big) \,.$$

De même, f *est de classe* $\mathcal{C}^1$ *si et seulement si tous les* f_j *le sont.*

En particulier, une fonction $f : \mathrm{U} \to \mathbb{R}^n$ *est différentiable au point* $a \in \mathrm{U}$ *si et seulement si chaque composante est différentiable.*

Remarque 3.7. — *Cette proposition nous dit que les problèmes de différentiabilité d'une fonction se ramène à ceux de fonctions à valeurs réelles.*

Démonstration. — Comme les projections $\pi_j : \mathrm{F}_1 \times \ldots \times \mathrm{F}_k \to \mathrm{F}_j$ sont linéaires, elles sont différentiables et $\mathrm{D}\pi_j = \pi_j$. Or $f_j = \pi_j \circ f$ donc si f st différentiable au point a, il est en de même pour les composantes et $\mathrm{D}_a f_j = \pi_j \circ \mathrm{D}_a f$.

Réciproquement, on suppose chaque composante différentiable. On écrit

$$\begin{aligned}
f(a+h) &= \big(f_1(a+h), \ldots, f_k(a+h)\big)\\
&= f(a) + \big(\mathrm{D}_a f_1(h) + \|h\|\varepsilon_1(h), \ldots, \mathrm{D}_a f_k(h) + \|h\|\varepsilon_k(h)\big)\\
&= f(a) + \big(\mathrm{D}_a f_1(h), \ldots, \mathrm{D}_a f_k(h)\big) + \|h\| \big(\varepsilon_1(h), \ldots, \varepsilon_k(h)\big) \,.
\end{aligned}$$

■

Proposition 3.8. — *On suppose que* $f : \mathrm{U} \to \mathrm{F}$ *est un homéomorphisme sur* $f(\mathrm{U})$*, que* f *est différentiable et que* $\mathrm{D}_a f$ *est un isomorphisme. Alors* f^{-1} *est différentiable au point* $f(a)$ *et* $\mathrm{D}_{f(a)} f^{-1} = (\mathrm{D}_a f)^{-1}$.

Démonstration. — On écrit $g = f^{-1}$ et $b = f(a)$. On cherche à estimer $g(b+h) - g(b)$. La différentiabilité de f nous dit

$$\begin{aligned}
h = (b+h) - b &= f(g(b+h)) - f(a)\\
&= \mathrm{D}_a f(g(b+h) - g(b)) + \|g(b+h) - a\|\varepsilon(g(b+h) - a) \,.
\end{aligned}$$

Du coup, on a

$$\mathrm{D}_a f(g(b+h) - g(b)) = h - \|g(b+h) - a\|\varepsilon(g(b+h) - a) \,.$$

Nous obtenons en appliquant $(\mathrm{D}_a f)^{-1}$

$$g(b+h) - g(b) = (\mathrm{D}_a f)^{-1}(h) - \|g(b+h) - a\|(\mathrm{D}_a f)^{-1}(\varepsilon(g(b+h) - a)) \,. \tag{3.1}$$

Donc, pour conclure, il suffit de montrer que

$$\lim_{h \to 0} \frac{\|g(b+h) - a\|(\mathrm{D}_a f)^{-1}(\varepsilon(g(b+h) - a))}{\|h\|} = 0 \,.$$

A partir de (3.1), on obtient

$$\begin{aligned}
\|g(b+h) - a\| &= \|(\mathrm{D}_a f)^{-1}(h - \|g(b+h) - a\|\varepsilon(g(b+h) - a))\|\\
&\leq \|(\mathrm{D}_a f)^{-1}\| \cdot (\|h\| + \|g(b+h) - a\| \cdot \|\varepsilon(g(b+h) - a)\|) \,.
\end{aligned}$$

Or, si h est assez petit, alors la continuité de g nous permet d'affirmer $\|\varepsilon(g(b+h) - a)\| \leq 1/(2\|(\mathrm{D}_a f)^{-1}\|)$ donc

$$\|g(b+h) - a\| \leq 2\|(\mathrm{D}_a f)^{-1}\| \cdot \|h\| \,.$$

Du coup,

$$\lim_{h \to 0} \frac{\|g(b+h) - a\|\varepsilon(g(b+h) - a)}{\|h\|} = \lim_{h \to 0} \varepsilon(g(b+h) - g(b)) = 0 \,.$$

Ceci montre que (3.1) nous donne la différentielle de f^{-1}. ■

Proposition 3.9. — *Soient $f, g : \mathrm{U} \to \mathbb{R}$ deux applications différentiables.*

1. *La fonction fg est différentiable au point a et $\mathrm{D}_a(fg) = g(a)\mathrm{D}_a f + f(a)\mathrm{D}_a g$.*
2. *Si $g(a) \neq 0$, alors f/g est différentiable au point a et $\mathrm{D}_a(f/g) = (1/g(a)^2)[g(a)\mathrm{D}_a f - f(a)\mathrm{D}_a g]$.*

Démonstration. — On écrit

$$\begin{aligned}(fg)(a+h) &= (f(a) + \mathrm{D}_a f(h) + \|h\|\varepsilon_f(h))(g(a) + \mathrm{D}_a g(h) + \|h\|\varepsilon_g(h)) \\ &= (fg)(a) + f(a)\mathrm{D}_a g(h) + g(a)\mathrm{D}_a f(h) + \|h\|[(f(a) + \mathrm{D}_a f(h))\varepsilon_g(h)) + (g(a) + \mathrm{D}_a g(h))\varepsilon_f(h))] \,.\end{aligned}$$

Pour f/g, on applique la formule du produit et la formule de composition à $x \mapsto (1/x)$ et g. ■

Exemple 3.10. — *Les fonctions polynomiales et les fractions rationnelles sont différentiables là où elles sont définies. En particulier, le déterminant d'une matrice carrée est différentiable.*

3.3.2 Autres propriétés élémentaires

Proposition 3.11. — *Une application $f : \mathrm{U} \to \mathrm{F}$ qui est différentiable au point a est aussi continue au point a.*

Démonstration. — En effet, on a

$$\|f(x) - f(a)\| \leq (\|\mathrm{D}_a f\| + \varepsilon(x-a)) \cdot \|x-a\| \,.$$

■

Proposition 3.12. — *Si $f : \mathrm{U} \to \mathbb{R}$ admet un extremum local au point $a \in \mathrm{U}$ et si f est différentiable au point a alors $\mathrm{D}_a f = 0$.*

Démonstration. — Sans perte de généralité, on peut supposer que a est un maximum local. Du coup, il existe un voisinage $\mathrm{V} \subset \mathrm{U}$ de a tel que $f(x) \leq f(a)$ pour tout $x \in \mathrm{V}$. Si $(a \pm h) \in \mathrm{V}$, $h \neq 0$, alors, comme $f(a+h) = f(a) + \mathrm{D}_a f(h) + \|h\|\varepsilon(h)$, on en déduit $\mathrm{D}_a f(h) + \|h\|\varepsilon(h) \leq 0$, impliquant ainsi

$$\limsup_{h \to 0} \frac{\mathrm{D}_a f(h)}{\|h\|} \leq 0 \,.$$

En considérant $x = a - h$, on obtient de même $\mathrm{D}_a f(h) \geq \|h\|\varepsilon(-h)$, impliquant donc

$$\liminf_{h \to 0} \frac{\mathrm{D}_a f(h)}{\|h\|} \geq 0 \quad \text{puis} \quad \lim_{h \to 0} \frac{\mathrm{D}_a f(h)}{\|h\|} = 0 \,.$$

Par conséquent, $\mathrm{D}_a f = 0$. ■

3.4 Dérivées directionnelles, dérivées partielles et matrice jacobienne

Soit U un ouvert d'un evn E et $f : \mathrm{U} \to \mathrm{F}$ une application à valeurs dans un evn F. Si $v \in \mathrm{E}$, alors la *dérivée au point $a \in \mathrm{U}$ dans la direction v* est définie, lorsqu'elle existe, par la limite

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

et on la note $\partial_v f(a)$, ou $\dfrac{\partial f}{\partial v}(a)$. C'est un vecteur de F.

Si on identifie E à $\mathbb{R}^p$ et que l'on note $\mathcal{B} = (\mathrm{e}_1, \ldots, \mathrm{e}_p)$ la base canonique, alors la dérivée dans la direction e_j s'appelle la *j-ième dérivée partielle de f*. On la note $\partial_j f(a)$, $f'_{x_j}(a)$ ou $\dfrac{\partial f}{\partial x_j}(a)$.

Lemme 3.13. — *Si f est différentiable au point a, alors toutes les dérivées directionnelles existent au point a et on a $\partial_v f(a) = \mathrm{D}_a f(v)$.*

Démonstration. — En effet, on a $f(a+tv)-f(a)=t\mathrm{D}_af(v)+|t|\cdot\|v\|\varepsilon(tv)$ donc

$$\lim_{t\to 0}\frac{f(a+tv)-f(a)}{t}=\lim_{t\to 0}\left(\mathrm{D}_af(v)+\frac{|t|\cdot\|v\|}{t}\varepsilon(tv)\right)=\mathrm{D}_af(v)\,.$$

■

Remarque 3.14. — *L'existence de dérivées partielles, voire de dérivées dans toutes les directions, n'impliquent pas la différentiabilité, ni même la continuité. En effet, prenons par exemple $f:\mathbb{R}^2\to\mathbb{R}$ définie par $f(x,y)=x^3/y$ si $y\neq 0$ et $f(x,0)=0$ sinon. On a $\partial_1 f(0,0)=\partial_2 f(0,0)=0$ car $f(x,0)=f(0,y)\equiv 0$ et plus généralement, pour $a,b\neq 0$,*

$$\partial_{(a,b)}f(0,0)=\lim_{t\to 0}\frac{(ta)^3/(tb)}{t}=0\,.$$

Or $f(x,x^4)=1/x$ qui n'admet pas de limite en 0.

Nous verrons plus loin que l'existence de dérivées partielles continues au voisinage du point a implique la différentiabilité au point a.

Identifions maintenant F à $\mathbb{R}^q$ et considérons les composantes de $f=(f_1,\dots,f_q)$. Si f est différentiable au point $a\in\mathbb{R}^p$, alors la *matrice jacobienne* de f au point a est la matrice de terme général $(\partial_j f_i(a))_{1\leq i\leq q,1\leq j\leq p}\in\mathcal{M}_{q,p}(\mathbb{R})$. Il s'agit donc de la représentation matricielle de D_af dans les bases canoniques. Le *jacobien de f* est le déterminant de la matrice jacobienne.

En particulier, si $f:\mathrm{U}\to\mathbb{R}$ alors

$$\mathrm{D}_af(h_1,\dots,h_p)=\sum_{j=1}^{p}\partial_j f(a)h_j\,.$$

De même, si V est un ouvert de $\mathbb{R}^q$ contenant $f(a)$ et $g:\mathrm{V}\to\mathbb{R}^m$ est une application différentiable au point $f(a)$ alors la formule de composition conduit à, pour $1\leq i\leq m$ et $1\leq j\leq p$,

$$\partial_j(g\circ f)_i(a)=\sum_{k=1}^{q}\partial_k g_i(f(a))\partial_j f_k(a)=\sum_{k=1}^{q}\frac{\partial g_i}{\partial x_k}(f(a))\frac{\partial f_k}{\partial x_j}(a)\,.$$

Plus généralement, si E se décompose en $\mathrm{E}_1\times\ldots\times\mathrm{E}_k$, alors on peut considérer les différentielles partielles par rapport à chaque sous-espace E_j. Plus précisément, on se fixe $a=(a_1,\dots,,a_k)\in\mathrm{U}$ et on considère $\iota_j:\mathrm{E}_j\to\mathrm{E}$ défini par $\iota_j(x)=(a_1,\dots,a_{j-1},x,a_{j+1},\dots,a_k)$ avec les conventions d'usage si $j\in\{1,k\}$. On définit $\partial_j f_a=\mathrm{D}_a(f\circ\iota_j)$ lorsqu'elle est bien définie.

Proposition 3.15. — *Si f est différentiable au point a, alors f admet des dérivées partielles par rapport à chaque sous-espace et*

$$\mathrm{D}_af:(h_1,\dots,h_k)\mapsto\sum_{j=1}^{k}\partial_j f_a(h_j)\,.$$

Démonstration. — Les applications ι_j sont affines, donc $f\circ\iota_j$ est différentiable pour tout $j\in\{1,\dots,k\}$ et, pour $h_j\in\mathrm{E}_j$,

$$\mathrm{D}_a(f\circ\iota_j)(h_j)=\mathrm{D}_af\begin{pmatrix}0\\ \vdots\\ 0\\ h_j\\ 0\\ \vdots\\ 0\end{pmatrix}\quad\text{impliquant ainsi}\quad\mathrm{D}_af\begin{pmatrix}h_1\\ \vdots\\ h_k\end{pmatrix}=\sum_{j=1}^{k}\partial_j f_a(h_j)\,.$$

■

3.5 Gradient

Dans ce paragraphe, on suppose que E est un espace euclidien, U un ouvert de E. On rappelle que, dans un espace euclidien, pour toute forme linéaire $\mathrm{L} \in \mathrm{E}^*$, il existe un vecteur $v \in \mathrm{E}$ tel que $\mathrm{L}(x) = \langle v, x\rangle$ pour tout $x \in \mathrm{E}$. Si $f : \mathrm{U} \to \mathbb{R}$ est une application différentiable en un point $a \in \mathrm{U}$, on définit le *gradient* de f au point a comme le vecteur $grad_a f = \nabla f(a)$ qui vérifie, pour tout $x \in \mathrm{E}$,

$$\mathrm{D}_a f(h) = \langle \nabla f(a), h\rangle\,.$$

Si $\mathrm{E} = \mathbb{R}^p$, alors

$$\nabla f(a) = \begin{pmatrix} \partial_1 f(a) \\ \vdots \\ \partial_p f(a) \end{pmatrix}.$$

On a les interprétations suivantes du gradient.

– Si on considère le graphe $\Gamma = \{(x, f(x)) \in \mathrm{U} \times \mathbb{R},\ x \in \mathrm{U}\}$ alors $\begin{pmatrix} \nabla f(a) \\ (-1) \end{pmatrix}$ est orthogonal à $\mathrm{T}_a\Gamma$ car

$$\left\langle \begin{pmatrix} \nabla f(a) \\ (-1) \end{pmatrix}, \begin{pmatrix} x \\ \mathrm{D}_a f(x) \end{pmatrix} \right\rangle = \langle \nabla f(a), x\rangle - \mathrm{D}_a f(x) = 0\,.$$

– Le gradient $\nabla f(a)$ est orthogonal à l'ensemble de niveau $\mathcal{N} = \{x \in \mathrm{E},\ f(x) = f(a)\}$. En effet, si $x \in \mathcal{N}$ est proche de a, alors $f(x) = f(a) + \mathrm{D}_a f(x-a) + \|x-a\|\varepsilon(x-a)$ donc $\mathrm{D}_a f(x-a) = -\|x-a\|\varepsilon(x-a)$ et

$$\lim_{x\to a} \left\langle \nabla f(a), \frac{x-a}{\|x-a\|} \right\rangle = 0\,.$$

– Le gradient $\nabla f(a)$ pointe vers la plus grande pente de f en a au sens suivant: on a $f(x) = f(a) + \langle \nabla f(a), x\rangle + \|x-a\|\varepsilon(x-a)$. En première approximation, $f(x)$ est maximale si $\langle \nabla f(a), x\rangle$ l'est. Or le produit scalaire est maximal si x est positivement colinéaire à $\nabla f(a)$.

4 Théorème de la moyenne et applications

On rappelle les énoncés pour les fonctions numériques avant de passer aux versions vectorielles.

Théorème 4.1 (Rolle). — *Soient $a, b \in \mathbb{R}$ tels que $a < b$ et $f : [a, b] \to \mathbb{R}$ une application continue, dérivable sur $]a, b[$. Si $f(a) = f(b)$ alors il existe $c \in]a, b[$ tel que $f'(c) = 0$.*

Démonstration. — Comme $[a, b]$ est compact, la fonction f atteint ses bornes. Si elle est constante, n'importe quel point $c \in]a, b[$ convient; sinon, comme $f(a) = f(b)$ un des extrema de f, que l'on notera c, n'est ni a ni b. Donc $c \in]a, b[$ et $f'(c) = 0$ d'après la proposition 3.12. ■

Théorème 4.2 (des accroissements finis). — *Soient $a, b \in \mathbb{R}$ tels que $a < b$ et $f : [a, b] \to \mathbb{R}$ une application continue, dérivable sur $]a, b[$. Il existe $c \in]a, b[$ tel que $f(b) - f(a) = f'(c)(b-a)$.*

Démonstration. — On applique le théorème de Rolle à la fonction auxiliaire $g : [a, b] \to \mathbb{R}$ définie par $g(t) = f(t)(b-a) - (f(b) - f(a))t$. ■

Théorème 4.3 (de la moyenne). — *Soient $a, b \in \mathbb{R}$ tels que $a < b$, F un evn, $f : [a, b] \to \mathrm{F}$ et $g : [a, b] \to \mathbb{R}$ deux applications continues, dérivables sur $]a, b[$. Si, pour tout $t \in]a, b[$, on a $\|f'(t)\| \leq g'(t)$ alors*

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

Démonstration. — Il suffit de montrer que, pour tout $\varepsilon > 0$, on a $\|f(b) - f(a)\| \leq g(b) - g(a) + \varepsilon(b-a)$. Une fois cette assertion établie, comme ε est arbitraire, on obtiendra $\|f(b) - f(a)\| \leq g(b) - g(a)$ et donc le théorème.

A cet effet, on considère $A = \{t \in [a,b] \ \|f(t) - f(a)\| \leq g(t) - g(a) + \varepsilon(t-a)\}$ pour $\varepsilon > 0$ fixé et on remarque que $A \neq \varnothing$ puisque $a \in A$ (car $0 \leq 0$). Notons $s = \sup A$. Par définition, il existe une suite croissante (s_n) contenue dans A qui tend vers s. Pour tout n, on a donc

$$\|f(s_n) - f(a)\| \leq g(s_n) - g(a) + \varepsilon(s_n - a)\,.$$

Comme f et g sont continues, on obtient à la limite $\|f(s) - f(a)\| \leq g(s) - g(a) + \varepsilon(s-a)$ montrant ainsi $s \in A$.

Pour conclure, on doit montrer $s = b$. Supposons donc $s < b$. Par définition de la dérivée au point s, on peut trouver $\delta > 0$ telle que $s + \delta < b$ et, pour $t \in]s, s+\delta]$, on a

$$\|f(t) - f(s)\| \leq \|D_s f\| \cdot (t-s) + (t-s)\frac{\varepsilon}{2}$$

et

$$g'(s)(t-s) \leq g(t) - g(s) + \frac{\varepsilon}{2}(t-s)$$

de sorte que

$$\begin{aligned} \|f(t) - f(s)\| &\leq \|D_s f\| \cdot (t-s) + (t-s)\frac{\varepsilon}{2} \\ &\leq g'(s) \cdot (t-s) + (t-s)\frac{\varepsilon}{2} \\ &\leq g(t) - g(s) + \varepsilon(t-s)\,. \end{aligned}$$

Or comme $s \in A$, on a, pour $t \in]s, s+\delta[$,

$$\begin{aligned} \|f(t) - f(a)\| &\leq \|f(t) - f(s)\| + \|f(s) - f(a)\| \\ &\leq g(t) - g(s) + \varepsilon(t-s) + g(s) - g(a) + \varepsilon(s-a) \\ &\leq g(t) - g(a) + \varepsilon(t-a) \end{aligned}$$

ce qui montre $t \in A$ et contredit le fait que $s = \sup A$. Du coup, $s = b$. ∎

Soit E un evn. Si $a, b \in E$, le *segment* $[a,b]$ est $\{(1-t)a + bt,\ t \in [0,1]\}$. Un sous-ensemble X de E est *convexe* si, pour tous $x, y \in X$, le segment $[x,y]$ est contenu dans X. Par exemple une boule, ouverte ou fermée, est convexe.

Théorème 4.4 (inégalité des accroissements finis). — *Soient $k \geq 0$, E, F deux evn, U un ouvert convexe et $f : U \to F$ une application différentiable. Si, pour tout $a \in U$, on a $\|D_a f\| \leq k$, alors, pour tous $x, y \in U$, on a*

$$\|f(x) - f(y)\| \leq k\|x - y\|\,.$$

Démonstration. — On applique le théorème de la moyenne à l'application auxiliaire $g : [0,1] \to F$ définie par $g(t) = f((1-t)x + ty)$: on a $g'(t) = D_{(1-t)x+ty} f(y-x)$ donc $\|g'(t)\| \leq k\|y - x\|$ et

$$\|g(1) - g(0)\| = \|f(y) - f(x)\| \leq k\|y - x\|\,.$$

∎

4.1 Applications de dérivée nulle

Un ouvert U est *connexe par arcs* si pour tous $x, y \in U$, il existe une fonction continue $c : [0,1] \to U$, autrement dit un *chemin*, telle que $c(0) = x$ et $c(1) = y$. Un ouvert convexe est en particulier connexe par arcs.

Proposition 4.5. — *Soient $U \subset E$ un ouvert connexe par arcs et $f : U \to F$ une application différentiable. Si, pour tout $a \in U$, on a $D_a f = 0$ alors f est constante.*

La fonction qui prend la valeur 0 sur $]0,1[$ et 1 sur $]2,3[$ est différentiable, de dérivée identiquement nulle, mais non constante. Ce n'est pas une contradiction puisque $]0,1[\cup]2,3[$ n'est pas connexe par arcs.

Démonstration. — Soient $x, y \in \mathrm{U}$ et $c : [0,1] \to \mathrm{U}$ un chemin reliant x à y. Notons $\mathrm{A} = \{t \in [0,1],\ (f \circ c)(t) = f(x)\}$. Cet ensemble est non vide car il contient $\{0\}$. Notons $s = \sup \mathrm{A}$ et montrons que $s = 1$ par l'absurde. Si $s < 1$, notons $a = c(s)$ et $r > 0$ de sorte que $\mathrm{B}(a,r) \subset \mathrm{U}$. La continuité de c nous donne l'existence de $\delta > 0$, $\delta \le \min\{s, 1-s\}$ telle que $c(]s-\delta, s+\delta[) \subset \mathrm{B}(a,r)$.

L'inégalité des accroissements finis affirme que, pour $b \in \mathrm{B}(a,r)$, on a $\|f(b)-f(a)\| \le \sup \|\mathrm{D}_x f\| \|b-a\| = 0$. Par définition de s, on peut choisir $t \in \mathrm{A} \cap]s-\delta, s]$ de sorte que $f(a) = f(c(s)) = f(x)$ car $t \in \mathrm{A}$; par ailleurs, si $t = s + \delta/2$, alors $f(c(t+\delta/2)) = f(a) = f(x)$, ce qui contredirait la définition de s. Donc $s = 1$ et $f(x) = f(y)$. ■

4.2 Caractérisation des fonctions de classe $\mathcal{C}^1$

On fournit une sorte de réciproque au lemme 3.13 qui donne un moyen assez simple d'établir en pratique la différentiabilité de fonctions.

Théorème 4.6. — *Une application $f : \mathrm{U}(\subset \mathbb{R}^n) \to \mathrm{F}$ est de classe $\mathcal{C}^1$ si et seulement si toutes les dérivées partielles de f existent et sont continues. Dans ce cas, on a*

$$\mathrm{D}_a f(h_1, \ldots, h_n) = \sum_{j=1}^{n} h_j \partial_j f(a)\,.$$

Dans cette écriture, on rappelle que $\partial_j f(a)$ désigne chaque fois des vecteurs de F.

Démonstration. — Le lemme 3.13 implique que si f est de classe $\mathcal{C}^1$, alors les dérivées partielles de f existent et sont continues, comme composées des fonctions $\mathrm{D}f$ et des évaluations sur les vecteurs de base $\mathrm{L} \in \mathcal{L}(\mathbb{R}^n, \mathrm{F}) \mapsto \mathrm{L}(\mathrm{e}_j)$. Passons maintenant à la réciproque.

On munit $\mathbb{R}^n$ de la norme $\|\cdot\|_\infty$. Montrons la différentiabilité de f au point $a \in \mathrm{U}$. Notons que l'on a déjà un candidat pour la différentielle. On suppose $\mathrm{B}(a,\mathrm{R}) \subset \mathrm{U}$. On considère $h_1, \ldots, h_n \in \mathbb{R}$ avec $|h_j| \le \mathrm{R}$ pour chaque $j \in \{1, \ldots, n\}$ et on notera $h = (h_1, \ldots, h_n)$.

