Chapitre 1 : DIFFÉRENTIELLE D'UNE APPLICATION

31 décembre 2005

Dans tout ce cours les espaces considérés sont des *espaces vectoriels* sur $\mathbb{R}$, munis d'une norme, notée $\|.\|$. Si E et F sont de tels espaces, l'image par une application *linéaire* $u : E \to F$ d'un vecteur $h \in E$ sera souvent notée $u \cdot h$ et parfois $u(h)$. Lorsque u est continue, sa norme est notée $\|u\|$. L'espace vectoriel des applications linéaires continues de E dans F est noté $L(E, F)$. Il est muni de la norme $\|u\|$.

1 Quelques définitions

Commençons par le

Lemme 1.1 *Si $u : E \to F$ est linéaire continue, et vérifie $\lim_{h\to 0} \dfrac{\|u \cdot h\|}{\|h\|} = 0$, alors $u = 0$.*

Preuve : L'hypothèse signifie que pour tout $\varepsilon > 0$, il existe $\alpha > 0$ tel que $\|h\| < \alpha$ implique $\|u \cdot h\| \leq \varepsilon \|h\|$. Soit $x \in E$. Si $x \neq 0$, le vecteur $h = \frac{\alpha x}{2\|x\|}$ a pour norme $\frac{\alpha}{2} < \alpha$, donc $\|u(\frac{\alpha x}{2\|x\|})\| \leq \varepsilon \left\| \frac{\alpha x}{2\|x\|} \right\|$, ce qui s'écrit, en utilisant la linéarité, $\|u \cdot x\| \leq \varepsilon \|x\|$. On en déduit $\|u\| \leq \varepsilon$ pour tout $\varepsilon > 0$, d'où $\|u\| = 0$ et donc $u = 0$. ∎

Définition 1.2 *Soit U un ouvert de E et $f : U \to F$ une application. L'application f est dite* différentiable *en $a \in U$ s'il existe une application* linéaire continue *$u : E \to F$ telle que*

$$\lim_{x\to a} \frac{f(x) - f(a) - u(x - a)}{\|x - a\|} = 0.$$

L'application f est dite différentiable dans U si elle l'est en chaque point de U.

Il est utile de remarquer que, contrairement à f qui n'est supposée définie que sur l'ouvert U, l'application u, comme toute application linéaire, est définie sur l'espace vectoriel E tout entier. On rappelle d'ailleurs que dans un espace

normé le sous-espace vectoriel engendré par un ouvert non vide est toujours l'espace entier.
Le lemme 1.1 justifie de donner un nom à l'application linéaire continue u.

Définition 1.3 *Si f est différentiable en a, l'application linéaire continue u est appelée* différentielle *de f en a. On la note $Df(a)$.*

On peut alors expliciter la définition sous la forme : si f est différentiable en a, pour tout $\varepsilon > 0$, il existe $r > 0$ tel que $\|h\| < r$ implique $a + h \in U$ et

$$\|f(a+h) - f(a) - Df(a) \cdot h\| < \varepsilon \|h\|.$$

Proposition 1.4 *Si $f : U \to F$ est différentiable en $a \in U$, alors pour tout $h \in E$,*

$$\lim_{t \to 0} \frac{f(a+th) - f(a)}{t} = Df(a) \cdot h$$

Preuve : On écrit

$$\frac{f(a+th) - f(a)}{t} = \frac{f(a+th) - f(a) - Df(a)(th)}{t} + Df(a) \cdot h$$

et on utilise la définition en remarquant que si $t \to 0$ alors $th \to 0$. ∎

Il est essentiel de signaler tout de suite que la réciproque n'est pas vraie. L'existence d'une limite pour chaque $h \in E$ permet de deviner $Df(a)$ mais ne dispense de montrer ni la linéarité ni la continuité.
Remarquons cependant que cette proposition permet de s'assurer que dans le cas particulier où $E = F = \mathbb{R}$, f est différentiable en a si et seulement si elle est dérivable en a. Si c'est le cas l'application linéaire $Df(a) : \mathbb{R} \to \mathbb{R}$ est $Df(a) \cdot t = f'(a)t$.

Proposition 1.5 *Si f est différentiable en a, elle est continue en a.*

Preuve En choisissant $\varepsilon = 1$, on voit qu'il existe $\alpha > 0$ tel que si $\|x - a\| < \alpha$, $x \in U$ et $\|f(x) - f(a) + Df(a)(x-a)\| < \|x - a\|$. D'où $\|f(x) - f(a)\| < \|Df(a)(x-a)\| + \|x - a\| \leq (1 + \|Df(a)\|)\|x - a\|$, ce qui montre que f est « localement lipschitzienne » en a, donc continue. ∎

La démonstration de la proposition suivante est laissée en exercice.

Proposition 1.6 *Si on remplace les normes choisies sur E et F par des normes* équivalentes*, la différentiabilité d'une application en un point et la valeur de la différentielle sont inchangées.*

Rappelons que sur un espace vectoriel de dimension finie *toutes* les normes sont équivalentes et que cette propriété est *caractéristique* des espaces de dimension finie.

La proposition suivante décrit le cas où l'espace F est un produit d'espaces normés, $F = F_1 \times F_2 \times \cdots \times F_n$, muni de la norme produit :

$$\|(y_1, \cdots, y_n)\| = \max_{j=1\cdots n} \|y_j\|$$

ou de toute norme équivalente. Définir une application f de U dans F c'est se donner n applications $f_j : U \to F_j$, de sorte que pour tout $x \in U$,

$$f(x) = (f_1(x), \cdots, f_n(x)).$$

Les applications f_j sont appelées *composantes* de f. On vérifie aisément que lorsque $U = E$, f est linéaire si et seulement si chaque composante f_j l'est. On sait que la continuité de f est équivalente à celle de chacune des composantes. La proposition suivante montre que la même propriété est vraie en ce qui concerne la différentiabilité.