Afin d'exploiter les hypothèses, on introduit les points suivants

$$\begin{cases} x_j = (a_1 + h_1, \ldots, a_{j-1} + h_{j-1}, a_j, \ldots, a_n) & \text{pour } j \in \{1, \ldots, n\} \\ x_{n+1} = a + h & \\ u_j = (0, \ldots, 0, h_j, 0, \ldots, 0) = h_j \mathrm{e}_j & \text{pour } j \in \{1, \ldots, n\}\,. \end{cases}$$

Il vient

$$f(a+h) - f(a) = \sum_{j=1}^{n} f(x_{j+1}) - f(x_j) = \sum_{j=1}^{n} f(x_j + u_j) - f(x_j)$$

On obtient ainsi

$$\begin{aligned} \left\| f(a+h) - f(a) - \sum_{j=1}^{n} h_j \partial_j f(a) \right\| &= \left\| \sum_{j=1}^{n} [f(x_{j+1}) - f(x_j) - h_j \partial_j f(x_j)] + \sum_{j=1}^{n} (\partial_j f(x_j) - \partial_j f(a)) h_j \right\| \\ &\le \sum_{j=1}^{n} \|f(x_{j+1}) - f(x_j) - h_j \partial_j f(x_j)\| + \|h\|_\infty \sum_{j=1}^{n} \|\partial_j f(x_j) - \partial_j f(a)\|\,. \end{aligned}$$

On introduit maintenant les fonctions auxiliaires

$$g_j : t \in [0,1] \mapsto f(x_j + t u_j) - t h_j \partial_j f(x_j)$$

auxquelles on souhaite appliquer le théorème de la moyenne. Vérifions donc que g_j est différentiable. Le second terme est différentiable car linéaire en t. Pour le premier, posons $y_j = x_j + t u_j$ et prenons $\alpha \in \mathbb{R}$. En se souvenant que $u_j = h_j \mathrm{e}_j$, on obtient

$$\lim_{\alpha \to 0} \frac{f(y_j + \alpha u_j) - f(y_j)}{\alpha} = \lim_{\alpha \to 0} \frac{f(y_j + (\alpha h_j)\mathrm{e}_j) - f(y_j)}{\alpha h_j} h_j = \partial_j f(y_j) h_j\,.$$

Du coup, g_j est bien dérivable et

$$g_j'(t) = \partial_j f(x_j + tu_j)h_j - h_j \partial_j f(x_j)\,.$$

Le théorème de la moyenne donne $\|g_j(1) - g_j(0)\| \leq \sup_{t\in[0,1]} \|g_j'(t)\|$ soit

$$\|f(x_{j+1}) - f(x_j) - h_j\partial_j f(x_j)\| \leq \sup_{t\in[0,1]} \|\partial_j f(x_j + tu_j) - \partial_j f(x_j)\| \cdot |h_j|\,.$$

On se fixe $\varepsilon > 0$. Par continuité des dérivées partielles, on peut trouver $r \in]0, \mathrm{R}[$ de sorte que, pour tous $x, y \in \mathrm{B}(a, r)$, on ait $\|\partial_j f(x) - \partial_j f(y)\| \leq \varepsilon/(2n)$. Du coup, il vient

$$\left\| f(a+h) - f(a) - \sum_{j=1}^n h_j \partial_j f(a) \right\| \leq \sum_{j=1}^n (\varepsilon/2n)|h_j| + \|h\|_\infty \sum_{j=1}^n \varepsilon/(2n) \leq \|h\|_\infty \varepsilon\,.$$

Ceci montre la différentiabilité au point a et identifie sa différentielle. Il reste à vérifier la continuité de $\mathrm{D}f$. Prenons $a, b \in \mathrm{U}$; pour tout $h \in \mathbb{R}^n$ tel que $\|h\|_\infty$, on a

$$\|(\mathrm{D}_a f - \mathrm{D}_b f)(h)\|_\mathrm{F} \leq \sum_{j=1}^n |h_j| \cdot \|\partial_j f(a) - \partial_j f(b)\|_\mathrm{F} \leq \sum_{j=1}^n \|\partial_j f(a) - \partial_j f(b)\|_\mathrm{F}$$

donc

$$\|\mathrm{D}_a f - \mathrm{D}_b f\| \leq \sum_{j=1}^n \|\partial_j f(a) - \partial_j f(b)\|_\mathrm{F}$$

et la continuité de $\mathrm{D}f$ découle de celle des dérivées partielles. ■

Le même argument conduit au résultat plus général suivant qui améliore la proposition 3.15.

Théorème 4.7. — *Soient* $\mathrm{E}_1, \ldots, \mathrm{E}_k, \mathrm{F}$ *des espaces de dimension finie,* $f : \mathrm{U} \to \mathrm{F}$ *une application définie sur un ouvert de* $\mathrm{E} \overset{def.}{=} \mathrm{E}_1 \times \ldots \times \mathrm{E}_k$. *L'application* f *est de classe* $\mathcal{C}^1$ *sur* U *si et seulement si* f *admet des différentielles partielles continues* $\partial_j f$, $j = 1, \ldots, k$, *le long de chaque sous-espace* E_j. *Dans ce cas,*

$$\mathrm{D}_a f(h) = \sum \partial_j f_a(h_j)\,.$$

4.3 Limite de suites de fonctions différentiables

On s'intéresse maintenant aux limites de fonctions différentiables.

Théorème 4.8 (d'inversion des limites et des dérivées). — *Soient* $\mathrm{U} \subset \mathrm{E}$ *un ouvert connexe par arcs et* $(f_n)_{n\geq 0}$ *une suite d'applications différentiables de* U *dans* F *telles que*

1. *il existe* $x_0 \in \mathrm{U}$ *tel que la suite* $(f_n(x_0))_{n\geq 0}$ *est convergente.*
2. *pour tout* $a \in \mathrm{U}$, *il existe* $r > 0$ *tel que la suite* $(\mathrm{D}f_n)_n$ *est uniformément convergente sur* $\mathrm{B}(a, r)$.

Alors la suite $(f_n)_n$ *est uniformément convergente sur chaque boule* $\mathrm{B}(a, r)$ *vers une fonction* f, *différentiable sur* U, *de différentielle la limite des différentielles* $(\mathrm{D}f_n)_n$*:*

$$\lim_n \mathrm{D}f_n = \mathrm{D}(\lim_n f_n)\,.$$

Remarque 4.9. — *Les deux conditions sont nécessaires. Prenons par exemple la suite de fonctions constantes à valeurs réelles* n*: les différentielles sont indentiquement nulles donc uniformément convergentes, mais la suite est divergente; ceci montre la nécessité de la première condition. Prenons sur* $]-1, 1[$, *la suite* $(f_n)_n$ *définies par* $x \mapsto \sqrt{x^2 + 1/(n+1)}$. *On a convergence uniforme vers* $|\cdot|$*; on peut vérifier que la suite de dérivées tend simplement vers* 1 *pour* $x > 0$, (-1) *pour* $x < 0$ *et* 0 *pour* $x = 0$.

Démonstration. — Prenons tout d'abord $a \in \mathrm{U}$ et $r > 0$ de sorte que **2.** soit vérifié. L'inégalité des accroissements finis appliquée à $f_p - f_q$ implique, pour $x, y \in \mathrm{B}(a,r)$,

$$\|(f_p(x) - f_q(x)) - (f_p(y) - f_q(y))\| \leq \sup_{z \in \mathrm{B}(a,r)} \|\mathrm{D}_z f_p - \mathrm{D}_z f_q\| \cdot \|x - y\| . \tag{4.1}$$

Comme on a convergence uniforme des dérivées, la convergence en un seul point de $\mathrm{B}(a,r)$ implique la convergence uniforme de (f_n) sur la boule. Justifions cette assertion: supposons que $(f_n(x))_n$ soit convergente. Alors $(f_n(x))_n$ est une suite de Cauchy, tout comme $(\mathrm{D}f_n)_n$ est une suite de Cauchy uniforme sur $\mathrm{B}(a,r)$. Prenons $\varepsilon > 0$, et fixons n_0 assez grand de sorte que si, $p, q \geq n_0$, alors

$$\|f_p(x) - f_q(x)\| \leq \varepsilon/2 \quad \text{et pour tout } z \in \mathrm{B}(a,r),\ \|\mathrm{D}_z f_p - \mathrm{D}_z f_q\| \leq \varepsilon/(2r) .$$

On en déduit, avec (4.1), pour $p, q \geq n_0$ et tout $y \in \mathrm{B}(a,r)$,

$$\|f_p(y) - f_q(y)\| \leq \|f_p(x) - f_q(x)\| + \|(f_p(x) - f_q(x)) - (f_p(y) - f_q(y))\| \leq \varepsilon/2 + (\varepsilon/2r)\|x - y\| \leq \varepsilon$$

puisque $\|x - y\| \leq \|x - a\| + \|a - y\| < 2r$. Donc la suite $(f_n)_n$ est de Cauchy uniforme sur $\mathrm{B}(a,r)$, impliquant la convergence uniforme de $(f_n)_n$ sur $\mathrm{B}(a,r)$.

Soit $x \in \mathrm{U}$ et prenons un chemin $c : [0,1] \to \mathrm{U}$ qui relie x_0 à x. Notons A l'ensemble des $t \in [0,1]$ pour lesquels on a convergence de la suite $(f_n(c(t)))_n$ et $s = \sup \mathrm{A}$. L'analyse précédente montre que $s = 1$. En effet, cet ensemble est non vide car il contient 0. Si $s < 1$, notons $a = c(s)$ et $r > 0$ de sorte que l'on ait convergence uniforme des dérivées sur $\mathrm{B}(a,r)$. La continuité de c nous donne l'existence de $\delta > 0$, $\delta \leq \min\{s, 1-s\}$, telle que $c(]s-\delta, s+\delta[) \subset \mathrm{B}(a,r)$. Par définition de s, il existe $t \in \mathrm{A} \cap]s-\delta, s]$. Donc on a convergence uniforme de (f_n) sur $\mathrm{B}(a,r)$, en particulier au point $t = s + \delta/2$, contredisant la définition de s. Comme $s = 1$, on trouve t assez proche de 1 pour que (4.1) implique la convergence au point x et sur la boule $\mathrm{B}(x, r_x)$ correspondante.

Notons f la limite de $(f_n)_n$ et L celle de $(\mathrm{D}_a f_n)_n$. Fixons-nous $\mathrm{B}(a,r)$ où ces convergences sont uniformes. Nous allons les exploiter pour établir la différentiabilité de f:

$$\begin{aligned}
\|f(a+h) - f(a) - \mathrm{L}(h)\| &= \|(f(a+h) - f_n(a+h)) - (f(a) - f_n(a)) \\
&\quad + (f_n(a+h) - f_n(a) - \mathrm{D}_a f_n(h)) + (\mathrm{L} - \mathrm{D}_a f_n)(h)\| \\
&\leq \|(f(a+h) - f_n(a+h)) - (f(a) - f_n(a))\| \\
&\quad + \|f_n(a+h) - f_n(a) - \mathrm{D}_a f_n(h)\| + \|\mathrm{L} - \mathrm{D}_a f_n\| \cdot \|h\| .
\end{aligned}$$

Fixons-nous $\varepsilon > 0$. Pour p, q assez grands, on a $\|\mathrm{D}_x f_p - \mathrm{D}_x f_q\| \leq \varepsilon/3$ pour tout $x \in \mathrm{B}(a,r)$. D'après (4.1), on obtient en passant à la limite $p \to \infty$ et en prenant $q = n$ assez grand,

$$\left\{\begin{array}{l}
\|(f(a+h) - f_n(a+h)) - (f(a) - f_n(a))\| \leq \varepsilon\|h\|/3 , \\[2ex]
\|\mathrm{L} - \mathrm{D}_a f_n\| \cdot \|h\| \leq \varepsilon\|h\|/3 .
\end{array}\right.$$

On fixe maintenant n assez grand, et on utilise la différentiabilité de f_n pour choisir h assez petit de sorte que

$$\|f_n(a+h) - f_n(a) - \mathrm{D}_a f_n(h)\| \leq \varepsilon\|h\|/3 .$$

En rassemblant ces trois estimées, on conclut

$$\|f(a+h) - f(a) - \mathrm{L}(h)\| \leq \varepsilon\|h\| .$$

■

En prenant des sommes partielles, on obtient

Théorème 4.10 (d'inversion des séries et des dérivées). — *Soient* $\mathrm{U} \subset \mathrm{E}$ *un ouvert connexe par arcs et* $(f_n)_{n \geq 0}$ *une suite d'applications différentiables de* U *dans* F *telles que*

1. *il existe* $x_0 \in \mathrm{U}$ *tel que la série de terme général* $(f_n(x_0))_{n \geq 0}$ *est convergente.*

2. *pour tout* $a \in \mathrm{U}$*, il existe* $r > 0$ *tel que la série de terme général* $(\mathrm{D}f_n)_n$ *est uniformément convergente sur* $\mathrm{B}(a, r)$*.*

Alors la série de terme général $(f_n)_n$ *est uniformément convergente sur chaque boule* $\mathrm{B}(a, r)$ *vers une fonction* f*, différentiable sur* U*, de différentielle la somme des différentielles* $(\mathrm{D}f_n)_n$*:*

$$\sum_n \mathrm{D}f_n = \mathrm{D}\left(\sum_n f_n\right).$$

On en déduit l'application suivante:

Proposition 4.11. — *Soit* $\sum_{n\geq 0} a_n z^n$ *une série entière de rayon de convergence* $\mathrm{R} > 0$*. La fonction* $f : \mathcal{L}(\mathbb{R}^n) \to \mathcal{L}(\mathbb{R}^n)$ *définie par* $f(\mathrm{A}) = \sum_{n\geq 0} a_n \mathrm{A}^n$ *est bien définie et différentiable, de classe* $\mathcal{C}^1$ *sur* $\mathrm{B}_{\mathcal{L}(\mathbb{R}^n)}(0, \mathrm{R})$*. De plus, on a*

$$\mathrm{D_A}f(\mathrm{H}) = \sum_{n\geq 0} a_n(\mathrm{HA}^{n-1} + \mathrm{AHA}^{n-2} + \ldots + \mathrm{A}^{n-1}\mathrm{H}).$$

Démonstration. — Si $\|\mathrm{A}\| < \mathrm{R}$ alors $\sum_{n\geq 0} \|a_n \mathrm{A}^n\| \leq \sum |a_n| \|\mathrm{A}\|^n$ qui est finie puisque $\|\mathrm{A}\| < \mathrm{R}$. Notons que $\mathrm{A} \mapsto \mathrm{A}^n$ est différentiable de dérivée $\mathrm{H} \mapsto \mathrm{HA}^{n-1} + \mathrm{AHA}^{n-2} + \ldots + \mathrm{A}^{n-1}\mathrm{H}$. Par conséquent, on obtient

$$\sum \|a_n(\mathrm{HA}^{n-1} + \mathrm{AHA}^{n-2} + \ldots + \mathrm{A}^{n-1}\mathrm{H})\| \leq \sum n|a_n| \|\mathrm{A}\|^{n-1}$$

qui est donc uniformément convergente sur $\mathrm{B}(0, \mathrm{R}')$ pour tout $\mathrm{R}' < \mathrm{R}$. Du coup le théorème précédent s'applique.

Enfin, l'expression de $\mathrm{D_A}f$ dépend continûment de A. ■

Cette proposition s'applique par exemple à

$$\begin{cases} (\mathrm{I} + \mathrm{A})^{-1} & = \displaystyle\sum_{n\geq 0} (-1)^n \mathrm{A}^n, & \|\mathrm{A}\| < 1 \\ \exp \mathrm{A} & = \displaystyle\sum_{n\geq 0} \frac{\mathrm{A}^n}{n!}, & \mathrm{A} \text{ quelconque}, \\ \mathrm{Log}(\mathrm{I} + \mathrm{A}) & = \displaystyle\sum_{n\geq 1} \frac{(-1)^{n+1}\mathrm{A}^n}{n}, & \|\mathrm{A}\| < 1. \end{cases}$$

Chapitre III. — DIFFERENTIELLES D'ORDRE SUPERIEUR

Comme d'habitude, tous nos espaces vectoriels sont de dimension finie. On note $(\mathrm{E}, ||.||_{\mathrm{E}})$ et $(\mathrm{F}, ||.||_{\mathrm{F}})$ des espaces vectoriels normés sur $\mathbb{R}$, U un ouvert non vide de E et $f\ :\mathrm{U}\to\mathrm{F}$ une application. Par un choix de bases pour E et F, on se ramène à $f:\mathbb{R}^p\to\mathbb{R}^q$.

5 Différentielles d'ordre 2

Définition 5.1. — *On dit que $f:\mathrm{U}\to\mathrm{F}$ est deux fois différentiable au point $a\in\mathrm{U}$ si f est différentiable sur un voisinage V de a et si l'application $\mathrm{D}f:\mathrm{V}\to\mathcal{L}(\mathrm{E},\mathrm{F})$ définie par $x\mapsto\mathrm{D}_xf$ est différentiable au point a. On la note $\mathrm{D}^2_af\in\mathcal{L}(\mathrm{E},\mathcal{L}(\mathrm{E},\mathrm{F}))$. L'application est deux fois différentiable sur U si f est deux fois différentiable en tout point de U. Elle est de classe $\mathcal{C}^2$ si f est deux fois différentiable et si la différentielle seconde $a\mapsto\mathrm{D}^2_af$ est continue.*

Si f est deux fois différentiable, alors $\mathrm{D}^2_af:\mathrm{E}\to\mathcal{L}(\mathrm{E},\mathrm{F})$ associe linéairement à un vecteur $k\in\mathrm{E}$ une application linéaire $\mathrm{L}_k:\mathrm{E}\to\mathrm{F}$. On a ainsi

$$\mathrm{D}_{a+k}f=\mathrm{D}_af+\mathrm{L}_k+\|k\|\varepsilon(k)$$

pour $a+k\in\mathrm{U}$ et où $\lim_{k\to0}\varepsilon(k)=0$. Autrement dit, on écrit $\mathrm{D}^2_af(k)(h)=\mathrm{L}_k(h)$. Le lemme 5.2 ci-dessous nous permet d'identifier $\mathcal{L}(\mathrm{E},\mathcal{L}(\mathrm{E},\mathrm{F}))$ avec l'espace vectoriel $\mathcal{L}^2(\mathrm{E},\mathrm{F})$ des applications bilinéaires $b:\mathrm{E}\times\mathrm{E}\to\mathrm{F}$, de sorte que l'on écrira $\mathrm{D}^2_af(k,h)=\mathrm{D}^2_af(k)(h)$.

Lemme 5.2. — *On a un isomorphisme canonique $\Phi:\mathcal{L}(\mathrm{E},\mathcal{L}(\mathrm{E},\mathrm{F}))\to\mathcal{L}^2(\mathrm{E},\mathrm{F})$ préservant les normes défini par $\Phi(f)(x,y)=f_x(y)$ où $(f:x\mapsto f_x)\in\mathcal{L}(\mathrm{E},\mathcal{L}(\mathrm{E},\mathrm{F}))$.*

Démonstration. — Tout d'abord on vérifie que si $f\in\mathcal{L}(\mathrm{E},\mathcal{L}(\mathrm{E},\mathrm{F}))$, alors $\Phi(f)$ est bien bilinéaire: prenons $x,x',y,y'\in\mathrm{E}$ et $\lambda,\mu\in\mathbb{R}$. On écrit

$$\begin{aligned}\Phi(f)(x+\lambda x',y+\mu y') &= f_{x+\lambda x'}(y+\mu y')\\ &= f_x(y+\mu y')+\lambda f_{x'}(y+\mu y')\quad\text{par linéarité de } x\mapsto f_x\\ &= f_x(y)+\mu f_x(y')+\lambda f_{x'}(y)+\lambda\mu f_{x'}(y')\quad\text{par linéarité de } f_x \text{ et } f_{x'}\\ &= \Phi(f)(x,y)+\lambda\Phi(f)(x',y)+\mu\Phi(x,y')+\lambda\mu\Phi(f)(x',y')\,.\end{aligned}$$

En prenant $y'=0$, on montre la linéarité par rapport à la première variable et en prenant $x'=0$, la linéarité par rapport à la seconde.

Enfin, on montre facilement que $\Phi(f+\lambda g)=\Phi(f)+\lambda\Phi(g)$. Si $\Phi(f)$ est l'application bilinéaire nulle, alors, pour chaque $x\in\mathrm{E}$, l'application $x\mapsto f_x$ est nulle, donc f est nulle. Ceci montre l'injectivité de Φ. Pour montrer la surjectivité, on considère une application bilinéaire $b:\mathrm{E}\times\mathrm{E}\to\mathrm{F}$ et on pose $f:x\mapsto(f_x:y\mapsto f(x,y))$ et on vérifie que $\Phi(f)=b$.

Enfin, si $f\in\mathcal{L}(\mathrm{E},\mathcal{L}(\mathrm{E},\mathrm{F}))$, alors

$$\|f\|=\sup_{\|x\|=1}\|f_x\|=\sup_{\|x\|=1}\left(\sup_{\|y\|=1}\|f_x(y)\|\right)=\sup_{\|x\|=\|y\|=1}\|\Phi(f)(x,y)\|\,.$$

■

Lemme 5.3. — *Supposons f deux fois différentiable au point $a\in\mathrm{U}$ et fixons-nous $h\in\mathrm{E}$. Alors l'application $g:x\mapsto\mathrm{D}_xf(h)$ est différentiable au point a et $\mathrm{D}_ag(k)=\mathrm{D}^2_af(k,h)$. Autrement dit, $\mathrm{D}_a(\mathrm{D}f(h))(k)=\mathrm{D}^2_af(k,h)$.*

Démonstration. — En effet, l'application d'évaluation $\mathrm{ev}:\mathrm{L}\in\mathcal{L}(\mathrm{E},\mathrm{F})\mapsto\mathrm{L}(h)\in\mathrm{F}$ est linéaire (donc différentiable) et $g=\mathrm{ev}\circ\mathrm{D}f$ puisque $g(x)=\mathrm{ev}\circ\mathrm{D}_xf=\mathrm{D}_xf(h)$. Donc

$$\mathrm{D}_ag(k)=(\mathrm{ev}\circ\mathrm{D}_a(\mathrm{D}f))(k)=\mathrm{D}^2_af(k)(h)=\mathrm{D}^2_af(k,h)\,.$$

■

Remarque 5.4. — *En prenant* $\mathrm{E} = \mathbb{R}^p$, *et en écrivant*

$$\partial^2_{i,j} f(a) = \frac{\partial^2 f}{\partial x_i \partial x_j}(a) \overset{def.}{=} \mathrm{D}^2_a f(\mathrm{e}_i, \mathrm{e}_j) = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)(a)$$

on obtient, à partir de $\mathrm{D}_a f(h) = \sum \partial_j f(a) h_j$,

$$\mathrm{D}^2_a f(k,h) = \sum_{i,j=1}^{p} \partial^2_{i,j} f(a) k_i h_j \,.$$

5.1 Théorème de Schwarz

Théorème 5.5 (Schwarz). — *Si* f *est deux fois différentiable au point* a, *alors* $\mathrm{D}^2_a f$ *est une application bilinéaire symétrique: pour tous* $h, k \in \mathrm{E}$, *on a* $\mathrm{D}^2_a f(h,k) = \mathrm{D}^2_a f(k,h)$.

Démonstration. — il suffit de montrer que $\|\mathrm{D}^2_a f(h,k) - \mathrm{D}^2_a f(k,h)\| = o((\|h\| + \|k\|)^2)$. En effet, si $u, v \in \mathrm{E}$ sont quelconques (mais non nuls) alors

$$\|\mathrm{D}^2_a f(u,v) - \mathrm{D}^2_a f(v,u)\| = \frac{1}{t^2}\|\mathrm{D}^2_a f(tu,tv) - \mathrm{D}^2_a f(tv,tu)\| = \frac{1}{t^2} o(t^2) = o(1)$$

donc $\mathrm{D}^2_a f(u,v) = \mathrm{D}^2_a f(v,u)$.

On se fixe $k, h \in \mathrm{E}$ assez petits de sorte que la fonction auxiliaire $g : [0,1] \to \mathrm{F}$ définie par

$$g(t) = f(a + th + k) - f(a + th) - t\mathrm{D}^2_a f(k,h)$$

soit effectivement bien définie. Nous souhaitons appliquer le théorème de la moyenne à g. Cette fonction est différentiable comme fonction composée d'applications différentiables et

$$g'(t) = \mathrm{D}_{a+th+k} f(h) - \mathrm{D}_{a+th} f(h) - \mathrm{D}^2_a f(k,h) = (\mathrm{D}_{a+th+k} f - \mathrm{D}_a f)(h) - (\mathrm{D}_{a+th} f - \mathrm{D}_a f)(h) - \mathrm{D}^2_a f(k,h)\,.$$

En remarquant que $\mathrm{D}^2_a f(k,h) = \mathrm{D}^2_a f(k + th, h) - \mathrm{D}^2_a f(th, h)$, on obtient

$$g'(t) = (\mathrm{D}_{a+th+k} f - \mathrm{D}_a f - \mathrm{D}^2_a f(k + th))(h) - (\mathrm{D}_{a+th} f - \mathrm{D}_a f - \mathrm{D}^2_a f(th))(h)\,.$$

Soit $\varepsilon > 0$. La différentiabilité de $\mathrm{D}f$ au point a implique l'existence de $r > 0$ tel que si $u \in \mathrm{B}_{\mathrm{E}}(0,r)$, alors

$$\|\mathrm{D}_{a+u} f - \mathrm{D}_a f - \mathrm{D}^2_a f(u)\| \leq \varepsilon \|u\|\,.$$

On se fixe $h, k \in \mathrm{B}_{\mathrm{E}}(0, r/2)$ de sorte que, pour tout $t \in [0,1]$, on ait $\|k + th\| \leq \|k\| + |t|\|h\| < r$. De même, on a $\|th\| = |t|\|h\| < r$. Donc, pour tout $t \in [0,1]$, on obtient

$$\begin{aligned}
\|g'(t)\| &\leq \|(\mathrm{D}_{a+th+k} f - \mathrm{D}_a f - \mathrm{D}^2_a f(k+th))(h)\| + \|(\mathrm{D}_{a+th} f - \mathrm{D}_a f - \mathrm{D}^2_a f(th))(h)\| \\
&\leq \|\mathrm{D}_{a+th+k} f - \mathrm{D}_a f - \mathrm{D}^2_a f(k+th)\|\|h\| + \|\mathrm{D}_{a+th} f - \mathrm{D}_a f - \mathrm{D}^2_a f(th)\|\|h\| \\
&\leq \varepsilon(\|th + k\|\|h\| + \|th\|\|h\|) \leq \varepsilon(\|h\| + \|k\| + \|h\|)\|h\| \\
&\leq 2\varepsilon(\|h\| + \|k\|)^2\,.
\end{aligned}$$

On applique maintenant le théorème de la moyenne à g afin d'obtenir

$$\|g(1) - g(0)\| \quad \leq \quad \sup_{0 \leq t \leq 1} \|g'(t))\| \leq 2\varepsilon(\|h\| + \|k\|)^2\,.$$

On obtient donc

$$\|f(a + h + k) - f(a + h) - f(a + k) + f(a) - \mathrm{D}^2_a f(k,h)\| \leq 2\varepsilon(\|h\| + \|k\|)^2\,.$$

Par symétrie du rôle de h et k, on en déduit

$$\|\mathrm{D}^2_a f(h,k) - \mathrm{D}^2_a f(k,h)\| \leq 4\varepsilon(\|h\| + \|k\|)^2\,.$$

■

Corollaire 5.6. — *Soit* $\mathrm{U} \subset \mathbb{R}^p$ *un ouvert. Si* $f : \mathrm{U} \to \mathrm{F}$ *est deux fois différentiable au point* $a \in \mathrm{U}$*, alors, pour tous* $i, j \in \{1, \dots, p\}$*, on a*

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

Démonstration. — D'après le théorème de Schwarz, on a

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \mathrm{D}^2_a f(\mathrm{e}_i, \mathrm{e}_j) = \mathrm{D}^2_a f(\mathrm{e}_j, \mathrm{e}_i) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a)\,.$$

■

Corollaire 5.7. — *Sous les mêmes conditions, si toutes les dérivées secondes* $(\partial^2_{i,j} f)$*,* $i, j \in \{1, \dots, p\}$*, existent et sont continues au voisinage au point* a *alors*

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

Démonstration. — Il découle des hypothèses que les dérivées partielles existent et sont continues, donc le théorème 4.6 implique que $f$est de classe $\mathcal{C}^1$. De plus, les dérivées partielles de $\mathrm{D}f$ correspondent aux dérivées partielles secondes de f, donc le même théorème implique que f est de classe $\mathcal{C}^2$ au voisinage de a. Du coup, le corollaire 5.6 permet de conclure. ■

Exemple 5.8. — *Si* f *n'est pas deux fois différentiable, alors l'ordre des dérivées partielles est important. Prenons par exemple* $f : \mathbb{R}^2 \to \mathbb{R}$ *définie par* $f(0,0) = 0$ *et si* $(x, y) \neq (0, 0)$*,*

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

On peut montrer que les dérivées partielles $\partial_1 f$ *et* $\partial_2 f$ *existent en tout point et* $\partial_1\partial_2 f(0) = 1$ *alors que* $\partial_2\partial_1 f(0) = -1$.