Proposition 1.7 *L'application f est différentiable en $a \in U$ si et seulement si chaque composante f_j l'est. Dans ce cas*

$$Df(a) = (Df_1(a), \cdots, Df_n(a))$$

où $Df_j(a) \in L(E, F_j)$ est la différentielle de f_j en a.

Preuve Si f est différentiable en a, $Df(a) : E \to F_1 \times \cdots \times F_n$ a n composantes : $Df(a) = (u_1, \cdots, u_n)$ où $u_j : E \to F_j$ est linéaire continue. De plus, pour tout $\varepsilon > 0$, il existe $\alpha > 0$ tel que, si $\|h\| < \alpha$, $a + h \in U$ et

$$\|f(a + h) - f(a) - Df(a) \cdot h\| < \varepsilon \|h\|.$$

En explicitant la norme de F et en utilisant les composantes de f et de $Df(a)$ ceci s'écrit :

$$\max_{j=1\cdots n} \|f_j(a + h) - f_j(a) - u_j \cdot h\| < \varepsilon \|h\|.$$

Cela prouve que, pour chaque j, f_j est différentiable en a et que $Df_j(a) = u_j$.

Réciproquement, supposons chaque f_j différentiable en a. L'application $u = (Df_1(a), \cdots, Df_n(a))$ est linéaire et continue d'après ce qui est rappelé. De plus en écrivant simultanément pour les n composantes la différentiabilité en a, on voit que pour tout $\varepsilon > 0$, il existe $\alpha > 0$ tel que, si $\|h\| < \alpha$, $a + h \in U$ et pour $j = 1 \cdots n$, $\|f_j(a + h) - f_j(a) - Df_j(a) \cdot h\| < \varepsilon \|h\|$. Alors

$$\|f(a + h) - f(a) - u.h\| = \max_{j=1\cdots n} \|f_j(a + h) - f_j(a) - Df_j(a) \cdot h\| < \varepsilon \|h\|$$

ce qui montre que f est différentiable en a et que $Df(a) = u$. ∎

Ceci s'applique en particulier au cas où F est de *dimension finie*. En appelant n cette dimension, le choix d'une base $\mathcal{W} = (w_1, \cdots, w_n)$ de F permet d'identifier (topologiquement) F à $\mathbb{R}^n$. Les composantes de f sont alors appelées *fonctions coordonnées*. Leurs différentielles (lorsqu'elles existent) sont des *formes* linéaires continues sur E.
L'utilisation de la base $\mathcal{W}$ conduit aux écritures suivantes. Pour tout $x \in U$,

$$f(x) = \sum_{j=1}^{n} f_j(x) w_j$$

et pour tout $h \in E$,

$$Df(a) \cdot h = \sum_{j=1}^{n} (Df_j(a) \cdot h) w_j$$

où $Df_j(a) \cdot h$ est un scalaire.

Si de plus $E = \mathbb{R}$, $f : U \subset \mathbb{R} \to \mathbb{R}^n$ est donnée par $f(x) = (f_1(x), \cdots, f_n(x))$. Elle est différentiable en a si et seulement si chaque fonction coordonnée est dérivable en a. Dans ce cas $Df(a) \cdot t = t(f_1'(a), \cdots, f_n'(a))$ est le produit du scalaire t par le vecteur de $\mathbb{R}^n$ dont les composantes sont les dérivées en a des composantes de f. On note $f'(a)$ ce vecteur que l'on appelle *vecteur dérivée* de f en a. On a donc $Df(a) \cdot t = tf'(a)$ et en particulier $f'(a) = Df(a) \cdot 1$.

Remarquons que, plus généralement, lorsque $E = \mathbb{R}$, si F est un espace normé quelconque, $f : U \subset \mathbb{R} \to F$ est différentiable en a si et seulement si $\lim_{t \to 0} \dfrac{f(a+t) - f(a)}{t}$ existe. C'est alors un vecteur de F que l'on note $f'(a)$ et qu'on appelle vecteur dérivée de f en a. La différentielle de f en a est alors l'application linéaire (automatiquement continue) $Df(a) : \mathbb{R} \to F$ définie par $Df(a) \cdot t = tf'(a)$. En particulier $Df(a) \cdot 1 = f'(a)$. Autrement dit lorsque $E = \mathbb{R}$ (et dans ce cas seulement) la proposition 1.4 a une réciproque.

2 Exemples

2.1 Application affine

Une application affine $A : E \to F$ est de la forme $A(x) = b + u(x)$ où $b \in F$ et u est une application linéaire de E dans F. Pour $a, x \in E$,

$$(1) \quad A(x) - A(a) = u(x) - u(a) = u(x - a)$$

par linéarité. Donc A est continue en a si et seulement si u est continue en 0, donc, puisque u est linéaire, en tout point. Dans ce cas la formule (1) montre que A est différentiable en tout point de E et que, pour tout $a \in E$,

$$DA(a) = u.$$

La différentielle d'une application affine continue est donc *indépendante* de a. En particulier si u est linéaire continue, pour tout $x \in E$, $Du(x) = u$.

2.2 Un exemple en dimension finie

Soit $f : \mathbb{R}^2 \to \mathbb{R}^2$ l'application définie par $f(x, y) = (x + y, xy)$. Soit $(a, b) \in \mathbb{R}^2$. Pour $(h, k) \in \mathbb{R}^2$, on a

$$f(a + h, b + k) - f(a, b) = (h + k, hb + ka + hk) = u(h, k) + \varphi(h, k)$$

où $u(h, k) = (h + k, hb + ka)$ et $\varphi(h, k) = (0, hk)$.
L'application u est linéaire continue (on est en dimension finie). Sa matrice dans les bases canoniques est $Jf(a, b) = \begin{pmatrix} 1 & 1 \\ b & a \end{pmatrix}$.
Choisissons sur $\mathbb{R}^2$ la norme $\|(h, k)\| = \max(|h|, |k|)$. La fonction $\varphi(h, k)$ vérifie $\|\varphi(h, k)\| = |hk| \leq \|(h, k)\|^2$. Donc $\lim\limits_{(h,k)\to 0} \dfrac{\varphi(h, k)}{\|(h, k)\|} = 0$, ce qui montre que l'application f est différentiable en (a, b), sa différentielle en (a, b) étant l'application linéaire de $\mathbb{R}^2$ dans $\mathbb{R}^2$ de matrice $Jf(a, b)$.

2.3 Un exemple en dimension infinie

Appelons E l'espace vectoriel des fonctions continûment dérivables sur $[\,0, 1\,]$ à valeurs dans $\mathbb{R}$, muni de la norme $\|f\| = \|f\|_\infty + \|f'\|_\infty$. Soit F l'espace vectoriel des fonctions continues sur $[\,0, 1\,]$ à valeurs dans $\mathbb{R}$, muni de la norme $\|.\|_\infty$. On considère l'application $H : E \to F$ donnée par $H(f) = f' + f^2$. Fixons $f \in E$. On a pour $h \in E$,

$$H(f + h) - H(f) = h' + 2fh + h^2$$

L'application définie sur E par $u(h) = h' + 2fh$ est visiblement linéaire, mais cette fois il faut montrer sa continuité. On a

$$\|u(h)\|_\infty = \|h' + 2fh\|_\infty \leq \|h'\|_\infty + 2\|f\|_\infty\|h\|_\infty \leq (1 + 2\|f\|_\infty)\|h\|$$

ce qui établit le résultat.
D'autre part $H(f + h) - H(f) - u(h) = h^2$ et $\|h^2\|_\infty \leq \|h\|_\infty^2 \leq \|h\|^2$, de sorte que $\lim\limits_{h\to 0} \dfrac{H(f + h) - H(f) - u(h)}{\|h\|} = 0$.
On a donc prouvé la différentiabilité de H en f, la différentielle étant l'application linéaire continue de E dans F définie par $DH(f) \cdot h = h' + 2fh$.
Remarquons que si on choisit sur E la norme $\|.\|_\infty$ au lieu de $\|.\|$, l'application u n'est pas continue et H n'est pas différentiable. Il est en effet impossible de déterminer une constante $C > 0$ telle que $\|h' + 2fh\|_\infty \leq C\|h\|_\infty$ pour tout $h \in E$, puisqu'en choisissant $h_n = x^n$, on devrait avoir pour tout n, $C \geq \dfrac{\|h_n' + 2fh_n\|_\infty}{\|h_n\|_\infty} = \sup\limits_{x\in[\,0,1\,]} |nx^{n-1} + 2x^n f(x)| \geq |n + 2f(1)|$ et que cette quantité tend vers l'infini avec n.

3 Différentielles partielles

3.1 Différentiabilité dans une « direction »

Soit $f : U \to F$ et $a \in U$. Soit $v \in E$, $v \neq 0$, la fonction d'une variable $\varphi_v : t \mapsto f(a+tv)$ est définie au voisinage de 0 (U est ouvert). Elle est différentiable en 0 si et seulement si $\varphi'_v(0) = \lim_{t\to 0} \dfrac{f(a+tv)-f(a)}{t}$ existe. On dit alors que f admet une *dérivée partielle* en a dans la direction v, égale à $\varphi'_v(0)$. C'est un vecteur de F.

D'après la Proposition 1.4, c'est le cas pour tout $v \neq 0$ si f est différentiable en a et $\varphi'_v(0) = Df(a) \cdot v$.

Il n'y a pas de réciproque à cette propriété sauf si $E = \mathbb{R}$, comme le montre l'exemple suivant en dimension 2.

Prenons $U = \mathbb{R}^2$, $F = \mathbb{R}$ et f définie par $f(x,y) = \dfrac{x^2 y}{x^2+y^2}$ si $(x,y) \neq (0,0)$ et $f(0,0) = 0$. Etudions f au point $(0,0)$.
En remarquant que pour tout $t \in \mathbb{R}$ et tout $v = (x,y) \in \mathbb{R}^2$, on a $f(tv) = tf(v)$, on en déduit, pour $t \neq 0$, $\dfrac{f(tv)-f(0)}{t} = f(v)$ et donc f admet $f(v)$ pour dérivée partielle en $(0,0)$ dans la direction v, pour tout $v \neq 0$. Pourtant f n'est pas différentiable en $(0,0)$. Si elle l'était, le calcul précédent montre (Proposition 1.4) que sa différentielle serait f. Mais f n'est pas linéaire et ne saurait donc être une différentielle.

Plus généralement, si $E_1 \neq 0$ est un sous-espace vectoriel de E, il existe un voisinage ouvert U_1 de 0 dans E_1 tel que si $h_1 \in U_1$, $a + h_1 \in U$: ceci parce que U est ouvert et l'application $E_1 \to E$ qui à h_1 associe $a + h_1$ continue. On peut alors définir l'application partielle $\varphi_1 : U_1 \to F$ donnée par $\varphi_1(h_1) = f(a+h_1)$ et se poser le problème de sa différentiabilité en 0.

Définition 3.1 *Si φ_1 est différentiable en 0, on dit que f admet une différentielle partielle en a dans la direction E_1, égale à $D\varphi_1(0)$.*

Bien sûr $D\varphi_1(0) \in L(E_1, F)$.