Plus généralement, si $f(x, y) = xyg(x, y)$ *si* $(x, y) \neq (0, 0)$ *et* $f(0) = 0$*, où* g *admet des dérivées partielles secondes en* 0*, alors les dérivées partielles secondes de* f *en zéro ne commutent pas si*

$$\lim_{x\to 0}\left(\lim_{y\to 0} g(x, y)\right) \neq \lim_{y\to 0}\left(\lim_{x\to 0} g(x, y)\right)\,.$$

5.2 Rappels sur les formes bilinéaires

En préparation pour la suite, on fait quelques rappels sur les formes bilinéaires. Une *forme bilinéaire* sur un espace vectoriel E est une application $b : \mathrm{E} \times \mathrm{E} \to \mathbb{R}$, linéaire par rapport à chaque variable. Si E est de dimension finie et $\mathcal{B} = (\mathrm{e}_1, \dots, \mathrm{e}_n)$ est une base de E, on lui associe la matrice

$$\mathrm{Mat}(b, \mathcal{B}) = (b(\mathrm{e}_i, \mathrm{e}_j))_{1\leq i,j\leq n} \in \mathcal{M}_n(\mathbb{R})\,.$$

On peut vérifier que b est entièrement déterminée par $\mathcal{B}$: si $x = \sum_i x_i \mathrm{e}_i$ et $y = \sum_j y_j \mathrm{e}_j$ alors

$$b(x, y) = \sum_{1\leq i,j\leq n} x_i y_j b(\mathrm{e}_i, \mathrm{e}_j)\,.$$

Matriciellement, cela s'écrit

$$b(x, y) = \begin{pmatrix} x_1 & \dots & x_n \end{pmatrix} \mathrm{Mat}(b, \mathcal{B}) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = {}^t\!x\,\mathrm{Mat}(b, \mathcal{B}) y\,.$$

Si $\mathcal{B}'$ est une autre base de E et P est la matrice de passage de $\mathcal{B}$ vers $\mathcal{B}'$, alors on a

$$\mathrm{Mat}(b, \mathcal{B}') = {}^t\mathrm{P}\,\mathrm{Mat}(b, \mathcal{B})\mathrm{P}\,.$$

Enfin, on dit que b est *symétrique* si $b(x,y)=b(y,x)$, ce qui est équivalent à la matrice de b à être symétrique. Si b est symétrique, alors il existe une base dans laquelle b est diagonale.

On dit qu'une forme bilinéaire symétrique b est *non dégénérée* si $\mathrm{Mat}(b,\mathcal{B})$ est inversible. La forme b, symétrique, est *positive* si, pour tout $x\in \mathrm{E}$, on a $b(x,x)\geq 0$; elle est *définie positive* si, pour tout $x\neq 0$, on a $b(x,x)>0$. On a des définitions analogues pour b négative et définie négative.

On remarque que b est positive si les valeurs propres de $\mathrm{Mat}(b,\mathcal{B})$ sont positives, et b est définie positive si les valeurs propres sont strictement positives.

5.3 Hessienne

On suppose dans ce paragraphe que U est un ouvert de $\mathbb{R}^n$ et $f:\mathrm{U}\to\mathbb{R}$ est une application à valeurs réelles.

Définition 5.9. — *Si $f:\mathrm{U}\to\mathbb{R}$ est deux fois différentiable au point $a\in\mathrm{U}$, alors on définit la hessienne $\mathrm{H}_f(a)$ au point a par la matrice de terme général $(\partial^2_{i,j}f(a))_{i,j}$.*

La hessienne est donc la matrice de $\mathrm{D}^2_a f$ dans la base canonique. Il découle du théorème de Schwarz (et de ses corollaires) que la hessienne est une forme bilinéaire symétrique et on a

$$\mathrm{D}^2_a f(h,k) = {}^t h \mathrm{H}_f(a) k$$

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

Proposition 5.10. — *Si $f:\mathrm{U}\to\mathbb{R}$ est deux fois différentiable au point a, alors il existe $\varepsilon:\mathrm{U}\to\mathbb{R}$ telle que*

$$f(x)=f(a)+\mathrm{D}_a f(x-a)+\frac{1}{2}\,{}^t(x-a)\mathrm{H}_f(a)(x-a)+\|x-a\|^2\varepsilon(x)$$

où $\lim_{x\to a}\varepsilon(x)=0$.

Démonstration. — On considère la fonction auxiliaire $g:\mathrm{U}\to\mathbb{R}$ définie par

$$g(x)=f(x)-f(a)-\mathrm{D}_a f(x-a)-\frac{1}{2}\,{}^t(x-a)\mathrm{H}_f(a)(x-a)$$

que l'on différentie pour x assez proche de a en se souvenant que le dernier terme est bilinéaire en x, cf. § 3.1:

$$\mathrm{D}_x g(h)=\mathrm{D}_x f(h)-\mathrm{D}_a f(h)-{}^t(x-a)\mathrm{H}_f(a)h=\mathrm{D}_x f(h)-\mathrm{D}_a f(h)-\mathrm{D}^2_a f(x-a,h)\,.$$

Soit $\varepsilon>0$; comme $\mathrm{D}f$ est différentiable au point a, il existe $\delta>0$ telle que, si $\|x-a\|<\delta$, alors $\|\mathrm{D}_x g\|\leq\varepsilon\|x-a\|$. Du coup, le théorème de la moyenne implique, pour $x\in\mathrm{B}(a,\delta)$,

$$\begin{aligned}\|f(x)-f(a)-\mathrm{D}_a f(x-a)-\frac{1}{2}\,{}^t(x-a)\mathrm{H}_f(a)(x-a)\| &= \|g(x)-g(a)\|\\ &\leq \sup_{y\in[x,a]}\|\mathrm{D}_y g\|\cdot\|x-a\|\\ &\leq \varepsilon\|x-a\|^2\,.\end{aligned}$$

Ceci conclut la démonstration. ■

Corollaire 5.11 (extrema locaux). — *Soit $f:\mathrm{U}\to\mathbb{R}$ une application deux fois différentiables. Si a est un maximum local (resp. un minimum local) alors $\mathrm{D}_a f=0$ et $\mathrm{H}_f(a)$ est négative (resp. positive). Réciproquement, si $\mathrm{D}_a f=0$ et $\mathrm{H}_f(a)$ est définie positive (resp. définie négative) alors a est un minimum local strict (resp. maximum local strict).*

Remarque 5.12. — *$\mathrm{H}_f(a)$ est (définie) positive si ses valeurs propres sont (strictement) positives, et $\mathrm{H}_f(a)$ est (définie) négative si ses valeurs propres sont (strictement) négatives. Si les valeurs propres sont de signe différent, alors a n'est pas un extremum.*

Démonstration. — Supposons que a est un maximum local. La proposition 3.12 nous affirme que $\mathrm{D}_a f = 0$. Du coup, sur un voisinage $\mathrm{B}(a, r)$ suffisamment petit de a, la proposition précédente nous affirme, que pour $h \in \mathrm{E}$ et $t \in [0, r/\|h\|]$,

$$f(a+th) = f(a) + \frac{t^2}{2}\,{}^t h \mathrm{H}_f(a) h + t^2\|h\|^2 \varepsilon(th) \leq f(a)$$

d'où

$$\frac{1}{2}\,{}^t h \mathrm{H}_f(a) h \leq -\|h\|^2 \varepsilon(th)\,.$$

En prenant la limite quand t tend vers 0, on obtient ${}^t h \mathrm{H}_f(a) h \leq 0$. Comme h est arbitraire, on en déduit que $\mathrm{H}_f(a)$ est négative.

Si a est un minimum local, le même argument montre ${}^t h \mathrm{H}_f(a) h \geq 0$ pour tout $h \in \mathrm{E}$, impliquant que $\mathrm{H}_f(a)$ est positive.

Supposons maintenant $\mathrm{D}_a f = 0$ et $\mathrm{H}_f(a)$ définie positive. Alors, pour h assez petit, on a

$$f(a+h) = f(a) + \frac{1}{2}\,{}^t h \mathrm{H}_f(a) h + \|h\|^2 \varepsilon(h)\,.$$

D'après le lemme 5.13 ci-dessous, il existe $c > 0$ tel que ${}^t h \mathrm{H}_f(a) h \geq c\|h\|^2$. Prenons maintenant h assez proche de 0 pour que $\|\varepsilon(h)\| \leq (c/4)$. Il vient

$$f(a+h) \geq f(a) + \frac{c}{4}\|h\|^2 > f(a)\,.$$

Donc a est un minimum local strict.

Si $\mathrm{H}_f(a)$ est définie négative, alors $-\mathrm{H}_f(a)$ est définie positive, donc il existe $c > 0$ telle que ${}^t h \mathrm{H}_f(a) h \leq -c\|h\|^2$. Du coup, l'argument précédent montre

$$f(a+h) \leq f(a) - \frac{c}{4}\|h\|^2 < f(a)\,.$$

Donc a est un maximum local strict. ∎

Lemme 5.13. — *Soit $b : \mathrm{E}\times\mathrm{E} \to \mathbb{R}$ une forme bilinéaire définie positive définie sur un evn de dimension finie. Il existe $c > 0$ telle que, pour tout $x \in \mathrm{E}$, on ait $b(x,x) \geq c\|x\|^2$.*

Démonstration. — Soit $\mathrm{S} = \{x \in \mathrm{E}, \|x\| = 1\}$. L'application $x \mapsto b(x,x)$ est continue d'après la proposition 2.14, donc atteint son minimum sur S qui est compact en un vecteur $x_0 \neq 0$. Du coup, on a, pour tout $x \in \mathrm{E}$, $b(x,x) \geq b(x_0,x_0)\|x\|^2$, avec $b(x_0,x_0) > 0$. ∎

Corollaire 5.14 (Lagrange). — *Soit U un ouvert de $\mathbb{R}^2$ et soit $f : \mathrm{U} \to \mathbb{R}$ une application deux fois différentiable en $a \in \mathrm{U}$ tel que $\mathrm{D}_a f = 0$. On pose*

$$r = \frac{\partial^2 f}{\partial x^2}(a), \quad s = \frac{\partial^2 f}{\partial x \partial y}(a), \quad t = \frac{\partial^2 f}{\partial y^2}(a)\,.$$

- *Si $rt - s^2 > 0$ alors f admet au point a un extremum local strict. Si $r > 0$, alors a est un minimum et si $r < 0$, alors a est un maximum.*
- *Si $rt - s^2 < 0$, alors f n'admet pas d'extremum local au point a.*

Démonstration. — On vérifie que $\mathrm{H}_f(a)$ est définie positive ou négative si $rt - s^2 > 0$, avec $\mathrm{H}_f(a)$ positive si $r > 0$ et négative si $r < 0$. En effet, le polynôme caractéristique de $\mathrm{H}_f(a)$ est

$$\chi(\lambda) = \lambda^2 - (r+t)\lambda + (rt - s^2)\,.$$

Du coup, le discriminant vaut $(r+t)^2 - 4(rt - s^2) = (r-t)^2 + 4s^2 \geq 0$ et on a bien deux valeurs propres réelles, dont le produit est $rt - s^2$. Si le produit est strictement positif, alors les deux valeurs propres sont du même signe (et non nul), donc $\mathrm{H}_f(a)$ est définie positive ou négative. Si le produit est strictement négatif, alors les valeurs propres sont de signe opposé, donc $\mathrm{H}_f(a)$ n'est ni positive ni négative. Du coup a ne peut être un extremum local d'après la condition nécesaire du corollaire 5.11.

Si $rt - s^2 > 0$ alors r et t sont de même signe, donc le signe de la somme des valeurs propres est celui de r. ∎

6 Différentielles d'ordre quelconque

On adapte le paragraphe précédent aux différentielles d'ordre supérieur. On commence par identifier leur nature.

Lemme 6.1. — *Soient* E,F *deux espaces vectoriels normés. On note* $\mathcal{L}_1(\mathrm{E},\mathrm{F})=\mathcal{L}(\mathrm{E},\mathrm{F})$*, puis par récurrence* $\mathcal{L}_{n+1}(\mathrm{E},\mathrm{F})=\mathcal{L}(\mathrm{E},\mathcal{L}_n(\mathrm{E},\mathrm{F}))$*. Pour tout* $n\geq 1$*, il existe un isomorphisme canonique entre* $\mathcal{L}_n(\mathrm{E},\mathrm{F})$ *et l'espace* $\mathcal{L}^n(\mathrm{E},\mathrm{F})$ *des applications* n*-linéaires* $f:\mathrm{E}^n\to\mathrm{F}$ *préservant les normes.*

On rappelle que $\mathcal{L}^n(\mathrm{E},\mathrm{F})$ est muni de la norme

$$\|\mathrm{L}\|=\sup_{\|x_1\|=\|x_2\|=\ldots=\|x_n\|=1}\|\mathrm{L}(x_1,\ldots,x_n)\|\,.$$

Démonstration. — On procède par récurrence. Les cas $n=1,2$ ont déjà été traités. Supposons que la conclusion du lemme soit vraie en rang n, et montrons-la au rang $n+1$. On définit $\Phi:\mathcal{L}_{n+1}(\mathrm{E},\mathrm{F})\to\mathcal{L}^{n+1}(\mathrm{E},\mathrm{F})$ en posant

$$\Phi(\mathrm{L})(x_1,x_2,\ldots,x_{n+1})=\mathrm{L}(x_1)(x_2,\ldots,x_{n+1})\,.$$

La linéarité par rapport à la première variable découle du fait que $\mathrm{L}\in\mathcal{L}(\mathrm{E},\mathcal{L}_n(\mathrm{E},\mathrm{F}))$, et la linéarité par rapport aux autres variables vient de l'hypothèse de récurrence.

Le reste de la démonstration suit celle du lemme 5.2. ■

Définition 6.2. — *Soit* $p\geq 2$*. On dit que* $f:\mathrm{U}\to\mathrm{F}$ *est* p *fois différentiable au point* $a\in\mathrm{U}$ *si* f *est* $(p-1)$ *différentiable sur un voisinage* V *de* a *et si la différentielle de rang* $(p-1)$ $\mathrm{D}^{p-1}f:\mathrm{V}\to\mathcal{L}_{p-1}(\mathrm{E},\mathrm{F})$ *est différentiable au point* a*. L'application est* p *fois différentiable sur* U *si* f *est* p *fois différentiable en tout point de* U*. Elle est de classe* $\mathcal{C}^p$ *si* f *est* p *fois différentiable et si la différentielle* p*-ième* $a\mapsto\mathrm{D}^p_af$ *est continue. On dit que* f *est de classe* $\mathcal{C}^\infty$ *sur* U *si* f *est différentiable* p *fois pour tout* $p\geq 1$*.*

Si f est p fois différentiable, on obtient par récurrence

$$\mathrm{D}_a(\mathrm{D}(\ldots(\mathrm{D}(\mathrm{D}f(h_1))(h_2))\ldots)(h_{p-1}))(h_p)=\mathrm{D}^p_af(h_p,\ldots,h_1)\,.$$

Remarque 6.3. — *Si* f *est* p*-différentiable au voisinage de* a *et si* D^pf *est* q *fois différentiable, alors* f *est* $(p+q)$ *fois différentiable et* $\mathrm{D}^{p+q}_af=\mathrm{D}^q_a(\mathrm{D}^pf)$*.*

6.1 Symétrie des différentielles et dérivées partielles

Une application k-linéaire $f:\mathrm{E}^k\to\mathrm{F}$ est *symétrique* si, pour toute permutation $\sigma\in\mathfrak{S}_k$ et tout $(x_1,\ldots,x_k)\in\mathrm{E}^k$, on a

$$f(x_1,\ldots,x_k)=f(x_{\sigma(1)},\ldots,x_{\sigma(k)})\,.$$

Théorème 6.4 (de Schwarz généralisé). — *Si* f *est* p *fois différentiable au point* a*, alors* D^p_af *est une application* p*-linéaire symétrique.*

Démonstration. — On procède par récurrence sur p. Le cas $p=2$ est connu, donc supposons $p\geq 3$ et les cas $j<p$ connus. Soit $(h_1,\ldots,h_p)\in\mathrm{E}^p$. Si f est p fois différentiable, alors l'application $x\mapsto\mathrm{D}^{p-2}_xf(h_3,\ldots,h_p)$ est deux fois différentiable au point a, donc le théorème de Schwarz implique

$$\mathrm{D}^p_af(h_1,h_2,\ldots,h_p)=\mathrm{D}^p_af(h_2,h_1,\ldots,h_p)\,.$$

Or, notre hypothèse de récurrence implique aussi que pour toute permutation $\sigma\in\mathfrak{S}_{p-1}$, on a

$$\mathrm{D}^{p-1}_xf(h_2,\ldots,h_p)=\mathrm{D}^{p-1}_xf(h_{\sigma(2)},\ldots,h_{\sigma(p)})$$

pour x assez proche de a, donc en différentiant au point a, on obtient

$$\mathrm{D}^p_af(h_1,h_2,\ldots,h_p)=\mathrm{D}^p_af(h_1,h_{\sigma(2)},\ldots,h_{\sigma(p)})\,.$$

Pour conclure, on remarque que $\mathfrak{S}_p$ est engendré par la transposition $(1,2)$ et les permutations qui fixent 1. On en déduit donc

$$\mathrm{D}^p_af(h_1,h_2,\ldots,h_p)=\mathrm{D}^p_af(h_{\sigma(1)},\ldots,h_{\sigma(p)})$$

pour toute permutation $\sigma\in\mathfrak{S}_p$. ■

Soient U un ouvert de $\mathbb{R}^n$ et $f : \mathrm{U} \to \mathrm{F}$ une application. Pour $p \geq 1$ et $j_1, \ldots, j_p \in \{1, \ldots, n\}$, on définit la dérivée partielle par le vecteur

$$\partial^p_{j_1,\ldots,j_p} f = \frac{\partial^p f}{\partial x_{j_1} \ldots \partial x_{j_p}} \stackrel{\text{def.}}{=} \partial_{j_1}(\partial_{j_2}(\ldots(\partial_{j_p} f)\ldots))$$

s'il est bien défini. Si f est p fois différentiable, alors on obtient par récurrence

$$\partial^p_{j_1,\ldots,j_p} f(a) = \mathrm{D}^p_a f(\mathrm{e}_{j_1}, \ldots, \mathrm{e}_{j_p}) \,.$$

De plus, si, pour $j \in \{1, \ldots, p\}$, on note

$$h_j = \begin{pmatrix} h_j^1 \\ \vdots \\ h_j^n \end{pmatrix}$$

alors on obtient, sans tenir compte des symétries,

$$\mathrm{D}^p_a f(h_1, \ldots, h_p) = \sum_{j_1,\ldots,j_p=1}^{n} \partial^p_{j_1,\ldots,j_p} f(a) h_1^{j_1} \ldots h_p^{j_p} \,.$$

Cette formule et le théorème de Schwarz généralisé impliquent

Corollaire 6.5. — *Si f est p fois différentiable au point a, alors, pour toute permutation $\sigma \in \mathfrak{S}_p$, on a*

$$\partial^p_{j_1,\ldots,j_p} f(a) = \partial^p_{\sigma(j_1),\ldots,\sigma(j_p)} f \,.$$

6.2 Propriétés et exemples

On énonce quelques propriétés qui découlent sans malice des propriétés déjà établies des applications différentiables en procédant par récurrence.

Proposition 6.6. — *Soient* U *un ouvert d'un evn* E, $a \in \mathrm{U}$, $p \geq 1$ *et* $f : \mathrm{U} \to \mathrm{F}$ *une application à valeurs dans un evn* F.

1. *Si* E *est un produit fini* $\mathrm{E}_1 \times \ldots \times \mathrm{E}_k$ *d'evn, alors* f *est de classe* $\mathcal{C}^k$ *si et seulement si les différentielles partielles d'ordre au plus* p *existent et sont continues le long de chaque sous-espace* E_j.
2. *Si* F *est un produit fini d'espaces* $\mathrm{F}_1 \times \ldots \times \mathrm{F}_k$, *alors* $f = (f_1, \ldots, f_k)$ *est de classe* $\mathcal{C}^p$ *si et seulement si* f_j *est de classe* $\mathcal{C}^p$ *pour chaque* $j \in \{1, \ldots, k\}$.
3. *La combinaison linéaire d'applications de classe* $\mathcal{C}^p$ *dans* U *est de classe* $\mathcal{C}^p$ *dans* U.
4. *Si* $\mathrm{L} : \mathrm{F} \to \mathrm{G}$ *est une application linéaire et* f *est différentiable* p *fois au point* a, *alors* $\mathrm{L} \circ f$ *est différentiable* p *fois et* $\mathrm{D}^p_a(\mathrm{L} \circ f) = \mathrm{L} \circ \mathrm{D}^p_a f$. *De même, si* $\mathrm{L} : \mathrm{G} \to \mathrm{E}$ *est linéaire,* $\mathrm{L}(b) = a$, *alors* $(f \circ \mathrm{L})$ *est* p *fois différentiable au point* b *et* $\mathrm{D}^p_b(f \circ \mathrm{L})(h_1, \ldots, h_p) = \mathrm{D}^p_a f(\mathrm{L}(h_1), \ldots, \mathrm{L}(h_p)))$.

Cela nous permet de montrer

Exemple 6.7. — *Une application k-linéaire* $f : \mathrm{E}_1 \times \ldots \times \mathrm{E}_k \to \mathrm{F}$ *est de classe* $\mathcal{C}^\infty$ *et les différentielles d'ordre* $p \geq k+1$ *sont toutes nulles.*

Démonstration. — On procède par récurrence sur k. Pour le cas $k = 1$, on sait déjà que $\mathrm{D}_a f = f$ pour tout $a \in \mathrm{E}_1$, donc $\mathrm{D}_{a+h} f - \mathrm{D}_a f = f - f = 0$. Supposons maintenant le cas $(k-1)$ connu et considérons une transformation k-linéaire. D'après le théorème 4.7, on obtient

$$\mathrm{D}_a f(h_1, \ldots, h_k) = \sum_{j=1}^{k} f(a_1, \ldots, a_{j-1}, \underbrace{\quad h_j \quad}_{j\text{ème place}}, a_{j+1}, \ldots, a_k)$$

soit

$$\mathrm{D}_a f(h_1, \ldots, h_k) = f(h_1, a_2, \ldots, a_k) + f(a_1, h_2, a_3, \ldots, x_k) + \ldots + f(a_1, \ldots, a_{k-1}, h_k) \,.$$

Donc $\mathrm{D}f$ est la somme finie de la composée de la projection de $\mathrm{E}_1 \times \ldots \times \mathrm{E}_k$ sur $\mathrm{E}_1 \times \ldots \times \mathrm{E}_{j-1} \times \mathrm{E}_{j+1} \times \ldots \times \mathrm{E}_k$ et de l'application $(k-1)$-linéaire de $\mathrm{E}_1 \times \ldots \times \mathrm{E}_{j-1} \times \mathrm{E}_{j+1} \times \ldots \times \mathrm{E}_k$ sur $\mathcal{L}(\mathrm{E}_j, \mathrm{F})$ définie par

$$(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k) \mapsto (h \in \mathrm{E}_j \mapsto f(x_1, , \ldots, x_{j-1}, h, x_{j+1}, \ldots, x_k)) \, .$$

Donc l'hypothèse de récurrence s'applique et la proposition précédente permet de conclure. ∎

Proposition 6.8. — *Soient $p \in \mathbb{N} \cup \{\infty\}$, $\mathrm{E}, \mathrm{F}, \mathrm{G}$ trois evns, U un ouvert de E, V un ouvert de F, a un point de U, et $f : \mathrm{U} \to \mathrm{F}$ et $g : \mathrm{V} \to \mathrm{G}$ deux applications telles que $f(\mathrm{U}) \subset \mathrm{V}$.*

1. *Si f est p fois différentiable au point a et g est p fois différentiable au point $f(a)$ alors $(g \circ f)$ est p fois différentiable au point a.*

2. *Si f et g sont de classe $\mathcal{C}^p$ alors $(g \circ f)$ est de classe $\mathcal{C}^p$.*

Démonstration. — On procède par récurrence: le cas $p = 1$ est connu, et de plus $\mathrm{D}_a(g \circ f) = \mathrm{D}_{f(a)}g \circ \mathrm{D}_a f$. Supposons maintenant la proposition connue au rang $(p-1) \geq 1$ et montrons-la au rang p. Sans perte de généralité, on peut supposer que f est p fois différentiable sur tout U et g sur tout V. On remarque que $x \mapsto \mathrm{D}_x(g \circ f)$ est la composée de $\varphi : \mathrm{U} \to \mathcal{L}(\mathrm{E}, \mathrm{F}) \times \mathcal{L}(\mathrm{F}, \mathrm{G})$ définie par $\varphi(x) = (\mathrm{D}_x f, \mathrm{D}_{f(x)} g)$ et de l'application bilinéaire $\psi : \mathcal{L}(\mathrm{E}, \mathrm{F}) \times \mathcal{L}(\mathrm{F}, \mathrm{G}) \to \mathcal{L}(\mathrm{E}, \mathrm{G})$ définie par $\psi(\mathrm{A}, \mathrm{B}) = \mathrm{B} \circ \mathrm{A}$. D'après l'exemple précédent, ψ est de classe $\mathcal{C}^\infty$, et les applications composantes de φ sont $(p-1)$ fois différentiables, donc φ aussi. On en déduit que $(g \circ f)$ est p fois différentiable. Si les différentielles d'ordre p sont continues, alors on en déduit que $(g \circ f)$ est aussi de classe $\mathcal{C}^p$. ∎

On obtient par récurrence

Proposition 6.9. — *Soient U un ouvert d'un evn et $f, g : \mathrm{U} \to \mathbb{R}$ des applications p fois différentiables (resp. de classe $\mathcal{C}^p$), alors le produit (fg) est aussi p fois différentiable (resp. de classe $\mathcal{C}^p$).*

Exemple 6.10. — *Toute application polynomiale ou rationnelle $f : \mathbb{R}^n \to \mathbb{R}$ est de classe $\mathcal{C}^\infty$ là où elle est définie.*

Exemple 6.11. — *L'application $f : \mathrm{GL}(\mathrm{E}, \mathrm{F}) \to \mathrm{GL}(\mathrm{F}, \mathrm{E})$ définie par $f(u) = u^{-1}$ est de classe $\mathcal{C}^\infty$.*

Démonstration. — Soit $\Phi : \mathrm{E} \to \mathrm{F}$ un isomorphisme. On considère $g : \mathrm{GL}(\mathrm{E}) \to \mathrm{GL}(\mathrm{E})$ définie par $g(u) = u^{-1}$ et les applications linéaires $\mathrm{A}_\Phi : \mathcal{L}(\mathrm{E}, \mathrm{F}) \to \mathcal{L}(\mathrm{E})$ et $\mathrm{B}_\Phi : \mathcal{L}(\mathrm{E}) \to \mathcal{L}(\mathrm{F}, \mathrm{E})$ définies par $\mathrm{A}_\Phi(u) = \Phi^{-1} \circ u$ et $\mathrm{B}_\Phi(u) = u \circ \Phi^{-1}$. On a vu $\mathrm{D}_u g(h) = -u^{-1} \circ h \circ u^{-1}$. Or $f = \mathrm{B}_\Phi \circ g \circ \mathrm{A}_\Phi$ de sorte que, par linéarité et composition, on obtient aussi $\mathrm{D}_u f(h) = -u^{-1} \circ h \circ u^{-1}$. On l'écrit sous la forme $\mathrm{D}_u f = \psi \circ \varphi$ où $\varphi : \mathrm{GL}(\mathrm{E}, \mathrm{F}) \to \mathcal{L}(\mathrm{F}, \mathrm{E}) \times \mathcal{L}(\mathrm{F}, \mathrm{E})$ est définie par $\varphi(u) = (u^{-1}, u^{-1})$ et $\psi : \mathcal{L}(\mathrm{F}, \mathrm{E}) \times \mathcal{L}(\mathrm{F}, \mathrm{E}) \to \mathcal{L}(\mathcal{L}(\mathrm{E}, \mathrm{F}), \mathcal{L}(\mathrm{F}, \mathrm{E}))$ est définie par $\psi(v_1, v_2)(h) = -v_1 \circ h \circ v_2$. Par récurrence, l'application φ est de classe $\mathcal{C}^p$ si f est de classe $\mathcal{C}^p$ et ψ est de classe $\mathcal{C}^\infty$ car bilinéaire. Donc $\mathrm{D}f$ est de classe $\mathcal{C}^p$ et f de classe $\mathcal{C}^{p+1}$. ∎

7 Développements limités

On adapte les formules de Taylor aux applications de plusieurs variables. Les hypothèses chaque fois sont différentes, et il convient de bien les retenir. On commence par le cas de polynômes où on obtient une formule exacte.

7.1 Polynômes généralisés

Définition 7.1. — *Soient E, F deux espaces vectoriels de dimension finie.*

1. *Soit $k \geq 0$; une application $f : \mathrm{E} \to \mathrm{F}$ est un polynôme homogène de degré k s'il existe une application k-linéaire symétrique $g : \mathrm{E}^k \to \mathrm{F}$ telle que $f(x) = g(x, \ldots, x)$. Si $k = 0$, f est constante.*

2. *Soit $d \geq 0$; un polynôme de degré d $\mathrm{P} : \mathrm{E} \to \mathrm{F}$ est une combinaison linéaire de polynômes homogènes de degré $k \leq d$ de sorte que le polynôme homogène de degré d est non identiquement nul.*

Théorème 7.2. — *Soit* $\mathrm{P} : \mathrm{E} \to \mathrm{F}$ *un polynôme de degré* $d \geq 0$. *Alors*

$$\mathrm{P}(x) = \sum_{k=0}^{d} \frac{1}{k!} \mathrm{D}_0^k \mathrm{P}(\underbrace{x, \ldots, x}_{k \text{ fois}}) .$$

On traite d'abord le cas des polynômes homogènes.

Proposition 7.3. — *Soit* $f : \mathrm{E} \to \mathrm{F}$ *un polynôme homogène de degré* $k \geq 0$, *alors* f *est de classe* $\mathcal{C}^\infty$, $\mathrm{D}^p f = 0$ *si* $p > k$ *et, si* $p \leq k$,

$$\mathrm{D}_x^p f(h, \ldots, h) = \frac{k!}{(k-p)!} g(\underbrace{x, \ldots, x}_{k-p \text{ fois}}, \underbrace{h, \ldots, h}_{p \text{ fois}})$$

où $f(x) = g(x, \ldots, x)$.

Démonstration. — On considère pour chaque $j \geq 1$, l'application linéaire $\Delta_j : \mathrm{E} \to \mathrm{E}^j$ définie par $\Delta_j(x) = (\underbrace{x, \ldots, x}_{j \text{ fois}})$ de sorte que $f = g \circ \Delta_k$ pour un polynôme homogène de degré k.

On procède par récurrence sur k. Prenons $k = 1$, de sorte que $f = g \in \mathcal{L}(\mathrm{E}, \mathrm{F})$. Si $p = 0$ alors $\mathrm{D}_x^0 f = f(x)$; si $p = 1$ alors $\mathrm{D}_x f = f = g$ pour tout $x \in \mathrm{E}$. Du coup, $\mathrm{D}_{x+h} f - \mathrm{D}_x f = 0$ impliquant ainsi $\mathrm{D}^2 f = 0$ et plus généralement $\mathrm{D}^p f = 0$ pour tout $p \geq 2$.

Supposons maintenant la proposition établie jusqu'au rang $(k-1) \geq 1$ et prenons un polynôme homogène de degré k de sorte que $f = g \circ \Delta_k$. D'après l'exemple 6.7, on obtient

$$\mathrm{D}_x f(h) = \mathrm{D}_{\Delta_k(x)} g(\Delta_k(h)) = k g(\underbrace{x, \ldots, x}_{k-1}, h) = k g(\Delta_{k-1}(x), \Delta_1(h))$$

par symétrie, donc $\mathrm{D}f$ est un polynôme homogène de degré $(k-1)$ à valeurs dans $\mathcal{L}(\mathrm{E}, \mathrm{F})$. L'hypothèse de récurrence s'applique de sorte que

- Si $p \leq k$, alors

  $$\mathrm{D}_x^{p-1}(\mathrm{D}f) \circ \Delta_{p-1} = \frac{k \times (k-1)!}{[(k-1)-(p-1)]!} g(\Delta_{k-p}(x), \Delta_{p-1}, \Delta_1)$$

  donc

  $$\mathrm{D}_x^p f \circ \Delta_p = \mathrm{D}_x^{p-1}(\mathrm{D}f) \circ \Delta_p = \frac{k!}{(k-p)!} g(\Delta_{k-p}(x), \Delta_p) .$$

- Si $p > k$ alors $p - 1 > k - 1$ et $\mathrm{D}_x^p f = \mathrm{D}_x^{p-1}(\mathrm{D}f) = 0$.

■

Démonstration du théorème 7.2. — On considère pour chaque $j \geq 1$, l'application linéaire Δ_j : $\mathrm{E} \to \mathrm{E}^j$ définie par $\Delta_j(x) = (\underbrace{x, \ldots, x}_{j \text{ fois}})$. Il existe des polynômes homogènes P_k et des applications k-linéaires et symétriques g_k de degré k pour k variant de 0 à d tels que $\mathrm{P}_k = g_k \circ \Delta_k$ et $\mathrm{P} = \sum_{k=0}^d \mathrm{P}_k$.

D'après la proposition précédente, on a

$$\mathrm{D}_x^p \mathrm{P}_k(\Delta_p(h)) \quad = \quad \begin{cases} \frac{k!}{(k-p)!} g_k(\Delta_{k-p}(x), \Delta_p(h)) & \text{si } 0 \leq p \leq k \\ 0 & \text{si } p > k \end{cases}$$

Par conséquent, $\mathrm{D}_0^k \mathrm{P} = k! g_k$ et

$$\sum_{k=0}^{d} \frac{1}{k!} \mathrm{D}_0^k \mathrm{P}(\Delta_k(x)) = \sum_{k=0}^{d} g_k(\Delta_k(x)) = \mathrm{P}(x) .$$

■

Si $f : \mathrm{U} \to \mathrm{F}$ est une application p fois différentiable, alors ses différentielles sont des applications multilinéaires symétriques, donc on peut appliquer ce qui précède. On introduit d'abord quelques notations.

Notations.— Afin de simplifier les notations, on introduit le formalisme suivant. Un *multi-indice* est un n-uplet d'entiers naturels $\alpha = (\alpha_1, \ldots, \alpha_n)$, $\alpha_j \in \mathbb{N}$. Sa *longueur* vaut $|\alpha| = \sum_{j=1}^{n} \alpha_j$.

On définit aussi $\alpha! = \alpha_1! \times \ldots \times \alpha_n!$ et $h^\alpha = h_1^{\alpha_1} \ldots h_n^{\alpha_n}$ si $h = (h_1, \ldots, h_n)$. Enfin, pour une application p fois différentiable définie sur un ouvert de $\mathbb{R}^n$, on écrit, pour un multi-indice de longueur p

$$\partial_\alpha f = \partial_1^{\alpha_1} \ldots \partial_n^{\alpha_n} f = \frac{\partial^p f}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}}$$

où le symbole $\partial_j^{\alpha_j}$ signifie que l'on dérive α_j fois dans la direction e_j.

Lemme 7.4. — *Étant donné un multi-indice $\alpha \in \mathbb{N}^n$ de longueur p, le nombre $m(\alpha)$ de p-uplet $(j_1, \ldots, j_p)$ qui comportent exactement α_k termes prenant la valeur k vaut*

$$m(\alpha) = \frac{p!}{\alpha!} = \frac{p!}{\alpha_1! \ldots \alpha_n!}\,.$$

Démonstration. — Le nombre $m(\alpha)$ est déterminé par le choix de α_1 indices parmi p, puis de α_2 indices parmi les $(p - \alpha_1)$ restant, etc. D'où

$$m(\alpha) = \binom{p}{\alpha_1} \times \binom{p-\alpha_1}{\alpha_2} \times \ldots \times \binom{p - \sum_{j=1}^{n-1} \alpha_j}{\alpha_n} = \frac{p!}{\alpha_1! \ldots \alpha_n!}\,.$$

■

Corollaire 7.5 (formule du binôme généralisé). — *On a, pour $x_1, \ldots, x_n \in \mathbb{R}$, $k \geq 1$,*

$$\left(\sum_{j=1}^{n} x_j\right)^k = \sum_{\alpha \in \mathbb{N}^n,\ |\alpha| = k} \frac{k!}{\alpha!} x^\alpha\,.$$

Corollaire 7.6 (expression développée de la différentielle). — *Soit $f : \mathrm{U} \to \mathrm{F}$ une application p fois différentiable en un point a d'un ouvert U de $\mathbb{R}^n$. On a, pour tout $h \in \mathbb{R}^n$,*

$$\frac{1}{p!} \mathrm{D}_a^p f(\underbrace{h, \ldots, h}_{p}) = \sum_{\alpha \in \mathbb{N}^n, |\alpha| = p} \frac{1}{\alpha!} \partial_\alpha f(a) h^\alpha\,.$$

7.2 Formule de Taylor-Young

Cette formule décrit le comportement local d'une fonction au voisinage d'un point.

Théorème 7.7 (Taylor-Young). — *Soit $f : \mathrm{U} \to \mathrm{F}$ une application différentiable p fois au point $a \in \mathrm{U}$. Soit $r > 0$ tel que $\mathrm{B}(a, r) \subset \mathrm{U}$. Il existe $\varepsilon : \mathrm{B}(0, r) \to \mathrm{F}$ telle que $\lim_0 \varepsilon = 0$ et, pour tout $h \in \mathrm{B}(0, r)$,*

$$f(a + h) = \sum_{k=0}^{p} \frac{1}{k!} \mathrm{D}_a^k f(\underbrace{h, \ldots, h}_{k}) + \|h\|^p \varepsilon(h)\,.$$

Démonstration. — On procède par récurrence. Le cas $p = 1$ est connu. On suppose que la formule est établi jusqu'au rang $(p - 1) \geq 1$. On considère la fonction auxiliaire $g : \mathrm{B}(0, r) \to \mathrm{F}$ définie par

$$g(x) = f(a + x) - \sum_{k=0}^{p} \frac{1}{k!} \mathrm{D}_a^k f(\underbrace{x, \ldots, x}_{k})$$

que l'on différentie grâce à la formule établies pour les polynômes homogènes:

$$\mathrm{D}_x g(h) = \mathrm{D}_{a+x} f(h) - \sum_{k=0}^{p} \frac{1}{(k-1)!} \mathrm{D}_a^k f(\underbrace{x, \ldots, x}_{k-1}, h)\,.$$

Soit $\varepsilon > 0$; comme $\mathrm{D}f$ est différentiable $(p-1)$ fois au point a, il existe $\delta > 0$ telle que, si $\|x\| < \delta$, alors $\|\mathrm{D}_x g\| \leq \varepsilon \|x\|^{p-1}$. Du coup, le théorème de la moyenne implique, pour $h \in \mathrm{B}(0, \delta)$,

$$\begin{aligned} \left\| f(a+h) - \sum_{k=0}^{p} \frac{1}{k!} \mathrm{D}_a^k f(\underbrace{h, \ldots, h}_{k}) \right\| &= \|g(h) - g(0)\| \\ &\leq \sup_{x \in \mathrm{B}(0, \|h\|)} \|\mathrm{D}_x g\| \cdot \|h\| \\ &\leq \varepsilon \cdot \|h\|^p\,. \end{aligned}$$

■

7.3 Formule de Taylor avec reste intégral

Par convention, si $f = (f_1, \ldots, f_n) : [a, b] \to \mathbb{R}^n$, est une application continue, on pose

$$\int_a^b f(x)dx = \begin{pmatrix} \int_a^b f_1(x)dx \\ \vdots \\ \int_a^b f_n(x)dx \end{pmatrix}$$

Théorème 7.8 (Taylor avec reste intégral). — *Soit $f : \mathrm{U} \to \mathrm{F}$ une application de classe $\mathcal{C}^{p+1}(\mathrm{U})$ définie sur un ouvert convexe. Pour tout $a, b \in \mathrm{U}$, on a*

$$f(b) = \sum_{k=0}^{p} \frac{1}{k!} \mathrm{D}_a^k f(\underbrace{b-a, \ldots, b-a}_{k}) + \frac{1}{p!} \int_0^1 (1-t)^p \mathrm{D}_{a+t(b-a)}^{p+1} f(\underbrace{b-a, \ldots, b-a}_{p+1})\, dt\,.$$

Démonstration. — Cela découle de la formule en dimension 1; Soit $g : \mathrm{I} \to \mathbb{R}$ une fonction de classe $\mathcal{C}^{p+1}(\mathrm{I})$ définie sur un intervalle ouvert contenant $[0, 1]$. Alors on a

$$g(1) = \sum_{k=0}^{p} \frac{1}{k!} g^{(k)}(0) + \frac{1}{p!} \int_0^1 (1-t)^p g^{(p+1)}(t)dt \tag{7.1}$$

Montrons (7.1) par récurrence sur p. Si $p = 0$, c'est le fameux théorème fondamental de l'analyse:

$$g(1) - g(0) = \int_0^1 g'(t)dt\,.$$

Supposons la formule (7.1) établie pour les fonctions de classe $\mathcal{C}^p$, et supposons g de classe $\mathcal{C}^{p+1}$. On part donc de

$$g(1) = \sum_{k=0}^{p-1} \frac{1}{k!} g^{(k)}(0) + \frac{1}{(p-1)!} \int_0^1 (1-t)^{p-1} g^{(p)}(t)dt$$

et on intègre par parties le reste intégral en prenant une primitive de $(1-t)^{p-1}$ de sorte que

$$\begin{aligned} \frac{1}{(p-1)!} \int_0^1 (1-t)^{p-1} g^{(p)}(t)dt &= \frac{-1}{p!} (1-t)^p g^{(p)}(t) \Big|_0^1 + \frac{1}{p!} \int_0^1 (1-t)^p g^{(p+1)}(t)dt \\ &= \frac{1}{p!} g^{(p)}(0) + \frac{1}{p!} \int_0^1 (1-t)^p g^{(p+1)}(t)dt\,. \end{aligned}$$

Du coup, on obtient bien (7.1).

La formule de Taylor avec reste intégral s'obtient en appliquant (7.1) à la fonction auxiliaire $g : [0, 1] \to \mathbb{R}^n$ définie par $g(t) = f(a + t(b-a))$. ■

Corollaire 7.9 (Majoration du reste). — *Soit $f : \mathrm{U} \to \mathrm{F}$ une application de classe $\mathcal{C}^{p+1}(\mathrm{U})$ définie sur un ouvert convexe. Si $\|\mathrm{D}^{p+1}f\| \leq \mathrm{M}$ sur U, alors, pour tout $a, b \in \mathrm{U}$, on a*

$$\left\| f(b) - \sum_{k=0}^{p} \frac{1}{k!}\mathrm{D}_a^k f(\underbrace{b-a, \ldots, b-a}_{k}) \right\| \leq \frac{\mathrm{M}}{(p+1)!}\|b-a\|^{p+1}.$$

DÉMONSTRATION. — On estime le reste intégral

$$\begin{aligned}
\left\| \frac{1}{p!}\int_0^1 (1-t)^p \mathrm{D}_{a+t(b-a)}^{p+1} f(\underbrace{b-a, \ldots, b-a}_{p+1})\, dt \right\| &\leq \frac{1}{p!}\int_0^1 (1-t)^p \|\mathrm{D}_{a+t(b-a)}^{p+1} f(\underbrace{b-a, \ldots, b-a}_{p+1})\| dt \\
&\leq \frac{1}{p!}\int_0^1 (1-t)^p \mathrm{M}\, \|b-a\|^{p+1}\, dt \\
&\leq \frac{\mathrm{M}}{(p+1)!}\|b-a\|^{p+1}
\end{aligned}$$

■

7.4 Formule de Taylor-Lagrange

Dans ce paragraphe, on s'intéresse à des applications à valeurs réelles $f : \mathrm{U} \to \mathbb{R}$ définies sur un ouvert convexe d'un espace vectoriel normé E.

Théorème 7.10 (Taylor-Lagrange). — *Soit $f : \mathrm{U} \to \mathbb{R}$ une application définie sur un ouvert convexe d'un evn E, différentiable $(p+1)$ fois sur U. Pour tous $a, b \in \mathrm{U}$, il existe $c \in]a, b[$ tel que*

$$f(b) = \sum_{k=0}^{p} \frac{1}{k!}\mathrm{D}_a^k f(\underbrace{b-a, \ldots, b-a}_{k}) + \frac{1}{(p+1)!}\mathrm{D}_c^{p+1} f(\underbrace{b-a, \ldots, b-a}_{p+1}).$$

DÉMONSTRATION. — On considère la fonction auxiliaire $g : [0,1] \to \mathbb{R}$ définie par

$$g(t) = \sum_{k=0}^{p} \frac{(1-t)^k}{k!}\mathrm{D}_{a+t(b-a)}^k f(\underbrace{b-a, \ldots, b-a}_{k}) + \mathrm{K}\frac{(1-t)^{p+1}}{(p+1)!} - f(b)$$

où l'on fixe la valeur de K de sorte que $g(0) = 0$. On différentie g, différentiable par hypothèses:

$$\begin{aligned}
g'(t) &= \sum_{k=0}^{p} \frac{(1-t)^k}{k!}\mathrm{D}_{a+t(b-a)}^{k+1} f(\underbrace{b-a, \ldots, b-a}_{k+1}) - \sum_{k=1}^{p} \frac{(1-t)^{k-1}}{(k-1)!}\mathrm{D}_{a+t(b-a)}^k f(\underbrace{b-a, \ldots, b-a}_{k}) - \mathrm{K}\frac{(1-t)^p}{p!} \\
&= \sum_{k=1}^{p+1} \frac{(1-t)^{k-1}}{(k-1)!}\mathrm{D}_{a+t(b-a)}^{k} f(\underbrace{b-a, \ldots, b-a}_{k}) - \sum_{k=1}^{p} \frac{(1-t)^{k-1}}{(k-1)!}\mathrm{D}_{a+t(b-a)}^k f(\underbrace{b-a, \ldots, b-a}_{k}) - \mathrm{K}\frac{(1-t)^p}{p!} \\
&= \frac{(1-t)^p}{p!}\mathrm{D}_{a+t(b-a)}^{p+1} f(\underbrace{b-a, \ldots, b-a}_{p+1}) - \mathrm{K}\frac{(1-t)^p}{p!}.
\end{aligned}$$

Or $g(1) = g(0) = 0$, donc le théorème de Rolle implique l'existence de $\theta \in]0,1[$ tel que $g'(\theta) = 0$, ce qui nous permet d'obtenir

$$\mathrm{K} = \mathrm{D}_{a+\theta(b-a)}^{p+1} f(\underbrace{b-a, \ldots, b-a}_{p+1}).$$

Le fait que $g(0) = 0$ implique, en notant $c = a + \theta(b-a)$,

$$f(b) = \sum_{k=0}^{p} \frac{1}{k!}\mathrm{D}_a^k f(\underbrace{b-a, \ldots, b-a}_{k}) + \frac{1}{(p+1)!}\mathrm{D}_c^{p+1} f(\underbrace{b-a, \ldots, b-a}_{p+1}).$$

■

Remarque 7.11. — *Cette formule n'est pas valide pour des fonctions à valeurs vectoriels. Autrement dit, il n'y a pas de raison que l'on puisse choisir la même valeur du segment pour chaque composante. Prenons par exemple* $g : [0,1] \to \mathbb{R}^2$ *définie par* $g(t) = (t^2, t^3)$. *On a donc* $g'(t) = (2t, 3t)$, $g(1) - g(0) = (1,1)$ *et il n'existe pas de valeurs de* t *vérifiant* $2t = 3t = 1$.

CHAPITRE IV. — INVERSION LOCALE, FONCTIONS IMPLICITES ET RANG CONSTANT

Comme d'habitude, tous nos espaces vectoriels sont de dimension finie. On note $(\mathrm{E}, ||.||_{\mathrm{E}})$ et $(\mathrm{F}, ||.||_{\mathrm{F}})$ des espaces vectoriels normés sur $\mathbb{R}$, U un ouvert non vide de E et $f\ : \mathrm{U} \to \mathrm{F}$ une application. Par un choix de bases pour E et F, on se ramène à $f : \mathrm{U}\,(\subset \mathbb{R}^p)\ \to \mathbb{R}^q$.

Ce chapitre est consacré à la démonstration de trois théorèmes fondamentaux: le théorème d'inversion locale, le théorème des fonctions implicites et le théorème du rang constant. On y trouvera aussi le lemme de Morse, et l'énoncé du théorème de Sard.

8 Difféomorphismes

Soient $\mathrm{U} \subset \mathrm{E}$ et $\mathrm{V} \subset \mathrm{F}$ des ouverts d'espaces vectoriels normés E et F respectivement. Un *difféomorphisme de classe* $\mathcal{C}^1$ $f : \mathrm{U} \to \mathrm{V}$ est un homéomorphisme $f : \mathrm{U} \to \mathrm{V}$ de classe $\mathcal{C}^1$ tel que son application réciproque $f^{-1} : \mathrm{V} \to \mathrm{U}$ est aussi de classe $\mathcal{C}^1$.

Si I et J sont des intervalles de $\mathbb{R}$, alors une application surjective $f : \mathrm{I} \to \mathrm{J}$ est un difféomorphisme si $f \in \mathcal{C}^1$ et si f' ne s'annule pas. Dans ce cas, f est strictement monotone, donc bijective sur son image qui est J. De plus, l'application inverse $g = f^{-1} : \mathrm{J} \to \mathrm{I}$ est dérivable de dérivée $g('x) = 1/f'(g(x))$ qui est bien continue.

Proposition 8.1. — *On a les propriétés suivantes.*

1. *Une application $f : \mathrm{U} \to \mathrm{V}$ est un difféomorphisme si et seulement si*
   a) *f est de classe $\mathcal{C}^1$;*
   b) *f est un homéomorphisme;*
   c) *pour tout $x \in \mathrm{U}$, $\mathrm{D}_x f$ est un isomorphisme.*
2. *Soient $f : \mathrm{U} \to \mathrm{V}$ un difféomorphisme et $u : \mathrm{V} \to \mathrm{G}$ une application à valeurs dans un espace vectoriel normé G. Alors u est de classe $\mathcal{C}^1$ si et seulement si $u \circ f$ est de classe $\mathcal{C}^1$.*

DÉMONSTRATION. — Supposons que f est un difféomorphisme de classe $\mathcal{C}^1$. Par définition, f est de classe $\mathcal{C}^1$ et un homéomorphisme. On écrit $g = f^{-1}$ de sorte que $g \circ f = \mathrm{Id}$. Du coup, pour tout $x \in \mathrm{U}$, $\mathrm{D}_{f(x)} g \circ \mathrm{D}_x f = \mathrm{Id}$, impliquant ainsi que $\mathrm{D}_x f$ est inversible. Réciproquement, si f est un homéomorphisme de classe $\mathcal{C}^1$ de sorte que $\mathrm{D}_x f$ est un isomorphisme pour tout $x \in \mathrm{U}$, alors il reste à montrer que $g = f^{-1}$ est de classe $\mathcal{C}^1$. La proposition 3.8 implique que g est différentiable en chaque point et $\mathrm{D}_y g = (\mathrm{D}_{g(y)} f)^{-1}$ pour tout $y \in \mathrm{V}$. Du coup, $\mathrm{D}g$ est la composée de $\varphi : u \in \mathrm{GL}(\mathrm{E}, \mathrm{F}) \mapsto u^{-1} \in \mathrm{GL}(\mathrm{F}, \mathrm{E})$, $x \mapsto \mathrm{D}_x f$ et g qui sont toutes continues par hypothèses (on a $\mathrm{D}g = \varphi \circ \mathrm{D}f \circ g$). Donc $\mathrm{D}g$ est continue et on conclut que g est de classe $\mathcal{C}^1$.