Proposition 3.2 *Si $f : U \to F$ est différentiable en $a \in U$, elle admet en a dans* toute *direction $E_1 \neq 0$ de E une différentielle partielle égale à la restriction à E_1 de la différentielle de f en a.*

Preuve On sait que quel que soit $\varepsilon > 0$, il existe $\delta > 0$ tel que si $h \in E$ vérifie $\|h\| < \delta$, alors $\|f(a+h)-f(a)-Df(a)\cdot h\| < \varepsilon\|h\|$. C'est vrai a fortiori si $h_1 \in E_1$ vérifie $\|h_1\| < \delta$ et cela s'écrit alors $\|\varphi_1(h_1) - \varphi_1(0) - Df(a)_{|_{E_1}} \cdot h_1\| < \varepsilon\|h_1\|$ et le résultat suit en remarquant que $Df(a)_{|_{E_1}} \in L(E_1, F)$. ∎

Bien sûr il n'y a pas de réciproque.

Si $E = E_1 \times E_2$ muni de la norme produit, E_1 et E_2 peuvent être vus comme des sous-espaces vectoriels de E en identifiant E_1 à $E_1 \times 0 = \{(h_1, 0) \mid h_1 \in E_1\}$ (*resp.* E_2 à $0 \times E_2$) et en remarquant que la norme dans E de $(h_1, 0)$ est égale à la norme dans E_1 de h_1. Si $f : U \to F$ admet en $a \in U$ une différentielle partielle dans la direction E_1 (*resp.* E_2), on la note $D_1 f(a)$ (*resp.* $D_2 f(a)$). Si f est différentiable en $a \in U$, la proposition 3.2 permet d'écrire l'utile formule suivante. Si $h = (h_1, h_2) \in E = E_1 \times E_2$, puisque $h = (h_1, 0) + (0, h_2)$,

$$Df(a) \cdot h = Df(a) \cdot (h_1, 0) + Df(a) \cdot (0, h_2) = D_1 f(a) \cdot h_1 + D_2 f(a) \cdot h_2.$$

3.2 Dérivées partielles et matrice jacobienne

Lorsque E est de dimension finie, on l'identifie à $\mathbb{R}^m$ en y choisissant une base $\mathcal{V} = (v_1, \cdots, v_m)$. Alors $f : U \to F$ s'identifie à une fonction de m *variables réelles* $(x_1, \cdots, x_m)$: les coordonnées dans la base $\mathcal{V}$. Soit $a = (a_1, \cdots, a_m) \in U$, la dérivée partielle de f en a dans la direction v_j (si elle existe) est le vecteur de F qui est la dérivée en a_j de la fonction d'une variable

$$x_j \mapsto f(a_1, \cdots, a_{j-1}, x_j, a_{j+1}, \cdots, a_m).$$

On le note $\dfrac{\partial f}{\partial x_j}(a)$. On a donc si $h = \displaystyle\sum_{j=1}^{m} h_j v_j$,

$$Df(a) \cdot h = \sum_{j=1}^{m} h_j \frac{\partial f}{\partial x_j}(a)$$

où les h_j sont des scalaires et les $\dfrac{\partial f}{\partial x_j}(a)$ des vecteurs de F. On a aussi la formule

$$\frac{\partial f}{\partial x_j}(a) = Df(a) \cdot v_j$$

Si de plus F est de dimension finie et identifié à $\mathbb{R}^n$ par le choix d'une base $\mathcal{W} = (w_1, \cdots, w_n)$, la fonction f se représente par les n fonctions coordonnées $(f_1, \cdots, f_n)$ où $f_i : U \to \mathbb{R}$ est une fonction de m variables à valeurs réelles. Les f_i sont différentiables en a si et seulement si f l'est et dans ce cas pour tout $h \in E$,

$$Df_i(a) \cdot h = \sum_{i=1}^{n} (Df_i(a) \cdot h)\, w_i$$

où, si $h = \displaystyle\sum_{j=1}^{m} h_j v_j$, alors $Df_i(a) \cdot h = \displaystyle\sum_{j=1}^{m} h_j \frac{\partial f_i}{\partial x_j}(a)$.

Cette fois $\dfrac{\partial f_i}{\partial x_j}(a)$ est un réel et on peut aussi écrire $Df_i(a) \cdot h = \displaystyle\sum_{j=1}^{m} \frac{\partial f_i}{\partial x_j}(a) h_j$.

L'application linéaire $Df(a)$ est donc celle qui associe au vecteur h de E le vecteur $Df(a)\cdot h = \sum_{i=1}^{n}\left(\sum_{j=1}^{m}\frac{\partial f_i}{\partial x_j}(a)h_j\right)w_i$ de F. En d'autres termes la matrice de l'application linéaire $Df(a)$ dans les bases $\mathcal{V}$ et $\mathcal{W}$ est la matrice à n lignes et m colonnes notée $Jf(a)$ donnée par

$$Jf(a) = \left(\frac{\partial f_i}{\partial x_j}(a)\right)_{1\leq i\leq n, 1\leq j\leq m}$$

où i est l'indice ligne et j l'indice colonne.

Définition 3.3 *La matrice $Jf(a)$ est appelée* matrice jacobienne *de f en a dans les bases $\mathcal{V}$ et $\mathcal{W}$.*

Une ligne de la matrice jacobienne est constituée des dérivées partielles d'une même composante de f et une colonne par les dérivées partielles par rapport à une même variable.
Remarquons que, si $E = E_1 \times E_2$, ce qui correspond, via l'identification déjà faite, à supposer que $(v_1, \cdots, v_r)$ est une base de E_1 et $(v_{r+1}, \cdots, v_m)$ une base de E_2, la matrice $Jf(a)$ est de la forme

$$Jf(a) = (J_1f(a), J_2f(a))$$

où $J_if(a)$ est la matrice jacobienne de $D_if(a)$, $(i = 1, 2)$.

Définition 3.4 *Si $n = m$, le déterminant de la matrice (carrée) $Jf(a)$ est appelé* jacobien *de f en a et noté $jf(a)$.*

La condition $jf(a) \neq 0$ exprime que $Df(a)$ est un isomorphisme de E sur F, ou que $Df(a) \in GL(E, F)$.

4 Différentiabilité et opérations classiques

4.1 Combinaison linéaire et composition

Le premier résultat est immédiat.

Proposition 4.1 *Soit f_1 et f_2 deux applications définies sur un ouvert U d'un espace normé E à valeurs dans un espace normé F, différentiables en $a \in U$. Alors pour tous scalaires λ et μ, l'application $f = \lambda f_1 + \mu f_2$ est différentiable en a et $Df(a) = \lambda Df_1(a) + \mu Df_2(a)$.*

La proposition suivante montre que la notion de différentielle se comporte vis-à-vis de la composition de la façon la plus simple possible.

Proposition 4.2 *Soit $f : U \to F$, $a \in U$, V un ouvert de F contenant $f(a) = b$ et $g : V \to G$. Si f est différentiable en a et g en b, alors $g \circ f$ est différentiable en a et*

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

Preuve On remarque d'abord que la composée de deux applications linéaires continues étant linéaire continue, puisque $Df(a) \in L(E, F)$ et $Dg(b) \in L(F, G)$, $Dg(b) \circ Df(a) \in L(E, G)$.
Puisque f est continue en a, il existe $r > 0$ tel que pour $x \in Bo(a, r)$, $f(x) \in V$ donc $g \circ f$ est définie sur cette boule. De plus si r est assez petit et $\|h\| < r$, la preuve de la proposition 1.5 montre que

$$\|f(a + h) - f(a)\| \leq (1 + \|Df(a)\|)\|h\|.$$

D'autre part, pour tout $\varepsilon > 0$, il existe $\alpha > 0$ tel que, pour tout $k \in F$, si $\|k\| < \alpha$,

$$\|g(b + k) - g(b) - Dg(b) \cdot k\| \leq \varepsilon\|k\|.$$

Or $g \circ f(a + h) = g(b + k)$ où $k = f(a + h) - b = f(a + h) - f(a)$.
Si $\|h\| < \alpha_1 = \min\left(r, \dfrac{\alpha}{1 + \|Df(a)\|}\right)$, on a donc $\|k\| < \alpha$ et par suite :
$\|g \circ f(a+h) - g \circ f(a) - Dg(b) \circ Df(a)(h)\| = \|g(b+k) - g(b) - Dg(b) \cdot k + Dg(b) \cdot [k - Df(a) \cdot h]\| \leq \varepsilon\|f(a + h) - f(a)\| + \|Dg(b)\| \, \|f(a + h) - f(a) - Df(a) \cdot h\|$.
Mais il existe $\eta > 0$ tel que $\|h\| < \eta$ implique $\|f(a+h) - f(a) - Df(a) \cdot h\| < \varepsilon\|h\|$.
Si donc $\|h\| < \min(\alpha_1, \eta)$, on a

$$\|g \circ f(a + h) - g \circ f(a) - Dg(b) \circ Df(a)(h)\| < \varepsilon(1 + \|Df(a)\| + \|Dg(b)\|) \, \|h\|.$$

C'est le résultat annoncé. ∎

Dans le cas où les espaces E et F sont tous deux de dimension finie, cette proposition s'explicite en termes de matrices jacobiennes :

$$J(g \circ f)(a) = Jg(b) \, Jf(a)$$

Il s'agit bien sûr du produit des deux matrices.

4.2 Applications bilinéaires et produits

Proposition 4.3 *Soit E, F et G des espaces normés et $B : E \times F \to G$ une application bilinéaire. Alors B est continue si et seulement s'il existe $C > 0$ tel que pour tout $x \in E$ et tout $y \in F$, $\|B(x, y)\| \leq C\|x\| \, \|y\|$. Dans ce cas B est différentiable dans $E \times F$ et pour tout $(x, y) \in E \times F$, $DB(x, y)$ est l'application* linéaire *continue de $E \times F$ dans G donnée par*

$$(h, k) \mapsto DB(x, y) \cdot (h, k) = B(x, k) + B(h, y).$$

Preuve L'affirmation concernant la continuité se montre comme son analogue pour les applications linéaires. Si B est continue, elle l'est en $(0,0)$ et il existe $\alpha > 0$ tel que si $\|(x,y)\| = \max(\|x\|, \|y\|) < \alpha$, alors $\|B(x,y)\| \leq 1$. Si $x \in E$, $x \neq 0$ et $y \in F$, $y \neq 0$, $(u,v) = \left(\frac{\alpha x}{2\|x\|}, \frac{\alpha y}{2\|y\|}\right)$ vérifie $\|(u,v)\| = \dfrac{\alpha}{2} < \alpha$. D'où $\|B(u,v)\| = \left\|B\left(\frac{\alpha x}{2\|x\|}, \frac{\alpha y}{2\|y\|}\right)\right\| \leq 1$ ou encore, en utilisant la bilinéarité et la propriété convenable d'une norme,

$$\|B(x,y)\| \leq \frac{4}{\alpha^2}\,\|x\|\,\|y\|$$

ce qui permet de choisir $C = \dfrac{4}{\alpha^2}$.