Comme on peut écrire $u = (u \circ f) \circ f^{-1}$, on déduit facilement que u est de classe $\mathcal{C}^1$ si et seulement si $u \circ f$ est de classe $\mathcal{C}^1$ par stabilité de la propriété d'être $\mathcal{C}^1$. ■

On remarque qu'être un homéomorphisme de classe $\mathcal{C}^1$ ne suffit pas à être un difféomorphisme. Prenons par exemple l'application $x \mapsto x^3$ sur $\mathbb{R}$ qui est bien un homéomorphisme, mais son application réciproque n'est pas différentiable en zéro. Cependant, il n'y a pas de problèmes pour passer à la régularité supérieure:

Proposition 8.2. — *Si un difféomorphisme $f : \mathrm{U} \to \mathrm{V}$ est de classe $\mathcal{C}^k$, $k \geq 2$, alors f^{-1} est aussi de classe $\mathcal{C}^k$.*

DÉMONSTRATION. — On écrit $g = f^{-1}$ de sorte que $f \circ g = \mathrm{Id}$. Du coup, $\mathrm{D}_{g(y)} f \circ \mathrm{D}_y g = \mathrm{Id}$, impliquant ainsi $\mathrm{D}_y g = (\mathrm{D}_{g(y)} f)^{-1}$. En reprenant les notations de la démonstration précédente, on a donc $\mathrm{D}_y g = \varphi \circ \mathrm{D}f \circ g$.

Montrons par récurrence sur $k \geq 1$ que si f est de classe $\mathcal{C}^k$ alors g aussi. Pour $k = 1$, cela découle de la définition de difféomorphisme. Si c'est vrai au rang k et que f est de classe $\mathcal{C}^{k+1}$, alors $\mathrm{D}f$ est de classe $\mathcal{C}^k$ et g de classe $\mathcal{C}^k$. Or comme φ est de classe $\mathcal{C}^\infty$ (voir l'exemple 6.11), on en déduit que $\mathrm{D}g$ est de classe $\mathcal{C}^k$, donc g est de classe $\mathcal{C}^{k+1}$. ■

On montre comment, dans la proposition 3.8, se passer de l'hypothèse d'être un homéomorphisme si on part d'une application de classe $\mathcal{C}^1$, comme en dimension 1.

Théorème 8.3 (d'inversion locale). — *Soit $f : \mathrm{U} \to \mathrm{F}$ une application de classe $\mathcal{C}^1$. S'il existe $a \in \mathrm{U}$ tel que $\mathrm{D}_a f$ est inversible, alors il existe des voisinages V de a et W de $f(a)$ tels que $f : \mathrm{V} \to \mathrm{W}$ est un difféomorphisme.*

Remarque 8.4. — *Les hypothèses du théorème sont équivalentes à f de classe $\mathcal{C}^1$ et jacobien au point a non nul i.e,* $\det \mathrm{D}_a f \neq 0$.

Une application $f : \mathrm{U} \to \mathrm{F}$ est un *difféomorphisme local* si, pour tout $a \in \mathrm{U}$, il existe un voisinage V de a tel que $f : \mathrm{V} \to f(\mathrm{V})$ est un difféomorphisme. Le théorème d'inversion locale donne une condition nécessaire (et suffisante) pour qu'une application de classe $\mathcal{C}^1$ soit un difféomorphisme local.

Corollaire 8.5. — *Une application $f : \mathrm{U} \to \mathrm{F}$ de classe $\mathcal{C}^1$ est un difféomorphisme local si et seulement si, pour tout $x \in \mathrm{U}$, on a* $\det \mathrm{D}_x f \neq 0$.

Méthode de Newton.— La démonstration s'appuie sur une généralisation de la méthode de Newton. Rappelons le principe de cette méthode. On se donne une fonction numérique $u : \mathrm{I} \to \mathbb{R}$ de classe $\mathcal{C}^1$ dont on cherche un zéro i.e., un point $t \in \mathrm{I}$ tel que $u(t) = 0$. On part d'une condition initiale x_0 et on remplace le graphe de u par la tangente au point x_0 à laquelle on trouve le zéro correpondant que l'on note x_1. On recommence avec la condition initiale x_1 et on construit une suite (x_n) en espérant que celle-ci soit convergente. On montre alors que la limite, si elle existe, est le zéro recherché.

Si x_n est construit, alors l'équation de la tangente est $y - u(x_n) = u'(x_n)(x - x_n)$. Cette droite coupe l'axe des abscisses au point
$$x_{n+1} = x_n - \frac{u(x_n)}{u'(x_n)}\,.$$
Ce point est bien défini si $u'(x_n) \neq 0$, c'est-à-dire si sa tangente n'est pas horizontale!

Pour montrer que la suite est bien convergente, on peut essayer d'appliquer le théorème du point fixe suivant à la fonction $\mathrm{N}_u(x) = x - u(x)/u'(x)$:

Théorème 8.6 (du point fixe). — *Soit (X, d) un espace métrique complet et considérons une transformation $f : \mathrm{X} \to \mathrm{X}$. Si f est contractante i.e., il existe $0 < k < 1$ telle que, pour $x, y \in \mathrm{X}$, on a $d(f(x), f(y)) \leq kd(x, y)$. Alors f admet un unique point fixe p i.e., vérifiant $f(p) = p$. De plus, le point p s'obtient comme suit. Pour tout choix initial $x_0 \in \mathrm{X}$, la suite (x_n) définie par récurrence en posant $x_{n+1} = f(x_n)$ est convergente de limite p.*

Démonstration. — On montre par récurrence que $d(x_n, x_{n+1}) \leq k^n d(x_0, x_1)$. En effet, c'est vrai rang $n = 0$ et si c'est vrai au rang n, alors
$$d(x_{n+1}, x_{n+2}) = d(f(x_n), f(x_{n+1})) \leq kd(x_n, x_{n+1}) \leq k^{n+1} d(x_0, x_1)\,.$$
Par conséquent, pour tout $n, p \geq 0$, on a
$$d(x_n, x_{n+p}) \leq \sum_{j=0}^{p-1} d(x_{n+j}, x_{n+j+1}) \leq \sum_{j=0}^{p-1} k^{n+j} d(x_0, x_1) \leq \frac{k^n}{1-k} d(x_0, x_1)$$
donc (x_n) est une suite de Cauchy. Comme on a supposé X complet, cette suite est convergente; on note sa limite p. Comme f est continue, l'identité $x_{n+1} = f(x_n)$ donne à la limite $p = f(p)$. L'unicité du point fixe découle de la contraction de f car si q était un autre point fixe, on aurait $d(p, q) = d(f(p), f(q)) \leq kd(p, q) < d(p, q)$ qui est impossible. ∎

Retournons à la méthode de Newton. Si $\mathrm{N}_u(p) = p$, cela implique
$$p = p - \frac{u(p)}{u'(p)}$$

soit $u(p) = 0$. Donc un point fixe pour la méthode de Newton fournit bien un zéro de u.

DÉMONSTRATION DU THÉORÈME D'INVERSION LOCALE.— Nous allons mettre en œuvre la stratégie ci-dessus. Pour simplifier les notations, on se ramène au cas où $\mathrm{E} = \mathrm{F}$, $a = f(a) = 0$ et $\mathrm{D}_a f = \mathrm{Id}_\mathrm{E}$. En effet, f est un difféomorphisme local au point a si et seulement si l'application: $u \mapsto (\mathrm{D}_a f)^{-1}[f(a+u) - f(a)]$ est un difféomorphisme local au point 0. C'est-à-dire que moyennant une translation sur a et sur $f(a)$ et un isomorphisme linéaire, on peut se ramener à cette situation simplifiée.

Pour $y \in \mathrm{E}$, on considère la méthode de Newton $\mathrm{N}_y : x \mapsto x - (f(x) - y)$ de sorte qu'un point fixe de N_y nous donne un antécédent de y par f. Nous allons montrer qu'il existe $\varepsilon > 0$ assez petit de sorte que

1. $\mathrm{B}(0, \varepsilon) \subset \mathrm{U}$;

2. si $y \in \mathrm{B}(0, \varepsilon/2)$, alors $\mathrm{N}_y(\mathrm{B}(0, \varepsilon)) \subset \mathrm{B}(0, \varepsilon)$;

3. si $y \in \mathrm{B}(0, \varepsilon/2)$, alors N_y est contractante.

Remarquons que N_y est différentiable sur U et $\mathrm{D}_x\mathrm{N}_y = \mathrm{Id} - \mathrm{D}_x f$. Puisque $\mathrm{D}f$ est continue et $\mathrm{D}_0 f = \mathrm{Id}$, il existe $\varepsilon > 0$ tel que $\mathrm{B}(0, \varepsilon) \subset \mathrm{U}$ et si $||x|| \leq \varepsilon$ alors $||\mathrm{D}_x f - \mathrm{Id}|| \leq 1/2$. D'après l'inégalité des accroissements finis, on a, pour tous x_1, x_2 dans la boule $\mathrm{B}(0, \varepsilon)$,

$$\|\mathrm{N}_y(x_1) - \mathrm{N}_y(x_2)\| \leq \frac{1}{2}\|x_1 - x_2\|. \tag{8.1}$$

Ceci montre en particulier que N_y est contractante. En remarquant que $\mathrm{N}_y(0) = y$, et en choisissant $y \in \mathrm{B}(0, \varepsilon/2)$, alors, pour $x \in \mathrm{B}(0, \varepsilon)$, on a

$$\|\mathrm{N}_y(x)\| \leq \|\mathrm{N}_y(x) - \mathrm{N}_y(0)\| + \|y\| \leq \frac{1}{2}\|x\| + \|y\| < \varepsilon$$

donc on obtient $\mathrm{N}_y(\mathrm{B}(0, \varepsilon)) \subset \mathrm{B}(0, \varepsilon)$ si $\|y\| < \varepsilon/2$. D'après le théorème du point fixe appliqué dans $\overline{\mathrm{B}(0, \varepsilon)}$, N_y admet un unique point fixe $g(y)$ dans $\overline{\mathrm{B}(0, \varepsilon)}$. Par conséquent, l'équation $f(x) = y$ admet une unique solution, $g(y)$, dans $\overline{\mathrm{B}(0, \varepsilon)}$. On pose $\mathrm{V} = \mathrm{B}(0, \varepsilon) \cap f^{-1}(\mathrm{B}(0, \varepsilon/2))$ qui est un ouvert non vide puisque $0 \in \mathrm{V}$; on note aussi $\mathrm{W} = f(\mathrm{V})$.

La restriction $f : \mathrm{V} \to \mathrm{W}$ est donc une bijection de classe $\mathcal{C}^1$ d'inverse g. De plus, $\mathrm{D}_x f$ est inversible sur V car $\mathrm{D}_x f = \mathrm{Id} - (\mathrm{Id} - \mathrm{D}_x f)$ et $\|\mathrm{Id} - \mathrm{D}_x f\| \leq 1/2 < 1$, cf. § 3.1 **5.**. Il reste à établir la continuité de g. D'après (8.1) et l'inégalité triangulaire inverse, on a, pour $y_1, y_2 \in \mathrm{W}$ et $x_1 = g(y_1)$, $x_2 = g(y_2)$,

$$\begin{aligned} \frac{1}{2}\|x_1 - x_2\| &\geq \|\mathrm{N}_0(x_1) - \mathrm{N}_0(x_2)\| \\ &= \|(x_1 - x_2) - (y_1 - y_2)\| \\ &\geq \|x_1 - x_2\| - \|y_1 - y_2\| \end{aligned}$$

ce qui implique $\|x_1 - x_2\| \leq 2\|y_1 - y_2\|$. Autrement dit, g est 2-lipschitzienne, donc continue. Le critère de la proposition 8.1 s'applique pour conclure que $f : \mathrm{V} \to \mathrm{W}$ est un difféomorphisme. ■

Corollaire 8.7 (inversion globale). — *Soient* E, F *des espaces vectoriels normés de dimension finie et soit* U *un ouvert de* E. *Si* $f : \mathrm{U} \to f(\mathrm{U})$ *est de classe* $\mathcal{C}^1$*, injective, et si* $\mathrm{D}_x f \in \mathrm{GL}(\mathrm{E}, \mathrm{F})$ *pour tout* $x \in \mathrm{U}$*, alors* f *est un difféomorphisme de classe* $\mathcal{C}^1$ *de* U *sur* $f(\mathrm{U})$.

DÉMONSTRATION. — D'après le théorème d'inversion locale, f est un difféomorphisme local. Mais comme f est globalement injective, on en déduit que f est un difféomorphisme global sur son image. ■

Corollaire 8.8 (application ouverte). — *Soient* E, F *des espaces vectoriels normés de dimension finie et soit* U *un ouvert de* E. *Si* $f : \mathrm{U} \to \mathrm{F}$ *est de classe* $\mathcal{C}^1$ *et si* $\mathrm{D}_x f \in \mathrm{GL}(\mathrm{E}, \mathrm{F})$ *pour tout* $x \in \mathrm{U}$*, alors* f *est une application ouverte: pour tout ouvert* $\mathrm{V} \subset \mathrm{U}$, $f(\mathrm{V})$ *est ouvert.*

DÉMONSTRATION. — Soit $\mathrm{V} \subset \mathrm{U}$ un ouvert et prenons $a \in \mathrm{V}$. Nous allons montrer que $f(a)$ admet un voisinage contenu dans $f(\mathrm{V})$. Cela découle du théorème d'inversion locale qui implique l'existence d'un voisinage ouvert W de a tel que $f : \mathrm{W} \to f(\mathrm{W})$ est un difféomorphisme; notons $g : f(\mathrm{W}) \to \mathrm{W}$ son inverse qui est donc continue. Du coup, $\mathrm{V} \cap \mathrm{W}$ est un ouvert non vide, et $f(\mathrm{V} \cap \mathrm{W}) = g^{-1}(\mathrm{V} \cap \mathrm{W})$ est ouvert par la continuité de g. ■

On conclut par un exemple important.

Coordonnées polaires dans le plan. — On considère l'application $f :]0,\infty[\times\mathbb{R} \to \mathbb{R}^2$ définie par $f(r,\theta) = (r\cos\theta, r\sin\theta)$. Cette application est de classe $\mathcal{C}^\infty$ et

$$\mathrm{D}f = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

donc $\det \mathrm{D}f = r\cos^2\theta + r\sin^2\theta = r \neq 0$ donc f est un difféomorphisme local de classe $\mathcal{C}^\infty$.

9 Fonctions implicites

On s'intéresse dans ce paragraphe aux solutions d'équations à paramètres. L'objet du théorème des fonctions implicites est de montrer comment de la connaissance d'une solution pour un paramètre fixé obtenir des solutions pour des paramètres voisins.

Théorème 9.1 (des fonctions implicites). — *Soient* $\mathrm{E}, \mathrm{F}, \mathrm{G}$ *des evn de dimension finie,* U *un ouvert de* $\mathrm{E}\times\mathrm{F}$ *et* $f : \mathrm{U} \to \mathrm{G}$ *de classe* $\mathcal{C}^k$, $k \geq 1$. *On suppose qu'en un point* $(x_0, y_0) \in \mathrm{U}$, *la différentielle partielle* $\partial_{\mathrm{F}} f(x_0, y_0)$ *est un isomorphisme.*

Alors il existe des ouverts V *et* W *de* E *et* F *respectivement et une application* $g : \mathrm{V} \to \mathrm{W}$ *de classe* $\mathcal{C}^k$ *tels que*

1. *on a* $\mathrm{V}\times\mathrm{W} \subset \mathrm{U}$*;*
2. *on a* $g(x_0) = y_0$ *et* $\mathrm{D}_{x_0} g = -(\partial_{\mathrm{F}} f(x_0,y_0))^{-1} \circ \partial_{\mathrm{E}} f(x_0, y_0)$*;*
3. *pour tout* $x \in \mathrm{V}$, *l'équation* $f(x,y) = f(x_0,y_0)$ *en* y *admet une unique solution* $g(x)$ *dans* W*; autrement dit, pour tout* $x \in \mathrm{V}$, *on a* $f(x, g(x)) = f(x_0, y_0)$ *et si* $f(x,y) = f(x_0,y_0)$ *avec* $y \in \mathrm{W}$ *alors* $y = g(x)$.

On se donne donc une application $f : \mathrm{E}\times\mathrm{F} \to \mathrm{G}$ et on s'intéresse à l'équation $f(x,y) = z$ où $z \in \mathrm{G}$ est fixé et (x,y) sont des inconnues. On cherche à résoudre cette équation de manière à trouver y en fonction de x. Graphiquement, on s'intéresse à l'ensemble de niveau $\mathrm{E}_z = \{(x,y) \in \mathrm{E}\times\mathrm{F},\ f(x,y) = z\}$ et on veut déterminer si, au voisinage d'un point $(x_0,y_0) \in \mathrm{E}_z$, on peut exprimer E_z comme le graphe d'une fonction $y = g(x)$. Le théorème répond par l'affirmative si le plan tangent de E_z dans $\mathrm{E}\times\mathrm{F}$ n'est pas parallèle à F. Par exemple, si on considère $f(x,y) = x^2 + y^2 - 1 = 0$, alors $\partial_y f = 2y$ qui est non nul si $y \neq 0$. Dans ce cas, on peut exprimer le cercle unité comme le graphe de $x \mapsto \pm\sqrt{1-x^2}$, $x \in]0,1[$.

Démonstration. — Posons $\varphi : \mathrm{U} \to \mathrm{E}\times\mathrm{G}$ défini par $\varphi(x,y) = (x, f(x,y))$. Cette application est de classe $\mathcal{C}^k$ et

$$\mathrm{D}_{(x_0,y_0)}\varphi = \begin{pmatrix} \mathrm{Id}_{\mathrm{E}} & 0 \\ \partial_{\mathrm{E}} f(x_0,y_0) & \partial_{\mathrm{F}} f(x_0,y_0) \end{pmatrix}.$$

Comme $\partial_{\mathrm{F}} f(x_0,y_0)$ est inversible, $\det \mathrm{D}_{(x_0,y_0)}\varphi \neq 0$, donc φ est un difféomorphisme au voisinage du point (x_0,y_0) d'après le théorème d'inversion locale. Il existe donc des ouverts $\mathrm{V} \subset \mathrm{E}$, $\mathrm{W} \subset \mathrm{F}$ et $\mathrm{Y} \subset \mathrm{G}$ tels que $\varphi : \mathrm{V}\times\mathrm{W} \to \mathrm{V}\times\mathrm{Y}$ est un difféomorphisme de classe $\mathcal{C}^k$.

L'égalité $f(x,y) = f(x_0,y_0)$ est équivalente au fait que $\varphi(x,y) = (x, f(x_0,y_0))$, donc $\varphi^{-1}(x, f(x_0,y_0)) = (x,y)$. Notons $\pi_{\mathrm{F}} : \mathrm{E}\times\mathrm{F} \to \mathrm{F}$ la projection canonique et $g : \mathrm{V} \to \mathrm{W}$ définie par $g(x) = \pi_{\mathrm{F}} \circ \varphi^{-1}(x, f(x_0,y_0))$, de sorte que $(x, g(x)) = \varphi^{-1}(x, f(x_0,y_0))$. On vérifie que $g(x_0) = y_0$, $g(\mathrm{V}) \subset \mathrm{W}$, g est de classe $\mathcal{C}^k$ comme composée de fonctions de classe au moins $\mathcal{C}^k$ et $f(x,g(x)) = f(x_0,y_0)$.

Enfin en différentiant $x \mapsto f(x, g(x))$, on obtient $\partial_{\mathrm{E}} f + \partial_{\mathrm{F}} f \circ \mathrm{D}g = 0$, d'où l'expression de $\mathrm{D}g$ au point x_0. ■

Interprétation analytique.— On écrit $f(x,y) = f_x(y)$ pour voir x comme un paramètre. Supposons que l'on connaisse une solution pour le paramètre x_0 de l'équation $f_{x_0}(y) = 0$. On étudie le problème linéarisé induit par $\partial_{\mathrm{F}} f_{x_0}(y)$: si celui-ci peut être résolu au sens que c'est un isomorphisme, alors l'équation perturbée admet une unique solution au voisinage de la solution y_0, qui dépend de manière $\mathcal{C}^1$ de la perturbation.

Exemple: racines simples de polynômes.— On prend $\mathrm{E} = \mathbb{C}_d[\mathrm{X}]$, $\lambda \in \mathbb{C}$ et $\mathrm{P}_0 \in \mathrm{E}$ tels que $\mathrm{P}_0(\lambda) = 0$. On suppose que λ est une racine simple impliquant ainsi $\mathrm{P}_0'(\lambda) \neq 0$. On étudie l'équation $\mathrm{P}(z) = 0$ en posant $f(\mathrm{P}, z) = \mathrm{P}(z)$. On a $\partial_{\mathbb{C}} f(\mathrm{P}_0, \lambda) = \mathrm{P}_0'(\lambda) \neq 0$, donc est inversible. Du coup, il existe un voisinage de U de P_0 et V de λ et une transformation $g : \mathrm{U} \to \mathrm{V}$ de classe $\mathcal{C}^\infty$ tels que, pour tout polynôme Q assez proche de P_0, $g(\mathrm{Q})$ est l'unique racine de Q dans V, cette racine étant aussi simple.

Equivalence entre les théorèmes d'inversion locale et des fonctions implicites.— On a déduit le théorème des fonctions implicites du théorème d'invesion locale, mais on aurait pu procéder à l'inverse. En effet, calculer l'inverse d'une application $\varphi : \mathrm{U} \to \mathrm{F}$ revient à étudier l'équation $\varphi(x) - y = 0$. En posant $f(x, y) = \varphi(x) - y$, on obtient $\partial_{\mathrm{E}} f(x, y) = \mathrm{D}_x\varphi$. Donc si $\mathrm{D}_{x_0}\varphi$ est inversible, alors on trouve des voisinages V de x_0 et W de $\varphi(x_0)$ et une (unique) application $\psi : \mathrm{W} \to \mathrm{V}$ tels que $\varphi(\psi(y)) = y$ pour tout $y \in \mathrm{W}$. Ceci montre que φ est un difféomorphisme local au point x_0.

10 Structure locale des applications différentiables

Dans ce paragraphe, on étudie la forme d'une application de classe $\mathcal{C}^k$ à difféomorphisme près. On étudie plusieurs cas de figure selon la régularité de f au voisinage d'un point a donné.

10.1 Rang constant

Si E et F sont des espaces vectoriels normés, le rang d'une application linéaire $\mathrm{L} \in \mathcal{L}(\mathrm{E}, \mathrm{F})$ est la dimension de $\mathrm{Im}\,\mathrm{L}$. Pour une application $\varphi : \mathrm{U} \to \mathrm{F}$ différentiable au point $a \in \mathrm{U}$, on définit

$$\mathrm{rg}_a\,\varphi \stackrel{\text{def.}}{=} \mathrm{rg}\,\mathrm{D}_a\varphi\,.$$

On rappelle le lemme suivant d'algèbre linéaire.

Lemme 10.1. — *Soient $p, q \geq 1$ et $r \geq 0$. Si une matrice $\mathrm{A} \in \mathcal{M}_{q,p}(\mathbb{R})$ est de rang r alors il existe $\mathrm{P} \in \mathrm{GL}_p(\mathbb{R})$ et $\mathrm{Q} \in \mathrm{GL}_q(\mathbb{R})$ telles que*

$$\mathrm{QAP}^{-1} = \begin{pmatrix} \mathrm{I}_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Démonstration. — On considère l'application linéaire $u_{\mathrm{A}} : \mathbb{R}^p \to \mathbb{R}^q$ définie par $u_{\mathrm{A}}(\mathrm{X}) = \mathrm{AX}$. Soit $(\mathrm{e}_1, \ldots \mathrm{e}_n)$ une base de $\mathrm{Ker}\, u_{\mathrm{A}}$ que l'on complète en une base $(\mathrm{e}_1, \ldots, \mathrm{e}_n, \varepsilon_1, \ldots, \varepsilon_m)$ de $\mathbb{R}^p$.

Montrons que $(u_{\mathrm{A}}(\varepsilon_1), \ldots, u_{\mathrm{A}}(\varepsilon_m))$ est une base de $\mathrm{Im}\, u_{\mathrm{A}}$. Tout d'abord, cette famille est génératrice par définition: si $y \in \mathrm{Im}\, u_{\mathrm{A}}$, alors il existe $x \in \mathbb{R}^p$ tel que $u_{\mathrm{A}}(x) = y$. Or, on peut trouver des scalaires $x_1, \ldots, x_n$ et $y_1, \ldots, y_m$ tels que

$$x = \sum_{k=1}^{n} x_k \mathrm{e}_k + \sum_{j=1}^{m} y_j \varepsilon_j\,.$$

Du coup, comme $\sum x_k \mathrm{e}_k \in \mathrm{Ker}\, u_{\mathrm{A}}$, on obtient

$$y = u_{\mathrm{A}}(x) = \sum_{j=1}^{m} y_j u_{\mathrm{A}}(\varepsilon_j)\,.$$

Il reste à vérifier que cette famille est aussi libre. Pour cela, on choisit $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ tels que $\sum \lambda_j u_{\mathrm{A}}(\varepsilon_j) = 0$. Cela s'écrit aussi $u_{\mathrm{A}}(\sum \lambda_j \varepsilon_j) = 0$ ce qui implique que $\sum \lambda_j \varepsilon_j \in \mathrm{Ker}\, u_{\mathrm{A}}$. Comme $(\mathrm{e}_1, \ldots, \mathrm{e}_n)$ est une base de $\mathrm{Ker}\, u_{\mathrm{A}}$ et $(\mathrm{e}_1, \ldots, \mathrm{e}_n, \varepsilon_1, \ldots, \varepsilon_m)$ est libre, on en déduit que $\sum \lambda_j \varepsilon_j = 0$ donc $\lambda_1 = \ldots = \lambda_m = 0$ et cette famille est bien une base de $\mathrm{Im}\, u_{\mathrm{A}}$.

Comme $\mathrm{rg}\,\mathrm{A} = r$, on en déduit que $m = r$ et $p = n + r$. Enfin, on note $f_j = u_{\mathrm{A}}(\varepsilon_j)$ et on complète $(f_1, \ldots, f_r)$ en une base $\mathcal{F} = (f_1, \ldots, f_q)$ de $\mathbb{R}^q$. On réordonne la base de $\mathbb{R}^p$ pour l'écrire sous la forme $\mathcal{E} = (\varepsilon_1, \ldots, \varepsilon_r, \mathrm{e}_1, \ldots, \mathrm{e}_n)$. La matrice de u_{A} dans ces bases est donc

$$\mathrm{Mat}(u_{\mathrm{A}}, \mathcal{E}, \mathcal{F}) = \begin{pmatrix} \mathrm{I}_r & 0 \\ 0 & 0 \end{pmatrix}.$$

En notant P la matrice de passage de la base canonique de $\mathbb{R}^p$ vers $\mathcal{E}$ et Q la matrice de passage de la base canonique de $\mathbb{R}^q$ vers $\mathcal{F}$, on conclut la démonstration du lemme. ∎

On montre une version « à géométrie variable » de ce résultat.