Montrons maintenant que cette propriété implique la différentiabilité de B. Cela établira a fortiori sa continuité. (Une peuve directe est un excellent exercice...) Tout d'abord une vérification facile montre que l'application $u : (h,k) \mapsto B(x,k) + B(h,y)$ est linéaire. La majoration

$$\|u \cdot (h,k)\| \leq C(\|x\|\,\|k\| + \|h\|\,\|y\|) \leq C(\|x\| + \|y\|)\|(h,k)\|$$

permet d'affirmer la continuité de u (et aussi $\|u\| \leq C(\|x\| + \|y\|)$).
Ensuite si $(h,k) \in E \times F$, la bilinéarité de B permet d'écrire

$$B(x+h, y+k) - B(x,y) - u \cdot (h,k) = B(h,k)$$

et la majoration $\|B(h,k)\| \leq C\|h\|\,\|k\| \leq C\|(h,k)\|^2$ montre que $\dfrac{B(h,k)}{\|(h,k)\|} \to 0$ quand $\|(h,k)\| \to 0$. ∎

Exemples

1. Soit E un espace normé et $L(E)$ l'espace vectoriel normé des applications linéaires continues de E dans E. L'application $B : L(E) \times L(E) \to L(E)$ définie par $(u,v) \mapsto u \circ v$ est bilinéaire continue car $\|u \circ v\| \leq \|u\|\,\|v\|$. Elle est donc différentiable dans $L(E) \times L(E)$ et sa différentielle en (u,v) est donnée par
$$(h,k) \mapsto DB(u,v) \cdot (h,k) = u \circ k + h \circ v.$$

2. Si E est muni d'un produit scalaire $E \times E \to \mathbb{R}$ noté $\langle x, y\rangle$ et si la norme sur E est la norme associée : $\|x\| = \sqrt{\langle x, x\rangle}$, la forme bilinéaire $B : E \times E \to \mathbb{R}$, $B(x,y) = \langle x, y\rangle$ est continue donc différentiable en tout point. Lorsque $E = \mathbb{R}^m$, la norme $\|.\|_2$ est obtenue de cette façon à partir du produit scalaire usuel.

En combinant les propositions 4.3 et 4.2 on obtient nombre de résultats classiques. En voici quelques uns.

1. **Produit de deux fonctions numériques**

   Proposition 4.4 *Si f et $g : U \to \mathbb{R}$ sont différentiables en $a \in U$, leur produit $\varphi = f\,g$, défini par $\varphi(x) = f(x)g(x)$, est différentiable en a. La différentielle est la forme linéaire continue donnée, pour $h \in E$, par*

   $$D\varphi(a) \cdot h = g(a)Df(a) \cdot h + f(a)Dg(a) \cdot h$$

   Preuve On a $\varphi = B \circ \Phi$ où $\Phi : U \to \mathbb{R}^2$ a pour composantes f et g et B est la forme bilinéaire continue $\mathbb{R}^2 \to \mathbb{R}$ donnée par $B(s,t) = st$. La proposition 1.7 montre que Φ est différentiable en a et que $D\Phi(a) = (Df(a), Dg(a))$. D'autre part la proposition 4.2 donne pour tout $h \in E$, $D\varphi(a)\cdot h = DB(\Phi(a)) \circ D\Phi(a)(h) = DB(f(a), g(a))\cdot(Df(a)\cdot h, Dg(a)\cdot h) = B(f(a), Dg(a)\cdot h) + B(Df(a)\cdot h, g(a)) = f(a)Dg(a)\cdot h + g(a)Df(a)\cdot h$. ∎

2. **Produit d'une application par une fonction numérique**

   Proposition 4.5 *Si $f : U \to F$ et $\varphi : U \to \mathbb{R}$ sont différentiables en a, alors l'application $g = \varphi\, f$, c'est-à-dire $g(x) = \varphi(x)f(x)$ est différentiable en a et pour tout $h \in E$,*

   $$Dg(a) \cdot h = \varphi(a)Df(a) \cdot h + (D\varphi(a) \cdot h)\, f(a).$$

   (Dans cette formule, $\varphi(a)$ et $D\varphi(a) \cdot h$ sont des scalaires tandis que $f(a)$ et $Df(a) \cdot h$ sont des vecteurs de F.)

   Preuve La preuve est identique en écrivant $g = B \circ \Phi$ où $\Phi : U \to \mathbb{R} \times F$ a pour composantes (φ, f) et B est l'application bilinéaire de $\mathbb{R} \times F \to F$ donnée par $B(\lambda, y) = \lambda y$. La continuité de B est conséquence de $\|\lambda y\| = |\lambda|\,\|y\|$. ∎

3. **Produit scalaire**

   Proposition 4.6 *Supposons l'espace F muni d'une norme associée à un produit scalaire. Soit $f, g : U \to F$ des fonctions différentiables en $a \in U$. L'application $\varphi : U \to \mathbb{R}$ définie par $\varphi(x) = \langle f(x), g(x)\rangle$ est différentiable en a et sa différentielle est la forme linéaire continue sur E donnée par*

   $$D\varphi(a) \cdot h = \langle Df(a) \cdot h, g(a)\rangle + \langle f(a), Dg(a) \cdot h\rangle$$

   Preuve On écrit $\varphi = B \circ \Phi$ où $\Phi = (f, g) : U \to F \times F$ et B est le produit scalaire. ∎

   Remarque Si E est de dimension finie et $F = \mathbb{R}$, la formule

   $$Df(a) \cdot h = \sum_{j=1}^{m} \frac{\partial f}{\partial x_j}(a)h_j$$

   peut se lire

   $$Df(a) \cdot h = \langle grad\, f(a), h\rangle$$

où $grad\, f(a)$ (appelé *gradient* de f en a) est le vecteur de E

$$grad\, f(a) = \sum_{j=1}^{m} \frac{\partial f}{\partial x_j}(a) v_j$$

si $\mathcal{V} = (v_1, \cdots, v_m)$ est la base choisie dans E.

4.3 Inverse

Ce mot peut désigner

– la fonction $\dfrac{1}{f(x)}$ lorsque la fonction $f : U \to \mathbb{R}$ ne s'annule pas,

– l'application f^{-1} lorsque f est une bijection.

Il ne faut pas confondre ces deux situations... et il y a un théorème pour chacune d'elles.