Théorème 10.2 (du rang constant). — *Soit $f : \mathrm{U} \to \mathrm{F}$ une application de classe $\mathcal{C}^k$, $k \geq 1$, définie sur* U *et supposons qu'il existe $r \geq 1$ tel que $\mathrm{rg}_x f = r$ pour tout x dans un voisinage d'un point $a \in \mathrm{U}$. Alors il existe un voisinage $\mathrm{V} \subset \mathrm{U}$ de a, un voisinage $\mathrm{W} \subset \mathrm{F}$ de $f(a)$, des difféomorphismes $\varphi : \mathrm{V} \to \mathrm{V}' (\subset \mathbb{R}^p)$ et $\psi : \mathrm{W} \to \mathrm{W}' (\subset \mathbb{R}^q)$ tels que $\psi \circ f \circ \varphi^{-1} = \mathrm{D}_a f : \mathrm{V}' \to \mathrm{W}'$. Dans des bases appropriées, on obtient*

$$(x_1, \ldots, x_p) \mapsto (x_1, \ldots, x_r, \underbrace{0, \ldots, 0}_{q-r})$$

Ce théorème affirme qu'une application de classe au moins $\mathcal{C}^1$ de rang constant est localement linéaire à changement de coordonnées près par des difféomorphismes. Autrement dit, une application de classe au moins $\mathcal{C}^1$ de rang constant n'est pas plus compliquée qu'une application linéaire.

Donnons les grandes étapes de la démonstration. Dans un premier temps, on construit φ. Pour cela, on part du fait que la fonction f se comporte en première approximation comme l'application affine $x \mapsto f(a) + \mathrm{D}_a f$, donc envoie U proche du sous-espace $f(a) + \mathrm{Im}\,\mathrm{D}_a f$. Cela nous conduit à considérer une projection $\pi : \mathrm{F} \to \mathrm{Im}\,\mathrm{D}_a f$ et d'étudier dans un premier temps l'application $\pi \circ f : \mathrm{U} \to \mathrm{Im}\,\mathrm{D}_a f$. On se place dans des repères adaptés de E et F qui tiennent compte de $\mathrm{D}_a f$ c'est-à-dire pour que $\mathrm{D}_a f$ soit la matrice donnée par le lemme 10.1. On décompose donc E et F sous la forme $\mathrm{E} = \mathrm{E}' \oplus \mathrm{Ker}\,\mathrm{D}_a f$ et $\mathrm{F} = \mathrm{Im}\,\mathrm{D}_a f \oplus \mathrm{Ker}\,\pi$. Dans cette situation, on montre l'existence d'un difféomorphisme local φ tel que $\pi \circ f \circ \varphi^{-1} = \mathrm{D}_a f$ est linéaire.

L'étape suivante consiste à montrer que $f \circ \varphi^{-1}$ est de la forme $(x, y) \mapsto (x, g(x))$ au voisinage de 0; on utilise ici que le rang est localement constant. De là, on construit ψ au voisinage de $f(a)$ en posant $\psi(x, z) = (x, z - g(x))$.

Démonstration. — Quitte à remplacer f par $x \mapsto (f(a+x) - f(a))$, on peut supposer que $a = f(a) = 0$. Nous donnons deux présentations en parallèle de la preuve: l'une intrinsèque et l'autre en coordonnées.

On choisit une base $\mathcal{B}_\mathrm{E} = (\mathrm{e}_1, \ldots, \mathrm{e}_p)$ de E telle que $(\mathrm{e}_{r+1}, \ldots, \mathrm{e}_p)$ soit une base de $\mathrm{Ker}\,\mathrm{D}_0 f$; notons E' l'espace engendré par $(\mathrm{e}_1, \ldots, \mathrm{e}_r)$. Du coup, si on pose $f_j = \mathrm{D}_0 f(\mathrm{e}_j)$ pour $j = 1, \ldots, r$, alors $(f_1, \ldots, f_r)$ est une famille libre que l'on peut compléter en une base de $\mathcal{B}_\mathrm{F} = (f_1, \ldots, f_q)$ de F. Cela nous permet d'identifier E avec $\mathbb{R}^p$ et F avec $\mathbb{R}^q$. On travaille maintenant dans ces repères.

Dans cette base, on a l'écriture en matrice par blocs

$$\mathrm{D}_0 f = \begin{pmatrix} \mathrm{I}_r & 0 \\ 0 & 0 \end{pmatrix}$$

Du coup, la restriction $\mathrm{D}_0 f|_{\mathrm{E}'} : \mathrm{E}' \to \mathrm{Im}\,\mathrm{D}_0 f$ est un isomorphisme dont on notera l'inverse

$$(\mathrm{D}_0 f|_{\mathrm{E}'})^{-1} : \mathrm{Im}\,\mathrm{D}_0 f \to \mathrm{E}'.$$

Notez que $\mathrm{D}_0 f \circ (\mathrm{D}_0 f|_{\mathrm{E}'})^{-1} = \mathrm{Id}$ et $(\mathrm{D}_0 f|_{\mathrm{E}'})^{-1} \circ \mathrm{D}_0 f = \pi_1$ où $\pi_1 : \mathrm{E} \to \mathrm{E}$ est la projection de E sur E' parallèlement à $\mathrm{Ker}\,\mathrm{D}_0 f$.

Un développement limité de f montre que f envoie un voisinage de 0 dans $\mathbb{R}^p$ très proche de $\mathrm{Im}\,\mathrm{D}_0 f = \mathbb{R}^r \times \{0\}_{q-r} \subset \mathrm{F}$. On considère la projection $\pi : \mathrm{F} \to \mathrm{F}$ sur $\mathrm{Im}\,\mathrm{D}_0 f$ parallèlement à $\mathrm{F}' = \mathrm{vec}(f_{r+1}, \ldots, f_q) = \{0\} \times \mathbb{R}^{q-r}$ définie par $\pi(x, z) = x$, pour $(x, z) \in \mathbb{R}^r \times \mathbb{R}^{q-r}$.

On définit $\varphi : \mathrm{E} \to \mathrm{E}$ par $\varphi = ((\mathrm{D}_0 f|_{\mathrm{E}'})^{-1} \circ \pi \circ f) + \pi_2$, où $\pi_2 : \mathrm{E} \to \mathrm{E}$ désigne la projection de E sur $\mathrm{Ker}\,\mathrm{D}_0 f$ parallèlement à E'. L'application φ s'écrit, dans les repères choisis, $\varphi : \mathbb{R}^r \times \mathbb{R}^{p-r} \to \mathbb{R}^r \times \mathbb{R}^{p-r}$ en posant $\varphi(x, y) = (((\mathrm{D}_0 f|_{\mathrm{E}'})^{-1} \circ \pi \circ f)(x, y), y)$. C'est une application de classe $\mathcal{C}^k$. On a

$$\mathrm{D}_0 \varphi = ((\mathrm{D}_0 f|_{\mathrm{E}'})^{-1} \circ \pi \circ \mathrm{D}_0 f) + \pi_2 = \pi_1 + \pi_2 = \mathrm{Id}$$

ou, plus prosaïquement, on calcule la matrice jacobienne. On obtient la matrice par blocs

$$\mathrm{D}_0 \varphi = \begin{pmatrix} \pi \circ \partial_1 f(0) & \pi \circ \partial_2 f(0) \\ 0 & \mathrm{I}_{p-r} \end{pmatrix} = \begin{pmatrix} \mathrm{I}_r & 0 \\ 0 & \mathrm{I}_{p-r} \end{pmatrix}$$

car $\pi = \mathrm{Id}$ sur $\mathbb{R}^r \times \{0\}_{q-r}$ et $\{0\}_r \times \mathbb{R}^{q-r} = \operatorname{Ker} \pi$.

Par conséquent $\mathrm{D}_0\varphi$ est inversible et le théorème d'inversion locale affirme que φ est un difféomorphisme $\mathcal{C}^k$ au voisinage de 0; notons V, V′ ces voisinages de sorte que $\varphi : \mathrm{V} \to \mathrm{V}'$; on choisit V′ convexe. On a

$$\mathrm{D}_0 f \circ \varphi = \pi \circ f + \mathrm{D}_0 f \circ \pi_2 = \pi \circ f$$

car $\operatorname{Im} \pi_2 \subset \operatorname{Ker} \mathrm{D}_0 f$. Donc on a $\pi \circ f \circ \varphi^{-1} = \mathrm{D}_0 f$. Autrement dit, on a $(\pi \circ f \circ \varphi^{-1})(x, y) = x$ par construction.

Or $\mathrm{D}_0 f = \pi \circ \mathrm{D}_0 f$ donc $\pi[(f \circ \varphi^{-1}) - \mathrm{D}_0 f] = 0$ car $\pi = \mathrm{Id}$ sur $\operatorname{Im} \mathrm{D}_0 f$; du coup, on a $((f \circ \varphi^{-1}) - \mathrm{D}_0 f)(\mathrm{V}') \subset \operatorname{Ker} \pi$. Par conséquent, il existe une application $g : \mathrm{V}' \to \operatorname{Ker} \pi = \mathrm{F}'$ de classe $\mathcal{C}^k$ telle que $(f \circ \varphi^{-1}) = \mathrm{D}_0 f + g$, soit $g : \mathbb{R}^p \to \mathbb{R}^{q-r}$ telle que $(f \circ \varphi^{-1})(x, y) = (x, g(x, y))$.

Nous allons montrer qu'en fait g ne dépend que de $x \in \mathrm{E}' = \mathbb{R}^r$. C'est là que l'hypothèse de rang constant va jouer. En effet, pour $x \in \mathrm{V}'$,

$$\begin{aligned} r &= \operatorname{rg}_x(f \circ \varphi^{-1}) \geq \dim \mathrm{D}_x(f \circ \varphi^{-1})(\mathrm{E}') \\ &\geq \dim \pi \circ \mathrm{D}_x(f \circ \varphi^{-1})(\mathrm{E}') = \dim \mathrm{D}_x(\pi \circ f \circ \varphi^{-1}) \\ &\geq \dim \mathrm{D}_0 f(\mathrm{E}') = r \end{aligned}$$

donc $\dim \mathrm{D}_x(f \circ \varphi^{-1})(\mathrm{E}') = r$ et on déduit d'une part que $\mathrm{D}_x(f \circ \varphi^{-1})(\operatorname{Ker} \mathrm{D}_0 f) \subset \mathrm{D}_x(f \circ \varphi^{-1})(\mathrm{E}')$ car le rang vaut r et d'autre part que la restriction $\pi : \mathrm{D}_x(f \circ \varphi^{-1})(\mathrm{E}') \to \operatorname{Im} \mathrm{D}_0 f$ est un isomorphisme car elle est surjective et les dimensions coïncident. Comme $\pi(\mathrm{D}_x(f \circ \varphi^{-1})(\operatorname{Ker} \mathrm{D}_0 f)) = 0$, il vient $\mathrm{D}_x(f \circ \varphi^{-1})(\operatorname{Ker} \mathrm{D}_0 f) = 0$. Alternativement, on considère la matrice jacobienne de $(f \circ \varphi^{-1})$ en un point $(x, y) \in \mathrm{V}'$:

$$\mathrm{D}_{(x,y)} g = \begin{pmatrix} \mathrm{I}_r & 0 \\ \partial_1 g & \partial_2 g \end{pmatrix}$$

Or, pour tout $(x, y) \in \mathrm{V}'$, $\operatorname{rg}_{(x,y)}(f \circ \varphi^{-1}) = r$, donc $\partial_2 g \equiv 0$ d'après le lemme 10.3 ci-dessous.

Autrement dit, comme V′ est convexe, pour $z, w \in \mathrm{V}'$ tel que $z - w \in \operatorname{Ker} \mathrm{D}_0 f$, on obtient $g(z) = g(w)$. Ceci implique que g ne dépend que de x, ce que l'on peut encore écrire sous la forme $g = g \circ \pi_1$. Il vient $(f \circ \varphi^{-1})(x, y) = (x, g(x))$ pour $(x, y) \in (\mathbb{R}^r \times \mathbb{R}^{q-r}) \cap \mathrm{V}'$.

Enfin, on note $\psi = \mathrm{Id} - g \circ (\mathrm{D}_0 f|_{\mathrm{E}'})^{-1} \circ \pi$ c'est-à-dire $\psi : (x, z) \mapsto (x, z - g(x))$, bien définie au voisinage de $0 \in \mathrm{F}$. Cette application est injective, de classe $\mathcal{C}^k$ et

$$\mathrm{D}\psi = \mathrm{Id} - \mathrm{D}_{(\mathrm{D}_0 f|_{\mathrm{E}'})^{-1} \circ \pi} g \circ (\mathrm{D}_0 f|_{\mathrm{E}'})^{-1} \circ \pi\,.$$

Montrons que c'est un isomorphisme: si $\mathrm{D}\psi(h) = 0$ alors on déduit d'abord que $h \in \operatorname{Ker} \pi$ car g y prend ses valeurs donc

$$h = \mathrm{D}_{(\mathrm{D}_0 f|_{\mathrm{E}'})^{-1} \circ \pi} g \circ (\mathrm{D}_0 f|_{\mathrm{E}'})^{-1} \circ \pi(h) = 0\,.$$

Ceci montre que $\mathrm{D}\psi$ est inversible. En coordonnées, on obtient

$$\mathrm{D}\psi = \begin{pmatrix} \mathrm{I}_r & 0 \\ -\partial_1 g & \mathrm{I}_{q-r} \end{pmatrix}$$

sur son ensemble de définition, donc on retrouve que $\mathrm{D}\psi$ est inversible en tout point où ψ est défini. Le théorème d'inversion globale implique que ψ est un difféomorphisme sur son image. On obtient ainsi

$$\psi \circ f \circ \varphi^{-1} = [\mathrm{D}_0 f + g] - g \circ (\mathrm{D}_0 f|_{\mathrm{E}'})^{-1} \circ \pi \circ [\mathrm{D}_0 f + g] = \mathrm{D}_0 f + g - g \circ \pi_1 = \mathrm{D}_0 f$$

car g prend ses valeurs dans $\mathrm{F}' = \operatorname{Ker} \pi$. On peut aussi conduire le calcul en coordonnées:

$$(\psi \circ f \circ \varphi^{-1})(x, y) = \psi(x, g(x)) = (x, 0)\,.$$

Cela termine la démonstration. ■

Lemme 10.3. — *On considère une matrice par blocs* $\mathrm{A} = \begin{pmatrix} \mathrm{I}_r & 0 \\ \mathrm{B} & \mathrm{C} \end{pmatrix}$. *Si* $\operatorname{rg} \mathrm{A} = r$ *alors* $\mathrm{C} = 0$.

Démonstration. — On note $v_1, \ldots, v_n$ les vecteurs colonnes de A. On constate que les r premières colonnes $(v_1, \ldots, v_r)$ forment une famille libre car elles sont échelonnées. Comme on a supposé que A est de rang r, tout autre vecteur colonne v_j, $r < j \leq n$, est combinaison linéaire des r premières. On a donc $v_j = \sum_{k=1}^r \lambda_k v_k$. En regardant la $k$ième ligne de A, on constate que l'on doit avoir $\lambda_k = 0$. Donc $v_j = 0$ pour tout $r < j \leq n$. Cela signifie $\mathrm{C} = 0$. ■

10.2 Points critiques

On s'intéresse ici au cas où f admet un point critique isolé, de différentielle seconde non dégénérée. On note $\mathcal{S}_n(\mathbb{R})$ l'espace des matrices symétriques. On rappelle le théorème de Gauss concernant les formes quadratiques non dégénérées définies sur $\mathbb{R}$.

Théorème 10.4. — *Si q est une forme quadratique non dégénérée définie sur un espace vectoriel de dimension finie, il existe un unique couple d'entiers naturels (s,t), appelé la signature de q tel que $s+t = \dim \mathrm{E} \overset{def.}{=} n$, et tel qu'il existe une base dans laquelle*

$$q(x_1,\ldots,x_n) = \sum_{1\leq j\leq s} x_j^2 - \sum_{s+1\leq j\leq n} x_j^2.$$

On montre une version « à géométrie variable » du résultat précédent.

Lemme de Morse. — *Soit $f : \mathbb{R}^n \to \mathbb{R}$ une application de classe $\mathcal{C}^k$, $k \geq 2$ telle que $f(0) = \mathrm{D}_0 f = 0$ et telle que $\mathrm{D}_0^2 f$ soit inversible. Alors il existe un voisinage V de l'origine et des applications de classe $\mathcal{C}^{k-2}$ $y_1,\ldots,y_r$ et $z_1,\ldots,z_s$ définies sur V telles que $r+s=n$ et, pour $x \in \mathrm{V}$, on ait*

$$f(x) = \sum_{1\leq j\leq r} y_j^2 - \sum_{1\leq j\leq s} z_j^2\,.$$

Démonstration. — On considère le développement limité avec reste intégrale de f au voisinage de l'origine. On a

$$f(x) = \int_0^1 (1-t)\mathrm{D}^2_{tx} f(x,x)dt = {}^t\mathrm{X}\cdot \int_0^1 (1-t)\mathrm{D}^2_{tx} f dt \cdot \mathrm{X}\,.$$

On note

$$\mathrm{A}(x) = \left(\int_0^1 (1-t)\frac{\partial^2 f}{\partial x_i \partial x_j}(tx)dt\right)_{i,j} \qquad \text{et} \qquad \mathrm{A}_0 = \mathrm{A}(0)\,.$$

On utilise alors le lemme suivant.

Lemme 10.5. — *Si $\mathrm{A}_0 \in \mathcal{S}_n(\mathbb{R}) \cap \mathrm{GL}_n(\mathbb{R})$, il existe un voisinage U de A_0 dans $\mathcal{S}_n(\mathbb{R})$ et une application infiniement différentiable $\psi : \mathrm{U} \to \mathcal{M}_n(\mathbb{R})$ tels que $\psi(\mathrm{A}_0) = \mathrm{I}$ et, pour tout $\mathrm{A} \in \mathrm{U}$,*

$${}^t\psi(\mathrm{A})\cdot \mathrm{A}_0 \cdot \psi(\mathrm{A}) = \mathrm{A}\,.$$

Du coup, si x est assez proche de l'origine, alors $f(x) = {}^tx{}^t\psi(\mathrm{A}(x))\mathrm{A}_0\psi(\mathrm{A}(x))x$. On pose $\psi_1(x) = \psi(\mathrm{A}(x))x$, et on obtient

$$f(x) = {}^t\psi_1(x)\mathrm{A}_0\psi_1(x)\,.$$

Or il existe une base P dans laquelle A_0 soit diagonale avec r valeurs sur la diagonale égales à 1 et s égales à -1. On note J cette matrice et on pose $\psi_2 = \mathrm{P}\cdot\psi_1$. Il vient $f(x) ={}^t \psi_2(x)\mathrm{J}\psi_2(x)$. Si on appelle $y_1,\ldots,y_r$ et $z_1,\ldots,z_s$ les coordonnées de ψ_2, on obtient la forme recherchée.

Ceci établit le lemme de Morse modulo le Lemme 10.5. ■

Démonstration du Lemme 10.5. — On considère l'application $h : \mathcal{M}_n(\mathbb{R}) \to \mathcal{S}_n(\mathbb{R})$ définie par $h(\mathrm{M}) = {}^t\mathrm{M}\mathrm{A}_0\mathrm{M}$. On calcule la différentielle à l'identité de h.

$$h(\mathrm{I}+\mathrm{M}) = {}^t(\mathrm{I}+\mathrm{M})\mathrm{A}_0(\mathrm{I}+\mathrm{M}) = \mathrm{A}_0 + ({}^t\mathrm{M}\mathrm{A}_0 + \mathrm{A}_0\mathrm{M}) +{}^t\mathrm{M}\mathrm{A}_0\mathrm{M} = \mathrm{A}_0 + ({}^t\mathrm{M}\mathrm{A}_0 + \mathrm{A}_0\mathrm{M}) + \mathrm{O}(\|\mathrm{M}\|^2)\,.$$

Donc $\mathrm{D}_\mathrm{I}h(\mathrm{M}) = {}^t\mathrm{M}\mathrm{A}_0 + \mathrm{A}_0\mathrm{M}$. Cette application n'est pas inversible puisque $\dim \mathcal{S}_n(\mathbb{R}) < \dim \mathcal{M}_n(\mathbb{R})$. En revanche, le noyau consiste en les matrices M telles que $\mathrm{A}_0\mathrm{M}$ soit antisymétrique. On considère l'espace E des matrices M telles que $\mathrm{A}_0\mathrm{M}$ soit symétrique. Cet espace est supplémentaire à $\mathrm{Ker}\,\mathrm{D}_\mathrm{I}h$, et la restriction de $\mathrm{D}_\mathrm{I}h$ à E devient maintenant inversible (injective car A_0 est inversible et espaces source et but de même dimension).

Le théorème d'inversion locale appliqué à $h|_\mathrm{E}$ montre qu'il existe des voisinages U de A_0 et V de I et un difféomorphisme infiniment différentiable $\psi : \mathrm{U} \to \mathrm{V}$ qui inverse $h|_\mathrm{E}$. ■

On conclut par un dernier énoncé important. On considère une fonction f définie sur un ouvert U de $\mathbb{R}^p$, à valeurs dans $\mathbb{R}^q$, et de classe $\mathcal{C}^k$.

On appelle *points critiques* les points en lesquels l'application différentielle de f est non surjective, et *valeurs critiques* les images des points critiques. Les valeurs non critiques sont dites *régulières* (qu'elles soient des valeurs effectivement prises par f ou non). On rassemble deux résultats en un, dûs à Sard et Brown respectivement.

Théorème 10.6 (Sard, Brown). — *Soit $f : \mathrm{U} \to \mathbb{R}^q$ de classe $\mathcal{C}^k$, où U est un ouvert de $\mathbb{R}^p$.*

1. *Si $k > \max(0, p-q)$, alors l'ensemble des valeurs critiques a une mesure de Lebesgue nulle.*

2. *L'ensemble des valeurs régulières est dense dans $\mathbb{R}^q$.*

En revanche, l'ensemble des points critiques peut être très important, par exemple si $p < q$, tous les points sont critiques, mais l'ensemble image de f sera quand même de mesure nulle.

Chapitre V. — SOUS-VARIETES DE $\mathbb{R}^n$

On introduit dans ce chapitre des sous-ensembles particuliers d'espaces vectoriels sur lesquels on peut développer une théorie intéressante du calcul différentiel. Ces ensembles, les sous-variétés, généralisent les graphes de fonctions, comme nous le verrons ci-dessous.

11 Définitions

Soit E un espace vectoriel normé de dimension finie. Un sous-ensemble $\mathrm{X} \subset \mathrm{E}$ est une *sous-variété* de classe $\mathcal{C}^k$, $k \geq 1$, si, pour tout $x \in \mathrm{X}$, il existe un voisinage ouvert U de x dans E, un voisinage ouvert V de 0 dans E, un sous-espace vectoriel fermé $\mathrm{F} \subset \mathrm{E}$ et un difféomorphisme $\varphi : \mathrm{U} \to \mathrm{V}$ tels que

$$\varphi^{-1}(\mathrm{F} \cap \mathrm{V}) = \mathrm{X} \cap \mathrm{U}\,.$$

Cela signifie que X est localement un sous-espace vectoriel quitte à faire un changement de coordonnées approprié. On dit que φ *redresse* X en un sous-espace vectoriel. Cette notion généralise celle de graphes de fonctions, cf. l'exemple 1.

Exemple 1.— On considère dans $\mathbb{R}^n$, muni de la norme euclidienne, le graphe d'une fonction $f : \mathbb{R}^p \to \mathbb{R}^q$, $p + q = n$, de classe $\mathcal{C}^k$:

$$\mathrm{X} = \{(x, y) \in \mathbb{R}^p \times \mathbb{R}^q,\ y = f(x)\}\,.$$

L'application $\psi : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^n$ définie par $(x, y) \mapsto (x, f(x) + y)$ est un difféomorphisme sur son image puisque ψ est injective, de classe $\mathcal{C}^k$ et

$$\mathrm{D}_{(x,y)}\psi = \begin{pmatrix} \mathrm{Id}_p & 0 \\ \mathrm{D}_x f & \mathrm{Id}_q \end{pmatrix}$$

De plus, on a $\psi(\mathbb{R}^p \times \{0\}) = \mathrm{X}$, donc l'inverse $\varphi \overset{\text{def.}}{=} \psi^{-1}$ convient (à translation près).

Exemple 2.— La sphère unité pour la norme euclidienne

$$\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n,\ \|x\| = 1\}\,.$$

Pour tout $a \in \mathbb{S}^{n-1}$, $a_n \neq 0$, on peut considérer l'application

$$\Phi : (x_1, \ldots, x_{n-1}, x_n) \mapsto \left(x_1, \ldots, x_{n-1}, x_n \pm \sqrt{1 - \sum_{j=1}^{n-1} x_j^2}\right)$$

où on choisit le même signe que a_n et qui est bien définie sur

$$\mathrm{U} = \left\{(x_1, \ldots, x_{n-1}, x_n) \in \mathbb{R}^n,\ \sum_{j=1}^{n-1} x_j^2 < 1\right\}\,.$$

On note $f(x) = \sqrt{1 - \sum_{j=1}^{n-1} x_j^2}$ et on calcule

$$\mathrm{D}_a\Phi = \begin{pmatrix} \mathrm{I}_{n-1} & 0 \\ \pm\mathrm{D}_a f & 1 \end{pmatrix}$$

On vérifie ainsi que $\Phi : \mathrm{U} \to \mathrm{V} \overset{\text{def.}}{=} \Phi(\mathrm{U})$ est un difféomorphisme local au voisinage de a et $\Phi(\mathbb{R}^{n-1} \times \{0\} \cap \mathrm{U}) = \mathbb{S}^{n-1} \cap \mathrm{V}$. Si $a_n = 0$, alors on peut trouver une autre composante a_j non nulle et procéder de la même manière.

L'espace tangent en un point d'une sous-variété est un objet fondamental qui nous permettra de développer le calcul différentiel sur les sous-variétés. Il généralise les tangentes des graphes de fonctions.

Définition 11.1 (espace tangent). — *Soit* X *une sous-variété d'un evn* E *et soit* $x \in \mathrm{X}$. *Le plan tangent* $\mathrm{T}_x\mathrm{X}$ *de* X *au point* x *est l'ensemble des vecteurs* v *de* E *tels que*

$$\lim_{t \to 0} \frac{1}{t} d(x + tv, \mathrm{X}) = 0\,.$$

On rappelle que si $\mathrm{A} \subset \mathrm{E}$ et $x \in \mathrm{A}$ alors

$$d(x.\mathrm{A}) \overset{\text{def.}}{=} \inf_{a \in \mathrm{A}} d(x, a) = \inf_{a \in \mathrm{A}} \|a - x\| \,.$$

Lemme 11.2. — *Soit* X *une sous-variété. Pour tout* $x \in \mathrm{X}$*, il existe* $r > 0$ *tel que, pour tout* $y \in \mathrm{B}(a, r)$*, il existe* $z \in \mathrm{X}$ *tel que* $d(y, z) = d(y, \mathrm{X})$*.*

Démonstration. — Prenons $x \in \mathrm{X}$. Il existe un voisinage U de x et V de 0, un sous-espace vectoriel F ainsi qu'un difféomorphisme $\varphi : \mathrm{U} \to \mathrm{V}$ tels que $\varphi^{-1}(\mathrm{F} \cap \mathrm{V}) = \mathrm{X} \cap \mathrm{U}$. Comme U est ouvert, il existe $r > 0$ tel que $\mathrm{B}(x, 3r) \subset \mathrm{U}$. Prenons maintenant $y \in \mathrm{B}(x, r)$. On a $d(y, \mathrm{X}) \leq d(x, y) < r$. Prenons une suite $(z_n)_n$ de X telle que $\lim \|y - z_n\| = d(y, \mathrm{X})$. Remarquons que si $z \in \mathrm{X} \setminus \overline{\mathrm{B}(x, 2r)}$, alors $\|y - z\| \geq \|z - x\| - \|x - y\| \geq 2r - r \geq r > d(y, \mathrm{X})$, donc (z_n) appartient à $\mathrm{B}(x, 2r)$. Or $\varphi(\overline{\mathrm{B}(x, 2r)} \cap \mathrm{X})$ est un compact de F, donc le théorème de Bolzano-Weierstrass nous permet d'extraire une sous-suite $(z_{n_k})_k$ de (z_n) telle que $\varphi(z_{n_k})$ tend vers un point $p \in \mathrm{F} \cap \mathrm{V}$ car $\varphi(\overline{\mathrm{B}(x, 2r)}) \subset \mathrm{V}$. Par continuité de φ^{-1}, la sous-suite $(z_{n_k})_k$ tend vers le point $z = \varphi^{-1}(p) \in \mathrm{X}$ et $\|z - y\| = \lim \|z_{n_k} - y\| = d(y, \mathrm{X})$. ■

On a les descriptions suivantes de l'espace tangent en un point.

Proposition 11.3. — *Soient* X *une sous-variété et* $x \in \mathrm{X}$*.*

1. *Un vecteur* $v \in \mathrm{E}$ *est tangent au point* x *si, et seulement si, il existe une application différentiable* $c : \mathrm{I} \to \mathrm{E}$ *définie sur un intervalle contenant* 0 *telle que* $c(\mathrm{I}) \subset \mathrm{X}$*,* $c(0) = x$ *et* $c'(0) = v$*.*

2. *Si* X *est donnée au voisinage de* x *par un difféomorphisme* $\varphi : \mathrm{V} \to \mathrm{U}$*, avec* $\mathrm{X} \cap \mathrm{V} = \varphi^{-1}(\mathrm{F})$*, alors*

$$\mathrm{T}_x\mathrm{X} = (\mathrm{D}_x\varphi)^{-1}(\mathrm{F}) \,.$$

Démonstration. — On considère d'abord X donnée par un difféomorphisme $\varphi : \mathrm{V} \to \mathrm{U}$. Soit $w \in \mathrm{F}$ et posons $c(t) = \varphi^{-1}(tw)$ qui définit un chemin de X. Un développement limité à l'ordre 1 montre

$$c(t) = \varphi^{-1}(tw) = x + t(\mathrm{D}_x\varphi)^{-1}(w) + o(t)$$

donc

$$d(x + t(\mathrm{D}_x\varphi)^{-1}(w), \mathrm{X}) \leq \|x + t(\mathrm{D}_x\varphi)^{-1}(w) - c(t)\| = o(t)$$

et on en déduit que $c'(0) = (\mathrm{D}_x\varphi)^{-1}(w) \in \mathrm{T}_x\mathrm{X}$. Donc $(\mathrm{D}_x\varphi)^{-1}(\mathrm{F}) \subset \mathrm{T}_x\mathrm{X}$.

Faisons l'observation suivante. Si $c : \mathrm{I} \to \mathrm{X}$ vérifie $c(0) = x$ et $c'(0)$ existe, alors $\mathrm{D}_x\varphi(c'(0)) \in \mathrm{F}$. En effet, comme $c(t) \in \mathrm{X}$ pour tout t, on a $\varphi(c(t)) \in \mathrm{F}$ et

$$\mathrm{D}_x\varphi(c'(0)) = (\varphi \circ c)'(0) = \lim_{t \to 0} \frac{1}{t}(\varphi \circ c)(t) \in \mathrm{F}$$

car F est fermé (les sous-espaces vectoriels d'un espace vectoriel normé de dimension finie sont fermés). Du coup, $c'(0) \in (\mathrm{D}_x\varphi)^{-1}(\mathrm{F})$ est tangent à X au point x.

Soit v un vecteur tangent à X au point x. D'après le lemme 11.2, il existe $c(t) \in \mathrm{X}$ tel que $d(x + tv, \mathrm{X}) = \|x + tv - c(t)\| = o(t)$, donc $c(t) = x + tv + o(t)$ ce qui signifie que $c(0) = x$ et $c'(0)$ existe et vaut v. L'observation précédente implique $v \in (\mathrm{D}_x\varphi)^{-1}(\mathrm{F})$. Ceci montre que $(\mathrm{D}_x\varphi)^{-1}(\mathrm{F}) = \mathrm{T}_x\mathrm{X}$, donc établit le point **2.**, mais aussi **1.** car on a vu que tout vecteur $v \in (\mathrm{D}_x\varphi)^{-1}(\mathrm{F})$ s'exprimait comme le vecteur dérivé d'un chemin contenu dans X. ■

12 Différentes présentations des sous-variétés

On donne deux nouvelles définitions équivalentes de sous-variétés. Pour cela, on introduit deux classes de fonctions: les immersions et les submersions.

Définition 12.1 (immersion). — *Une application* $j : \mathrm{U}\,(\subset \mathrm{E}) \to \mathrm{F}$ *de classe* $\mathcal{C}^k$*,* $k \geq 1$*, est une immersion en* $x \in \mathrm{U}$ *si* $\mathrm{D}_x j \in \mathcal{L}(\mathrm{E}, \mathrm{F})$ *est injective.*

Remarque 12.2. — *En dimension finie, si* $\mathrm{L} \in \mathcal{L}(\mathrm{E}, \mathrm{F})$ *est injective, alors* L *admet un inverse à gauche, c'est-à-dire une application linéaire* $\mathrm{L}' \in \mathcal{L}(\mathrm{F}, \mathrm{E})$ *tel que* $\mathrm{L}' \circ \mathrm{L} = \mathrm{Id}_\mathrm{E}$*, cf. le lemme 10.1.*

Etant donnée une application différentiable $f : \mathbb{R}^p \to \mathbb{R}^q$, l'application $j : x\,(\in \mathbb{R}^p) \mapsto (x, f(x)) \in \mathbb{R}^{p+q}$ est une immersion car

$$\mathrm{D}j = \begin{pmatrix} \mathrm{Id}_p \\ \mathrm{D}f \end{pmatrix}$$

qui est bien de rang p.

Proposition 12.3. — *Si* X *est une sous-variété de* E *de classe* $\mathcal{C}^k$*,* $k \geq 1$*, et si* $x \in \mathrm{X}$*, alors il existe un voisinage* V *de* x *une immersion* $j : \mathrm{U} \to \mathrm{E}$ *définie au voisinage de l'origine telles que*

$$\mathrm{X} \cap \mathrm{V} = j(\mathrm{U})\,.$$

Démonstration. — On considère un voisinage W de $0 \in \mathbb{R}^n$, un sous-espace vectoriel fermé $\mathrm{F} \subset \mathbb{R}^n$ et un difféomorphisme $\varphi : \mathrm{V} \to \mathrm{W}$ tels que $\varphi^{-1}(\mathrm{F} \cap \mathrm{W}) = \mathrm{X} \cap \mathrm{V}$. On pose $p = \dim \mathrm{F}$ et $q = n - p$. Notons $\psi = \varphi^{-1} : \mathrm{W} \to \mathrm{U}$. Quitte à faire un changement de base de $\mathbb{R}^n$, on peut supposer que $\mathrm{F} = \mathbb{R}^p \times \{0\}$.

On note $j : \mathrm{F} \to \mathrm{E}$ définie par $j = \psi|_\mathrm{F}$, qui est de classe $\mathcal{C}^k$ car j est une restriction de ψ. Comme ψ est un difféomorphisme local, $\mathrm{D}\psi$ est injective, donc $\mathrm{D}j$ aussi. Par construction, $j(\mathrm{F} \cap \mathrm{W}) \subset \mathrm{X}$. ∎

On s'intéresse maintenant à la réciproque de cette proposition qui nous fournira une autre manière de définir une sous-variété.

Proposition 12.4. — *Soit* $j : \mathrm{U} \to \mathrm{E}\,(\simeq \mathbb{R}^n)$ *de classe* $\mathcal{C}^k$*, où* U *est un ouvert d'un evn* $\mathrm{G} \simeq \mathbb{R}^p$*. Si* j *est une immersion en* $x \in \mathrm{G}$*, alors il existe un voisinage* V *de* x *tel que* $\mathrm{X} = j(\mathrm{V})$ *soit une sous-variété de* F *et* $\mathrm{T}_{j(x)}\mathrm{X} = \mathrm{Im}\,\mathrm{D}_x j$*.*

Démonstration. — Comme j est une immersion en x on a $p \leq n$. Le rang de j est constant au voisinage de x puisque l'on peut trouver un mineur de $\mathrm{D}_x j$ de taille p non nul. Par continuité de $\mathrm{D}j$ et du déterminant, ce mineur reste non nul au voisinage de x. Comme $\mathrm{rg}\, j \leq p$, le rang ne peut augmenter.

Par conséquent, le théorème du rang constant implique l'existence de difféomorphismes locaux φ et ψ, tels que $\varphi(x) = 0$, $\psi(j(x)) = 0$ et $\psi \circ j \circ \varphi^{-1} = \mathrm{D}_x j$.

Notons $\mathrm{F} = \mathrm{Im}\,\mathrm{D}_x j$. L'application $\psi : \mathrm{W} \to \mathrm{W}'$ est un difféomorphisme sur son image et

$$\psi^{-1}(\mathrm{F} \cap \mathrm{W}') = (\psi^{-1} \circ \mathrm{D}_x j)(\mathrm{V}') = (j \circ \varphi^{-1})(\mathrm{V}') = j(\mathrm{V})\,.$$

Par ailleurs

$$\mathrm{T}_{j(x)}\mathrm{X} = (\mathrm{D}_{j(x)}\psi)^{-1}(\mathrm{F}) = \mathrm{D}_0\psi^{-1} \circ \mathrm{D}_x j(\mathrm{G}) = \mathrm{D}_0(\psi^{-1} \circ \mathrm{D}_x j)(\mathrm{G}) = \mathrm{D}_0(j \circ \varphi^{-1})(\mathrm{G}) = \mathrm{Im}\,\mathrm{D}_x j\,.$$

∎

Cette démonstration montre en particulier qu'une immersion est localement injective. Dans ce contexte, on dit que j fournit un *paramétrage local* de X.

Remarque 12.5. — *Si* $j : \mathrm{U}\,(\subset \mathbb{R}^p) \to \mathrm{X}$ *est un paramétrage local et* φ *un redressement de* X *en* $j(0)$*, alors* $\varphi \circ j : \mathrm{U} \to \mathrm{F}$ *est différentiable, et* $\dim \mathrm{F} = p$*. Cela découle du fait que* $j(\mathrm{U}) = \varphi^{-1}(\mathrm{F} \cap \mathrm{W})$*. Donc le théorème d'inversion locale montre que* $\varphi \circ j$ *est un difféomorphisme au voisinage de* 0 *et* $j = \varphi^{-1} \circ (\varphi \circ j)$ *est un homéomorphisme local.*

Du coup, si on a deux paramétrages $j_1, j_2 : \mathrm{U} \to \mathrm{X}$ *alors il existe un difféomorphisme* $\psi : \mathrm{U} \to \mathrm{U}$ *tel que* $j_2 \circ \psi = j_1$*. En effet, on peut écrire*

$$j_1 = j_2 \circ (\varphi \circ j_2)^{-1} \circ (\varphi \circ j_1)\,.$$

Définition 12.6 (submersion). — *Une application* $g : \mathrm{U}\,(\subset \mathrm{E}) \to \mathrm{F}$ *de classe* $\mathcal{C}^k$*,* $k \geq 1$*, est une submersion en* $x \in \mathrm{U}$ *si* $\mathrm{D}_x g \in \mathcal{L}(\mathrm{E}, \mathrm{F})$ *est surjective. On dit que* g *est une submersion au-dessus de* $y \in \mathrm{F}$ *si* $g^{-1}(\{y\})$ *est non vide et si* g *est une submersion en chaque point* $x \in g^{-1}(\{y\})$*.*

Remarque 12.7. —

1. *En dimension finie, si* $\mathrm{L} \in \mathcal{L}(\mathrm{E}, \mathrm{F})$ *est surjective, alors* L *admet un inverse à droite, c'est-à-dire une application linéaire* $\mathrm{L}' \in \mathcal{L}(\mathrm{F}, \mathrm{E})$ *tel que* $\mathrm{L} \circ \mathrm{L}' = \mathrm{Id}_\mathrm{F}$*, cf. le lemme 10.1.*

2. *Si* $g : \mathrm{U} \to \mathbb{R}$*, il suffit de vérifier que* $\mathrm{D}_x g \neq 0$ *pour s'assurer que* g *est une submersion en* x*.*

Proposition 12.8. — *Si* X *est une sous-variété de* E *de classe* $\mathcal{C}^k$*,* $k \geq 1$*, et si* $x \in \mathrm{X}$*, alors il existe un voisinage* V *de* x*, une submersion* $g : \mathrm{V} \to \mathbb{R}^q$ *au-dessus de l'origine telle que*

$$\mathrm{X} \cap \mathrm{V} = g^{-1}(\{0\}) \cap \mathrm{V}\,.$$

L'application $g : \mathbb{R}^n \to \mathbb{R}$ définie par $g(x) = \sum x_j^2 - 1$ est une submersion au-dessus de 0 puisque

$$\mathrm{D}_x g = \begin{pmatrix} 2x_1 & \dots & 2x_n \end{pmatrix}$$

et si $\mathrm{D}_x g = 0$ alors $x = 0$, mais $g(0) \neq 0$. Cela remontre que $\mathbb{S}^{n-1}$ est une sous-variété.

Démonstration. — On considère un voisinage W de $0 \in \mathbb{R}^n$, un sous-espace vectoriel fermé $\mathrm{F} \subset \mathbb{R}^n$ et un difféomorphisme $\varphi : \mathrm{V} \to \mathrm{W}$ tels que $\varphi^{-1}(\mathrm{F} \cap \mathrm{W}) = \mathrm{X} \cap \mathrm{V}$. On note $p = \dim \mathrm{F}$ et $q = n - p$.

Quitte à faire un changement de base de $\mathbb{R}^n$, on peut supposer que $\mathrm{F} = \mathbb{R}^p \times \{0\}$. Notons $\pi : \mathbb{R}^n \to \mathbb{R}^p \times \mathbb{R}^q$ la projection sur $\{0\} \times \mathbb{R}^q$, que l'on identifie à $\mathbb{R}^q$, parallèlement à $\mathbb{R}^p \times \{0\}$. Posons $g = \pi \circ \varphi : \mathbb{R}^n \to \mathbb{R}^q$ qui est de classe $\mathcal{C}^k$. On vérifie que $g(z) = 0$ si et seulement si $\varphi(z) \in \operatorname{Ker} \pi$, soit $\varphi(z) \in \mathrm{F}$. Par conséquent, on a $\mathrm{X} \cap \mathrm{V} = g^{-1}(\{0\}) \cap \mathrm{V}$ pour V assez petit. De plus, $\mathrm{D}g = \pi \circ \mathrm{D}\varphi$. Comme $\mathrm{D}\varphi$ est un isomorphisme et $\operatorname{rg} \pi = q$, on en déduit que $\mathrm{D}g$ est surjective sur V, donc g est bien une submersion. ■

On s'intéresse maintenant à la réciproque de cette proposition.

Proposition 12.9. — *Soit* $g : \mathrm{U} \to \mathrm{G}\,(\simeq \mathbb{R}^q)$ *de classe* $\mathcal{C}^k$*, où* U *est un ouvert d'un evn* $\mathrm{E} \simeq \mathbb{R}^n$*. Si* g *est une submersion au-dessus de* y*, alors* $\mathrm{X} = g^{-1}(\{y\})$ *est une sous-variété et* $\mathrm{T}_a\mathrm{X} = \operatorname{Ker} \mathrm{D}_a g$ *pour tout* $a \in \mathrm{X}$*.*

Démonstration. — Comme g est une submersion au-dessus de y, on a $n \geq q$. Soit $a \in \mathrm{U}$ tel que $g(a) = y$. On remarque que $\operatorname{rg} g$ est constant au voisinage de a, car ce rang est maximal, donc on peut trouver un mineur de $\mathrm{D}_a g$ non nul de taille $q = \dim \mathrm{G}$. Ce mineur restera non nul au voisinage de a par continuité de $\mathrm{D}g$ et du déterminant.

Par conséquent, le théorème du rang constant implique l'existence de difféomorphismes locaux φ et ψ, tels que $\varphi(a) = 0$, $\psi(y) = 0$ et $\psi \circ g \circ \varphi^{-1} = \mathrm{D}_a g$.

Notons $\mathrm{F} = \operatorname{Ker} \mathrm{D}_a g$. L'application $\varphi : \mathrm{V} \to \mathrm{V}'\,(\subset \mathrm{G})$ est un difféomorphisme et $\varphi^{-1}(\mathrm{F} \cap \mathrm{V}') = g^{-1}(\{y\}) \cap \mathrm{V}$ puisque si $x \in \mathrm{V} \cap \mathrm{X}$, alors $\mathrm{D}_a g(\varphi(x)) = (\psi \circ g \circ \varphi^{-1})(\varphi(x)) = \psi(y) = 0$, donc $\varphi(x) \in \mathrm{F}$, et l'on a

$$\psi \circ g \circ \varphi^{-1}(\mathrm{F} \cap \mathrm{V}') = \mathrm{D}_a g(\mathrm{F} \cap \mathrm{V}') = 0 = \psi(y)\,.$$

De plus, on a $g = \psi^{-1} \circ \mathrm{D}_a g \circ \varphi$ donc

$$\operatorname{Ker} \mathrm{D}_a g = \operatorname{Ker} \mathrm{D}_a(\psi^{-1} \circ \mathrm{D}_a g \circ \varphi) = \operatorname{Ker}(\mathrm{D}_a g \circ \mathrm{D}_a \varphi) = (\mathrm{D}_a \varphi)^{-1}(\operatorname{Ker} \mathrm{D}_a g) = (\mathrm{D}_a \varphi)^{-1}(\mathrm{F})\,.$$

■

Notion de dimension. — On remarque que si X est connexe, alors tous les sous-espaces F qui redressent X ont même dimension. Du coup, on définit la *dimension de* X comme la dimension de ces sous-espaces F. Si X est présentée par une immersion $j : \mathrm{U}\,(\subset \mathbb{R}^p) \to \mathrm{X}$, alors $\dim \mathrm{X} = p$ et si X est présentée par une submersion $g : \mathbb{R}^n \to \mathbb{R}^q$ au-dessus d'un point, alors $\dim \mathrm{X} = n - q$. On dit aussi que X est de *codimension* q.

On illustre ces notions en dimension 3. Posons donc $\mathrm{E} = \mathbb{R}^3$.

1. Une sous-variété de dimension 0 est un ensemble localement fini de points.

2. Une sous-variété de dimension 1 est une courbe. Elle peut être donnée par une immersion $j : \mathrm{I}\,(\subset \mathbb{R}) \to \mathrm{E}$, c'est-à-dire telle que $j'(t) \neq 0$; on peut aussi considérer une submersion $g : \mathbb{R}^3 \to \mathbb{R}^2$, ce qui signifie que les différentielles $\mathrm{D}_x g_1$ et $\mathrm{D}_x g_2$ des fonctions coordonnées de g sont linéairement indépendantes.

3. Une sous-variété de dimension 2 est une surface. Elle peut être donnée par une immersion $j : \mathrm{U}\,(\subset \mathbb{R}^2) \to \mathrm{E}$, c'est-à-dire telle que les dérivées partielles $\partial_1 j$ et $\partial_2 j$ sont linéairement indépendantes; on peut aussi considérer une submersion $g : \mathbb{R}^3 \to \mathbb{R}$, ce qui signifie que $\mathrm{D}g \neq 0$.

4. Une sous-variété de dimension 3 est un ouvert de $\mathbb{R}^3$.

Proposition 12.10. — *Si* X_1 *et* X_2 *sont des surfaces de* $\mathbb{R}^3$ *et si, pour tout* $x \in \mathrm{X}_1 \cap \mathrm{X}_2$*, les plans tangents* $\mathrm{T}_x\mathrm{X}_1$ *et* $\mathrm{T}_x\mathrm{X}_2$ *sont différents, alors* $\mathrm{X}_1 \cap \mathrm{X}_2$ *est une courbe et* $\mathrm{T}_x(\mathrm{X}_1 \cap \mathrm{X}_2) = \mathrm{T}_x\mathrm{X}_1 \cap \mathrm{T}_x\mathrm{X}_2$.

Démonstration. — Soit $x \in \mathrm{X}_1 \cap \mathrm{X}_2$ et supposons que X_1 soit définie au voisinage de x par une submersion $g_1 : \mathrm{V}_1 \to \mathbb{R}$ au-dessus de 0 et X_2 par une submersion $g_2 : \mathrm{V}_2 \to \mathbb{R}$ au-dessus de 0. On note $\mathrm{V} = \mathrm{V}_1 \cap \mathrm{V}_2$ et on considère $g : \mathrm{V} \to \mathbb{R}^2$ définie par $g(y) = (g_1(y), g_2(y))$. On a $g(y) = 0$ si et seulement si $g_1(y) = g_2(y) = 0$ donc si et seulement si $y \in (\mathrm{X}_1 \cap \mathrm{X}_2) \cap \mathrm{V}$. Par ailleurs, $\mathrm{D}g_1$ et $\mathrm{D}g_2$ sont des formes linéairement indépendantes puisque les plans tangents sont distincts: si $\lambda \mathrm{D}_y g_1 + \mu \mathrm{D}_y g_2 = 0$, alors, en prenant un élément $v \in \mathrm{T}_x\mathrm{X}_2 \setminus \mathrm{T}_x\mathrm{X}_1$, on obtient $\lambda \mathrm{D}_y g_1(v) = 0$ donc $\lambda = 0$; de même $\mu = 0$. Donc g est une submersion et $\mathrm{X}_1 \cap \mathrm{X}_2$ définit une courbe.

De plus $\operatorname{Ker} \mathrm{D}g = \operatorname{Ker} \mathrm{D}g_1 \cap \operatorname{Ker} \mathrm{D}g_2$. ■

13 Applications différentiables

Soit $\mathrm{X} \subset \mathrm{E}$ une sous-variété de classe $\mathcal{C}^k$, on souhaite définir la différentiabilité de $f : \mathrm{X} \to \mathrm{F}$ et sa différentielle en un point $x \in \mathrm{X}$ qui sera une application linéaire $\mathrm{D}_x f \in \mathcal{L}(\mathrm{T}_x\mathrm{X}, \mathrm{F})$. Nous proposons différentes suggestions selon la manière dont est présentée X, et nous montrerons qu'elles sont toutes équivalentes.

Soient $x \in \mathrm{X}$, V un voisinage ouvert de x dans E et $f : (\mathrm{V} \cap \mathrm{X}) \to \mathbb{R}^n$ une application.

1. On dira que f est différentiable en x si f admet un prolongement $\tilde{f} : \mathrm{V} \to \mathbb{R}^n$ différentiable en x. On notera $\mathrm{D}_x f = \mathrm{D}_x \tilde{f}|_{\mathrm{T}_x\mathrm{X}}$. On dira que f est de classe $\mathcal{C}^m$ sur $\mathrm{X} \cap \mathrm{V}$, $1 \leq m \leq k$, si on peut choisir $\tilde{f}$ de classe $\mathcal{C}^m$ sur V.

2. Si X est donnée par un difféomorphisme local $\varphi : \mathrm{V} \to \mathrm{U}$ tel que $\varphi(x) = 0$ et $\mathrm{X} \cap \mathrm{V} = \varphi^{-1}(\mathrm{F})$, alors on dira que f est différentiable en x si $f \circ \varphi^{-1}|_{\mathrm{F}}$ est différentiable en x. On notera $\mathrm{D}_x f = \mathrm{D}_0(f \circ \varphi^{-1}) \circ \mathrm{D}_x\varphi$ sur $\mathrm{T}_x\mathrm{X}$. On dira que f est de classe $\mathcal{C}^m$ sur V, $1 \leq m \leq k$, si $f \circ \varphi^{-1}|_{\mathrm{F}}$ est de classe $\mathcal{C}^k$ sur $\mathrm{U} \cap \mathrm{F}$.

3. Si X est donnée par un paramétrage local $j : \mathrm{U}\,(\subset \mathbb{R}^p) \to \mathrm{X} \cap \mathrm{V}$, c'est-à-dire une immersion telle que $j(0) = x$, alors on dira que f est différentiable en x si $f \circ j$ est différentiable en 0. On notera $\mathrm{D}_x f = \mathrm{D}_0(f \circ j) \circ (\mathrm{D}_0 j)^{-1}$. On dira que f est de classe $\mathcal{C}^m$ sur V, $1 \leq m \leq k$, si $f \circ j$ est de classe $\mathcal{C}^m$ sur U.

La dernière définition a besoin d'un peu d'explication: la différentielle $\mathrm{D}_0 j \in \mathcal{L}(\mathbb{R}^p, \mathrm{E})$ est injective et son image est $\mathrm{T}_x\mathrm{X}$. Par conséquent, $\mathrm{D}_0 j \in \mathrm{GL}(\mathbb{R}^p, \mathrm{T}_x\mathrm{X})$, donc $(\mathrm{D}_0 j)^{-1}$ désigne l'inverse $(\mathrm{D}_0 j)^{-1} : \mathrm{T}_x\mathrm{X} \to \mathbb{R}^p$.