Proposition 4.7 *Soit $f : U \to \mathbb{R}$ une fonction différentiable en $a \in U$ telle que $f(a) \neq 0$. Alors la fonction $g(x) = \dfrac{1}{f(x)}$ est définie au voisinage de a, différentiable en a et, pour tout $h \in E$,*

$$Dg(a) \cdot h = -\frac{1}{f(a)^2}\, Df(a) \cdot h$$

Preuve Puisque $f(a) \neq 0$ et que f est continue en a (puisque différentiable), il existe un voisinage V de a tel que, pour tout $x \in V$, $f(x) \neq 0$. L'application g est définie sur V et s'écrit $g = \imath \circ f$ où $\imath : \mathbb{R}^* \to \mathbb{R}^*$ est donnée par $\imath(y) = \dfrac{1}{y}$. C'est une fonction dérivable en tout point et $\imath'(y) = -\dfrac{1}{y^2}$. Autrement dit la différentielle de $\imath$ en y est l'application linéaire $D\imath(y) : \mathbb{R} \to \mathbb{R}$ définie par $D\imath(y) \cdot t = -\dfrac{t}{y^2}$ et tout découle de la proposition 4.2.∎

Proposition 4.8 *Soit U un ouvert de E, V un ouvert de F et $f : U \to V$ un* homéomorphisme. *On suppose que f est différentiable en a et que $Df(a) \in GL(E, F)$, c'est-à-dire qu'elle est linéaire, continue, bijective et que son inverse est continue. Alors $f^{-1} : V \to E$ est différentiable en $b = f(a)$ et*

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

Remarquons que la condition $Df(a) \in GL(E, F)$ suppose déjà qu'il y a des isomorphismes entre E et F et ceci est exigeant : en dimension finie cela suppose que F et E ont la même dimension, en dimension infinie que F (*resp.* E) soit un Banach dès que E (*resp.* F) l'est.

Preuve Si f^{-1} est différentiable en b, la proposition 4.2 donne pour sa différentielle la formule indiquée. En effet la proposition 4.2 permet de déduire de l'égalité $f \circ f^{-1} = id$ (sur V) et de $f^{-1}(b) = a$ l'égalité $Df(a) \circ Df^{-1}(b) = id$ (sur F). De même les égalités $f^{-1} \circ f = id$ (sur U) et $f(a) = b$ donnent $Df^{-1}(b) \circ Df(a) = id$ (sur E). Il faut donc établir qu'en posant, pour $k \in F$ de norme assez petite,

$$\varphi(k) = f^{-1}(b+k) - f^{-1}(b) - Df(a)^{-1} \cdot k$$

on a $\lim_{k \to 0} \dfrac{\varphi(k)}{\|k\|} = 0$.
Comme f est différentiable en a, si $\varepsilon > 0$ est fixé, il existe $\alpha > 0$ tel que si $h \in E$ vérifie $\|h\| < \alpha$, alors $a + h \in U$ et

$$(\star) \quad \|f(a+h) - f(a) - Df(a) \cdot h\| < \varepsilon \|h\|$$

Comme f^{-1} est continue en b, il existe $\eta > 0$ tel que $\|k\| < \eta$ implique $b + k \in V$ et $\|f^{-1}(b+k) - f^{-1}(b)\| < \alpha$ et donc, en posant $h = f^{-1}(b+k) - f^{-1}(b)$ de sorte que $a + h = f^{-1}(b+k)$, $(\star)$ s'écrit $\|b + k - b - Df(a) \cdot h\| < \varepsilon \|h\|$. D'où

$$\|Df(a) \cdot h\| \leq \|k\| + \varepsilon \|h\|.$$

Mais $h = Df(a)^{-1} \circ Df(a) \cdot h$ et donc

$$\|h\| \leq \|Df(a)^{-1}\| \, \|Df(a) \cdot h\| \leq \|Df(a)^{-1}\| (\|k\| + \varepsilon \|h\|)$$

ou encore

$$(1 - \varepsilon \|Df(a)^{-1}\|) \|h\| \leq \|Df(a)^{-1}\| \, \|k\|.$$

Si on prend $\varepsilon < \dfrac{1}{2\|Df(a)^{-1}\|}$, on a donc

$$(\star\star) \quad \|h\| \leq 2\|Df(a)^{-1}\| \, \|k\|.$$

Mézalor

$$\begin{aligned} \varphi(k) &= h - Df(a)^{-1} \cdot k = h - Df(a)^{-1} \cdot [(f(a+h) - f(a)] \\ &= -Df(a)^{-1} \cdot [f(a+h) - f(a) - Df(a) \cdot h] \end{aligned}$$

et donc en utilisant $(\star)$ et $(\star\star)$:

$$\|\varphi(k)\| \leq \|Df(a)^{-1}\| \varepsilon \|h\| \leq 2\|Df(a)^{-1}\|^2 \, \varepsilon \|k\| \quad \blacksquare$$

Chapitre 2 : THÉORÈME DES ACCROISSEMENTS FINIS ET APPLICATIONS DE CLASSE $\mathcal{C}^1$

31 décembre 2005

1 Théorème des accroissements finis

C'est un résultat qui concerne les fonctions d'une seule variable, même s'il a des conséquences générales. Pour les fonctions d'une variable on emploie de façon équivalente les mots dérivable et différentiable. Rappelons que si I est un ouvert de $\mathbb{R}$ et $f : I \to F$ une fonction dérivable en $t \in I$, on note $f'(t)$ le vecteur de F qui est la dérivée de f en t.