Théorème 13.1. — *Toutes ces définitions sont équivalentes, et la différentielle ne dépend pas des différents choix.*

Démonstration. — Soit $\tilde{f}$ un prolongement différentiable sur V et prenons $v \in \mathrm{T}_x\mathrm{X}$. Il existe un chemin $c : \mathrm{I} \to \mathrm{X}$ tel que $c(0) = x$ et $c'(0) = v$. L'application $\tilde{f} \circ c$ est donc différentiable et

$$\mathrm{D}_x \tilde{f}(v) = \mathrm{D}_0(\tilde{f} \circ c) = \lim_{t \to 0} \frac{1}{t}(f(c(t)) - f(x))$$

car $c(\mathrm{I}) \subset \mathrm{X}$ et donc $\tilde{f} \circ c = f \circ c$. Donc la différentiabilité de f en x ne dépend pas du prolongement $\tilde{f}$, ni $\mathrm{D}_x f$.

Prenons maintenant un difféomorphisme $\varphi : \mathrm{V} \to \mathrm{U}$. Si $\tilde{f}$ est différentiable, alors $\tilde{f} \circ \varphi^{-1}$ est aussi différentiable en 0 et

$$\mathrm{D}_0(\tilde{f} \circ \varphi^{-1}) = \mathrm{D}_x \tilde{f} \circ (\mathrm{D}_x \varphi)^{-1}\,.$$

Si on se restreint à F, alors $\tilde{f} \circ \varphi^{-1} = f \circ \varphi^{-1}$ et $(\mathrm{D}_x\varphi)^{-1}(\mathrm{F}) = \mathrm{T}_x\mathrm{X}$, donc

$$\mathrm{D}_0(\tilde{f} \circ \varphi^{-1}) \circ \mathrm{D}_x\varphi|_{\mathrm{T}_x\mathrm{X}} = \mathrm{D}_x\tilde{f}|_{\mathrm{T}_x\mathrm{X}} = \mathrm{D}_x f\,.$$

Réciproquement, supposons $f \circ \varphi^{-1}$ différentiable. Notons $\pi : \mathbb{R}^n \to \mathbb{R}^n$ la projection orthogonale sur F et posons $\tilde{f} = (f \circ \varphi^{-1}) \circ \pi \circ \varphi$. Il s'agit donc de la composée d'applications différentiables. Du coup

$$\mathrm{D}_x\tilde{f} = \mathrm{D}_0(f \circ \varphi^{-1}) \circ \pi \circ \mathrm{D}_x\varphi$$

donc

$$\mathrm{D}_x\tilde{f}|_{\mathrm{T}_x\mathrm{X}} = \mathrm{D}_0(f \circ \varphi^{-1}) \circ \pi \circ \mathrm{D}_x\varphi|_{\mathrm{T}_x\mathrm{X}} = \mathrm{D}_0(f \circ \varphi^{-1}) \circ \mathrm{D}_x\varphi|_{\mathrm{T}_x\mathrm{X}}$$

car $\mathrm{D}_x\varphi(\mathrm{T}_x\mathrm{X}) = \mathrm{F}$ et $\pi|_{\mathrm{F}} = \mathrm{Id}$.

Enfin, si j est une immersion, alors $\tilde{f} \circ j$ est différentiable et $\mathrm{D}_0(\tilde{f} \circ j) = \mathrm{D}_x\tilde{f} \circ \mathrm{D}_0 j$. Comme $\mathrm{D}_0 j : \mathbb{R}^p \to \mathrm{T}_x\mathrm{X}$ est un isomorphisme, on obtient

$$\mathrm{D}_x\tilde{f}|_{\mathrm{T}_x\mathrm{X}} = \mathrm{D}_0(\tilde{f} \circ j) \circ (\mathrm{D}_0 j)^{-1}|_{\mathrm{T}_x\mathrm{X}}\,.$$

Si $(f \circ j)$ est différentiable alors on peut construire une extension $\tilde{f}$ comme pour les redressements puisque j est la restriction de l'inverse d'un redressement sur F. ■

Il existe plusieurs difficultés pour donner du sens aux différentielles d'ordre supérieur, notamment à cause des propriétés de composition. La différentielle de f en x est une application $\mathrm{D}_x f \in \mathcal{L}(\mathrm{T}_x\mathrm{X}, \mathbb{R}^n)$, donc on devrait avoir $\mathrm{D}^2_x f \in \mathcal{L}^2(\mathrm{T}_x\mathrm{X}, \mathbb{R}^n)$. Pour donner du sens à cette application bilinéaire, on devrait vérifier son invariance par difféomorphisme de sorte qu'elle ne dépende pas du paramétrage local, c'est-à-dire on souhaiterait

$$\mathrm{D}^2_x(f \circ \varphi)(v,w) = \mathrm{D}^2_{\varphi(x)} f(\mathrm{D}_x\varphi(v), \mathrm{D}_x\varphi(w))\,.$$

Or $\mathrm{D}_x(f \circ \varphi)(w) = \mathrm{D}_{\varphi(x)} f \circ \mathrm{D}_x\varphi(w)$ et si on différentie à nouveau, on obtient

$$\mathrm{D}^2_x(f \circ \varphi)(v,w) = \mathrm{D}^2_{\varphi(x)} f(\mathrm{D}_x\varphi(v), \mathrm{D}_x\varphi(w)) + \mathrm{D}_{\varphi(x)} f \circ \mathrm{D}^2_x\varphi(v,w)$$

donc la propriété d'invariance n'est pas préservée. Elle le serait si $\mathrm{D}_{\varphi(x)} f = 0$. Nous verrons au paragraphe suivant que cela suffit pour les points critiques.

14 Extrema liés

On se propose d'étudier les extrema de fonctions différentiables $f : \mathrm{X} \to \mathbb{R}$. Par un paramétrage local, on peut se ramener au cas standard, cf. la proposition 3.12 et le corollaire 5.11. Nous déclinons cette approche lorsque X est donnée par une submersion.

14.1 Calcul à l'ordre 1

On adapte la proposition 3.12 aux sous-variétés.

Théorème 14.1 (extrema liés). — *Soient $\mathrm{X} \subset \mathbb{R}^n$ une sous-variété de dimension p donnée par une submersion $\mathrm{X} = (g = 0)$ et $f : \mathrm{X} \to \mathbb{R}$ une application différentiable; notons $q = n - p$. Si un point x est un extremum local alors, pour toute extension $\tilde{f}$ de f, il existe des scalaires $\lambda_1, \ldots \lambda_q$ tels que*

$$\mathrm{D}_x\tilde{f} = \sum_{j=1}^{q} \lambda_j \mathrm{D}_x g_j$$

où les g_j désignent les fonctions composantes de g. Une fois l'extension $\tilde{f}$ fixée, les scalaires $\lambda_1, \ldots, \lambda_q$ sont uniques et sont appelés les multiplicateurs de Lagrange.

Démonstration. — Décrivons X au voisinage de x par une immersion $j : \mathrm{U} \to \mathrm{X}$. Dire que x est un extremum local de f revient à dire que 0 est un extremum local de $f \circ j$, donc un point critique: $\mathrm{D}_0(f \circ j) = 0$. Soit $\tilde{f}$ une extension différentiable de f au voisinage de x. On a

$$\mathrm{D}_0(f \circ j) = \mathrm{D}_x\tilde{f} \circ \mathrm{D}_0 j = 0$$

donc

$$\mathrm{T}_x\mathrm{X} = \mathrm{D}_0 j(\mathbb{R}^p) \subset \operatorname{Ker} \mathrm{D}_x\tilde{f}\,.$$

Or $\mathrm{T}_x\mathrm{X} = \operatorname{Ker} \mathrm{D}_x g = \cap \operatorname{Ker} \mathrm{D}_x g_j$, donc $\cap \operatorname{Ker} \mathrm{D}_x g_j \subset \operatorname{Ker} \mathrm{D}_x\tilde{f}$. Le lemme 14.2 ci-dessous d'algèbre linéaire permet de conclure. ■

Lemme 14.2. — *Soient* E *un espace vectoriel de dimension finie et* $f_1, \dots, f_p \in \mathrm{E}^*$ *linéairement indépendantes. Alors* $(f_1, \dots, f_p)$ *est une base du sous-espace* $\mathrm{P} = \{\mathrm{L} \in \mathrm{E}^*,\ \cap \operatorname{Ker} f_j \subset \operatorname{Ker} \mathrm{L}\}$.

Pour ce lemme, on rappelle que si $\mathcal{B} = (\mathrm{e}_1, \dots, \mathrm{e}_n)$ est une base d'un espace vectoriel réel E, alors son dual E^* admet comme base la famille de formes linéaires $\mathcal{B}^* = (\mathrm{e}_1^*, \dots, \mathrm{e}_n^*)$ où chaque $\mathrm{e}_j^* : \mathrm{E} \to \mathbb{R}$ est définie par

$$\mathrm{e}_j^*(\mathrm{e}_i) = \begin{cases} 0 & \text{si } i \neq j \\ 1 & \text{si } i = j\,. \end{cases}$$

Si $\mathrm{F} \subset \mathrm{E}$ est un sous-espace de E, son orthogonal $\mathrm{F}^\circ \subset \mathrm{E}^*$ est le sous-espace

$$\mathrm{F}^\circ = \{\mathrm{L} \in \mathrm{E}^* \text{ telle que } \mathrm{L}(x) = 0 \text{ pour tout } x \in \mathrm{F}\}\,.$$

Nous aurons recours au lemme suivant:

Lemme 14.3. — *Avec ces notations, on a* $\dim \mathrm{F} + \dim \mathrm{F}^\circ = \dim \mathrm{E}$.

Démonstration. — Prenons une base $(\mathrm{e}_1, \dots, \mathrm{e}_p)$ de F que l'on complète en une base $\mathcal{B} = (\mathrm{e}_1, \dots, \mathrm{e}_n)$ de E. Notons $\mathcal{B}^*$ sa base duale. Le lemme sera établi si on montre que F° est engendré par $(\mathrm{e}_{p+1}^*, \dots, \mathrm{e}_n^*)$.

Prenons $\mathrm{L} \in \mathrm{E}^*$. On écrit $\mathrm{L} = \sum a_j \mathrm{e}_j^*$. Si $\mathrm{L} \in \mathrm{F}^\circ$, alors on a $\mathrm{L}(\mathrm{e}_i) = 0$ pour tout $i \in \{1, \dots, p\}$. Par conséquent, $\mathrm{L}(\mathrm{e}_i) = a_i = 0$ et donc $\mathrm{L} = \sum_{j=p+1}^n a_j \mathrm{e}_j^*$. Ceci montre $\mathrm{F}^\circ \subset \operatorname{vec}(\mathrm{e}_{p+1}^*, \dots, \mathrm{e}_n^*)$.

Réciproquement, si $\mathrm{L} \in \operatorname{vec}(\mathrm{e}_{p+1}^*, \dots, \mathrm{e}_n^*)$, alors $\mathrm{L}(\mathrm{e}_i) = 0$ pour tout $i \in \{1, \dots, p\}$, donc $\mathrm{L}(x) = 0$ pour tout $x \in \mathrm{F}$. Ceci montre l'inclusion inverse et conclut la démonstration du lemme. ■

Démonstration du lemme 14.2. — On définit $f : \mathrm{E} \to \mathbb{R}^p$ par $f(x) = (f_1(x), \dots, f_p(x))$. On a $p = \operatorname{rg}(f_1, \dots, f_p) = \dim \operatorname{Im} f$ et $\operatorname{Ker} f = \cap \operatorname{Ker} f_j$, donc $\dim \cap \operatorname{Ker} f_j = \dim \mathrm{E} - p$. Par définition P est l'orthogonal de $\operatorname{Ker} f$ donc $\dim \mathrm{P} = \dim \mathrm{E} - \dim \operatorname{Ker} f = p$. Comme les (f_j) sont indépendants et dans P, c'est une base. ■

Corollaire 14.4. — *Soient* X *une sous-variété,* $f : \mathrm{X} \to \mathbb{R}$ *une application de classe* $\mathcal{C}^k$, $k \geq 1$, *et* $x \in \mathrm{X}$ *un point. Les propositions suivantes sont équivalentes.*

1. *Si* $\tilde{f}$ *est un prolongement différentiable de* f, $\mathrm{T}_x\mathrm{X} \subset \operatorname{Ker} \mathrm{D}_x\tilde{f}$.

2. *Si* X *est donnée par un redressement* φ *sur un sous-espace* F, $\mathrm{F} \subset \operatorname{Ker} \mathrm{D}_0(f \circ \varphi^{-1})$.

3. *Si* X *est donnée par une submersion* g *au-dessus de* 0, $\mathrm{D}_x\tilde{f} = \sum \lambda_j \mathrm{D}_x g_j|_{\mathrm{T}_x\mathrm{X}}$.

4. *Si* X *est donnée par une immersion,* $\mathrm{D}_0(f \circ j) = 0$.

Définition 14.5 (point critique). — *Soient* X *une sous-variété et* $f : \mathrm{X} \to \mathbb{R}$ *une application de classe* $\mathcal{C}^k$, $k \geq 1$. *Un point* $x \in \mathrm{X}$ *est un point critique de* f *si les conditions du corollaire 14.4 sont vérifiées.*

Démonstration. — L'essence du théorème des extrema liés est d'établir les implications **4.** implique **1.** implique **3.**. Si **3.** est vérifié, alors

$$\mathrm{T}_x\mathrm{X} = \operatorname{Ker} \mathrm{D}_x g = \cap \operatorname{Ker} \mathrm{D}_x g_j \subset \operatorname{Ker} \mathrm{D}_x\tilde{f}$$

impliquant ainsi **1.** Comme $\mathrm{T}_x\mathrm{X} = \operatorname{Im} \mathrm{D}_0 j$, **1.** implique aussi $\mathrm{D}_0(f \circ j) = \mathrm{D}_x\tilde{f} \circ \mathrm{D}_0 j = 0$, c'est-à-dire **4.**.

Enfin, l'équivalence avec **2.** se déduit facilement du fait que j est la restriction de l'inverse d'un redressement sur F. ■

Enfin, on a le corollaire suivant:

Corollaire 14.6. — *Soient* X *une variété de classe* $\mathcal{C}^1$ *et* $f : \mathrm{X} \to \mathbb{R}$ *de classe* $\mathcal{C}^1$. *Si* $x \in \mathrm{X}$ *est un point critique, alors il existe un prolongement* $\widehat{f}$ *de classe* $\mathcal{C}^1$ *au voisinage de* x *tel que* $\mathrm{D}_x \widehat{f} = 0$.

Démonstration. — Soit $\tilde{f}$ un prolongement et notons $\lambda_1, \ldots, \lambda_q$ les multiplicateurs de Lagrange associés à $\tilde{f}$ en x. Posons $\widehat{f} = \tilde{f} - \sum \lambda_j g_j$ et remarquons que cela définit un autre prolongement de f qui vérifie $\mathrm{D}_x \widehat{f} = 0$ par le théorème des extrema liés. ■

Application au calcul de la distance à une droite. — On se place dans $\mathbb{R}^2$ que l'on munit de la distance euclidienne. On se donne une droite D d'équation $ax + by + c = 0$ avec $(a, b) \neq (0, 0)$ et un point $p = (x_0, y_0)$ et on s'intéresse à $\mathrm{dist}(p, \mathrm{D}) = \inf\{\|p - m\|_2,\ m \in \mathrm{D}\}$. On a

$$\mathrm{dist}(p, \mathrm{D}) = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}.$$

Pour montrer cela, on définit $g(x, y) = ax + by + c$ de sorte que $\mathrm{D} = g^{-1}(0)$. Montrons que g est une submersion au-dessus de 0. On a $\mathrm{D}g = \begin{pmatrix} a & b \end{pmatrix}$, donc $\mathrm{D}g$ est surjective en tout point puisque $(a, b) \neq (0, 0)$. Par conséquent D est bien une sous-variété définie par une submersion. Posons maintenant $f(x, y) = (x - x_0)^2 + (y - y_0)^2$. On cherche un minimum absolu de f sur D. On a $\mathrm{D}_{(x,y)} f = 2(x - x_0, y - y_0)$. D'après le théorème des extrema liés, pour tout extremum local $m = (x, y) \in \mathrm{D}$, il existe $\lambda \in \mathbb{R}$ tel que $\mathrm{D}_m f = \lambda \mathrm{D}_m g$. Autrement dit, nous devons résoudre le système d'inconnues x, y, λ suivant:

$$\left\{\begin{array}{rcl} 2(x - x_0) & = & \lambda a \\ 2(y - y_0) & = & \lambda b \\ ax + by + c & = & 0 \end{array}\right. \quad \text{d'où l'on tire} \quad \left\{\begin{array}{rcl} x & = & x_0 + \lambda a/2 \\ y & = & y_0 + \lambda b/2 \end{array}\right.$$

Du coup, on arrive à

$$ax_0 + by_0 + c + \frac{a^2 + b^2}{2}\lambda = 0 \quad \text{et} \quad \lambda = -2\frac{ax_0 + by_0 + c}{a^2 + b^2}.$$

Nous avons donc un seul extremum local, qui doit être le minimum que l'on recherche. Il vient

$$f(x, y) = \left(\frac{ax_0 + by_0 + c}{a^2 + b^2}\right)^2 a^2 + \left(\frac{ax_0 + by_0 + c}{a^2 + b^2}\right)^2 b^2 = \frac{(ax_0 + by_0 + c)^2}{a^2 + b^2}$$

donc

$$d(p, \mathrm{D}) = \sqrt{\frac{(ax_0 + by_0 + c)^2}{a^2 + b^2}} = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}.$$

Application à la diagonalisation des matrices symétriques. — On munit $\mathbb{R}^n$ de la norme euclidienne et on considère la sous-variété

$$\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n,\ \sum x_j^2 = 1\}$$

qui est compacte. On note $g(x) = \sum x_j^2 - 1$. Prenons $\mathrm{A} \in \mathcal{M}_n(\mathbb{R})$ une matrice symétrique et posons $f : \mathbb{S}^{n-1} \to \mathbb{R}$ définie par $f(x) = \langle \mathrm{A}x, x \rangle$.

Un point $a \in \mathbb{S}^{n-1}$ est critique s'il vérifie les conclusions du théorème des extrema liés, donc s'il existe $\lambda \in \mathbb{R}$ tel que $\mathrm{D}_a f = \lambda \mathrm{D}_a g$. Or $\mathrm{D}_a g(h) = 2\langle a, h \rangle$ et

$$f(a + h) = f(a) + \langle \mathrm{A}a, h \rangle + \langle \mathrm{A}h, a \rangle + f(h) = f(a) + 2\langle \mathrm{A}a, h \rangle + o(h)$$

donc $\mathrm{D}_a f(h) = 2\langle \mathrm{A}a, h \rangle$. On obtient ainsi $\langle \mathrm{A}a, h \rangle = \lambda \langle a, h \rangle$ pour tout $h \in \mathbb{R}^n$, soit $\langle \mathrm{A}a - \lambda a, h \rangle = 0$. En prenant $h = \mathrm{A}a - \lambda a$, on trouve finalement $\mathrm{A}a = \lambda a$ et

$$f(a) = \langle \mathrm{A}a, a \rangle = \lambda \langle a, a \rangle = \lambda$$

car $a \in \mathbb{S}^{n-1}$.

Les vecteurs propres unitaires de A sont donc les points critiques de f et leurs valeurs propres sont données par la valeur de f.

Supposons maintenant que F est un sous-espace vectoriel invariant par A ($\mathrm{A}(\mathrm{F}) \subset \mathrm{F}$) et que a est un point critique de la restriction de f sur $\mathrm{F} \cap \mathbb{S}^{n-1}$, alors a est aussi un point critique pour la restriction de f sur $\mathbb{S}^{n-1}$. En effet, on peut trouver une famille orthonormale $\mathrm{e}_1, \dots, \mathrm{e}_k$ tels que $x \in \mathrm{F}$ si et seulement si $\langle x, \mathrm{e}_j \rangle = 0$ pour tout j. Du coup, si a est critique alors

$$\langle \mathrm{A}a, h \rangle = \lambda \langle a, h \rangle + \sum_{1 \leq j \leq k} \mu_j \langle \mathrm{e}_j, h \rangle$$

pour tout $h \in \mathbb{R}^n$. Or, comme $a \in \mathrm{F}$, on a aussi $\mathrm{A}a \in \mathrm{F}$, donc, en prenant $h = \mathrm{e}_j$ pour chaque j, on trouve $\mu_j = 0$ et $\langle \mathrm{A}a, h \rangle = \lambda \langle a, h \rangle$, établissant ainsi que a est aussi critique sur $\mathbb{S}^{n-1}$.

Comme la sphère est compacte, f atteint son maximum en un point a, ce qui montre l'existence d'une valeur propre. Supposons que l'on ait construit par rćeurrence une famille orthonormée $\mathrm{e}_1, \dots, \mathrm{e}_k$ de vecteurs propres. Notons F_k l'orthogonal de ces vecteurs. Comme A est symétrique, on a bien $\mathrm{A}(\mathrm{F}_k) \subset \mathrm{F}_k$, donc un maximum de f sur $\mathbb{S}^{n-1} \cap \mathrm{F}_k$ sera un vecteur propre de A et on pourra procéder par récurrence pour montrer que A admet une base orthonormée de vecteurs propres.

14.2 Hessienne aux points critiques

Nous avons vu précédemment qu'il n'était en général pas possible de donner du sens à la différentielle seconde d'une application f définie sur une sous-variété. On montre que l'on peut définir proprement la hessienne $\mathrm{H}_f(a)$ en un point critique, ce qui nous permettra d'étendre le corollaire 5.11.

On suppose que X est au moins de classe $\mathcal{C}^2$, $f : \mathrm{X} \to \mathbb{R}$ est aussi de classe au moins $\mathcal{C}^2$ et x un point critique de f. Comme d'habitude, φ, g et j désigneront respectivement un redressement local de X, une submersion au-dessus de 0 et une immersion. De même, on note $\tilde{f}$ un prolongement $\mathcal{C}^2$ de f au voisinage de x.

Proposition 14.7. — *Soit $c : \mathrm{I} \to \mathrm{X}$ un chemin de classe $\mathcal{C}^2$ tel que $c(0) = x$; notons $v \overset{def.}{=} c'(0) \in \mathrm{T}_x\mathrm{X}$. On a*

$$(f \circ c)''(0) = \mathrm{D}_0^2(\tilde{f} \circ \varphi^{-1})(\mathrm{D}_x\varphi(v), \mathrm{D}_x\varphi(v)) = \mathrm{D}_0^2(f \circ j)((\mathrm{D}_0 j)^{-1}(v), (\mathrm{D}_0 j)^{-1}(v))$$

et

$$(f \circ c)''(0) = \left(\mathrm{D}_x^2 \tilde{f} - \sum_{j=1}^{q} \lambda_j \mathrm{D}_x^2 g_j \right)(v, v),$$

où $\lambda_1, \dots, \lambda_q$ sont les multiplicateurs de Lagrange associés au prolongement $\tilde{f}$.

Démonstration. — On écrit $f \circ c = \tilde{f} \circ c$ de sorte que $(f \circ c)'(t) = \mathrm{D}_{c(t)}\tilde{f}(c'(t))$ et

$$(f \circ c)''(0) = \mathrm{D}_x^2 \tilde{f}(v, v) + \mathrm{D}_x \tilde{f}(c''(0))\,.$$

Or, on a aussi

$$\sum \lambda_j (g_j \circ c)(t) \equiv 0$$

donc en différentiant deux fois, on obtient

$$\sum \lambda_j \mathrm{D}_{c(t)} g_j(c'(t)) \equiv 0$$

puis

$$\sum \lambda_j \mathrm{D}_x^2 g_j(v, v) + \sum \lambda_j \mathrm{D}_x g(c''(0)) \equiv 0\,.$$

En retranchant cette quantité à $(f \circ c'')(0)$, on obtient par le théorème des extrema liés

$$(f \circ c)''(0) = \mathrm{D}_x^2 \tilde{f}(v, v) - \sum \lambda_j \mathrm{D}_x^2 g_j(v, v)\,.$$

Comme

$$\mathrm{D}_{c(t)}\tilde{f}(c'(t)) = \mathrm{D}_{\varphi(c(t))}(\tilde{f} \circ \varphi^{-1}) \circ \mathrm{D}_{c(t)}\varphi(c'(t))$$

on obtient aussi

$$\begin{aligned}(f\circ c)''(0) &= \mathrm{D}_x^2(\tilde f\circ\varphi^{-1})(\mathrm{D}_0\varphi(v),\mathrm{D}_0\varphi(v)) + \mathrm{D}_0(\tilde f\circ\varphi^{-1})\circ \mathrm{D}_x^2\varphi(v,v)\\ &= \mathrm{D}_x^2(\tilde f\circ\varphi^{-1})(\mathrm{D}_x\varphi(v),\mathrm{D}_x\varphi(v))\end{aligned}$$

car x est un point critique.

Comme $v\in \mathrm{T}_x\mathrm{X}$, on a $\mathrm{D}_x\varphi(v) = (\mathrm{D}_0 j)^{-1}(v)$ et $\tilde f\circ\varphi^{-1}|_{\mathrm{F}} = \tilde f\circ j$. Ceci montre

$$\mathrm{D}_0^2(\tilde f\circ\varphi^{-1})(\mathrm{D}_x\varphi(v),\mathrm{D}_x\varphi(v)) = \mathrm{D}_0^2(f\circ j)((\mathrm{D}_0 j)^{-1}(v),(\mathrm{D}_0 j)^{-1}(v))\,.$$

■

On remarque que $(f\circ c)''(0) = \mathrm{D}_x^2\widehat{f}(v,v)$ où $\widehat{f}$ est donné par le corollaire 14.6.

Définition 14.8 (hessienne en un point critique). — *Si* X *est de classe* $\mathcal{C}^2$ *et* $f:\mathrm{X}\to\mathbb{R}$ *aussi avec point critique* x*, on définit la hessienne de* f *en* x *comme la forme bilinéaire symétrique* $\mathrm{H}_f(x):\mathrm{T}_x\mathrm{X}\times\mathrm{T}_x\mathrm{X}\to\mathbb{R}$ *définie par* $\mathrm{H}_f(x)(v,w) = \mathrm{D}_x^2\widehat{f}(v,w)$*, où* $\widehat{f}$ *est un prolongement tel que* $\mathrm{D}_x\widehat{f} = 0$.

La proposition précédente montre que la hessienne est bien définie et elle donne plusieurs formules pour calculer $\mathrm{H}_f(x)(v,v)$. Le corollaire 5.11 appliqué à $\tilde f\circ j$ implique

Corollaire 14.9 (extrema locaux). — *Soit* $f:\mathrm{X}\to\mathbb{R}$ *une application deux fois différentiables. Si* a *est un maximum local (resp. un minimum local) alors* $\mathrm{D}_a f = 0$ *et* $\mathrm{H}_f(a)$ *est négative (resp. positive). Réciproquement, si* $\mathrm{D}_a f = 0$ *et* $\mathrm{H}_f(a)$ *est définie positive (resp. définie négative) alors* a *est un minimum local strict (resp. maximum local strict).*