Définition 1.1 *Soit $[\alpha, \beta]$ un segment de $\mathbb{R}$ et $f : [\alpha, \beta] \to F$. On dit que f est dérivable sur $[\alpha, \beta]$ si elle l'est dans $]\alpha, \beta[$ et si $\lim_{t \to \alpha^+} \dfrac{f(t) - f(\alpha)}{t - \alpha}$ et $\lim_{t \to \beta^-} \dfrac{f(t) - f(\beta)}{t - \beta}$ existent. On note (abusivement) $f'(\alpha)$ et $f'(\beta)$ les vecteurs de F ainsi définis.*

Proposition 1.2 (Théorème des accroissements finis) *Soit $[\alpha, \beta]$ un segment de $\mathbb{R}$ et $f : [\alpha, \beta] \to F$ une application dérivable sur ce segment. Soit $\varphi : [\alpha, \beta] \to \mathbb{R}$ une fonction dérivable sur ce même segment. On suppose que, pour tout $t \in [\alpha, \beta]$, $\|f'(t)\| < \varphi'(t)$. Alors*

$$\|f(\beta) - f(\alpha)\| \leq \varphi(\beta) - \varphi(\alpha).$$

Preuve Commençons par remarquer que la fonction φ est strictement croissante sur l'intervalle $[\alpha, \beta]$ puisque sa dérivée est strictement positive, ceci permet d'être rassuré malgré l'absence de valeur absolue dans le membre de droite. Supposons $\|f(\beta) - f(\alpha)\| > \varphi(\beta) - \varphi(\alpha)$. Soit $m = \frac{\alpha+\beta}{2}$ le milieu de $[\alpha, \beta]$. L'une des deux inégalités $\|f(\beta) - f(m)\| > \varphi(\beta) - \varphi(m)$ ou $\|f(m) - f(\alpha)\| > \varphi(m) - \varphi(\alpha)$ a lieu (utiliser l'inégalité triangulaire). Soit $[\alpha_1, \beta_1]$ l'intervalle $[\alpha, m]$ ou $[m, \beta]$ sur lequel l'inégalité est vérifiée. On a donc $\|f(\beta_1) - f(\alpha_1)\| > \varphi(\beta_1) - \varphi(\alpha_1)$, $[\alpha_1, \beta_1] \subset [\alpha, \beta]$ et $\beta_1 - \alpha_1 = \frac{1}{2}(\beta - \alpha)$. On recommence avec

le segment $[\,\alpha_1, \beta_1\,]$ au lieu de $[\,\alpha, \beta\,]$. Par récurrence on construit une suite de segments emboîtés $[\,\alpha_n, \beta_n\,]$ tels que $\beta_n - \alpha_n = \frac{1}{2^n}(\beta - \alpha)$ et $\|f(\beta_n) - f(\alpha_n)\| > \varphi(\beta_n) - \varphi(\alpha_n)$. Les suites (α_n) et (β_n) sont adjacentes et convergent vers une même limite $\gamma \in [\,\alpha, \beta\,]$. Par hypothèse $\|f'(\gamma)\| < \varphi'(\gamma)$ et il est possible de choisir $\varepsilon > 0$ tel que $\|f'(\gamma)\| + 2\varepsilon < \varphi'(\gamma)$. En exprimant la dérivabilité de f et de φ en γ on peut trouver $\eta > 0$ tel que, si $t \in [\,\alpha, \beta\,]$ vérifie $|t - \gamma| < \eta$, alors

$$\|f(t) - f(\gamma) - (t - \gamma)f'(\gamma)\| < \varepsilon\,|t - \gamma|$$

et

$$|\varphi(t) - \varphi(\gamma) - (t - \gamma)\varphi'(\gamma)| < \varepsilon\,|t - \gamma|.$$

Puisque les suites (α_n) et (β_n) convergent vers γ, pour n convenable on a $|\alpha_n - \gamma| < \eta$ et $|\beta_n - \gamma| < \eta$. Alors :

$$\|f(\alpha_n) - f(\gamma) - (\alpha_n - \gamma)f'(\gamma)\| < \varepsilon\,(\gamma - \alpha_n)$$

$$\|f(\beta_n) - f(\gamma) - (\beta_n - \gamma)f'(\gamma)\| < \varepsilon\,(\beta_n - \gamma)$$

$$|\varphi(\alpha_n) - \varphi(\gamma) - (\alpha_n - \gamma)\varphi'(\gamma)| < \varepsilon\,(\gamma - \alpha_n)$$

$$|\varphi(\beta_n) - \varphi(\gamma) - (\beta_n - \gamma)\varphi'(\gamma)| < \varepsilon\,(\beta_n - \gamma).$$

On en déduit (inégalité triangulaire) :

$$\|f(\beta_n) - f(\alpha_n)\| \leq \varepsilon(\beta_n - \alpha_n) + \|(\beta_n - \alpha_n)f'(\gamma)\| = (\beta_n - \alpha_n)(\|f'(\gamma)\| + \varepsilon)$$

Vu le choix de ε, on a aussi

$$\|f(\beta_n) - f(\alpha_n)\| < (\beta_n - \alpha_n)(\varphi'(\gamma) - \varepsilon).$$

D'autre part

$$(\gamma - \alpha_n)\varphi'(\gamma) < \varphi(\gamma) - \varphi(\alpha_n) + \varepsilon(\gamma - \alpha_n)$$

et

$$(\beta_n - \gamma)\varphi'(\gamma) < \varphi(\beta_n) - \varphi(\gamma) + \varepsilon(\beta_n - \gamma)$$

d'où

$$(\beta_n - \alpha_n)\varphi'(\gamma) < \varphi(\beta_n) - \varphi(\alpha_n) + \varepsilon(\beta_n - \alpha_n)$$

En combinant avec l'inégalité précédente on obtient

$$\|f(\beta_n) - f(\alpha_n)\| < \varphi(\beta_n) - \varphi(\alpha_n)$$

ce qui contredit la définition du segment $[\,\alpha_n, \beta_n\,]$. ∎

Si E est un espace vectoriel et $a, b \in E$, on définit le *segment* $[\,a, b]$ comme l'image de l'application affine $\gamma : [\,0, 1\,] \to E$ donnée par $\gamma(t) = (1 - t)a + tb = a + t(b - a)$.

Corollaire 1.3 *Soit U un ouvert de E et $f : U \to F$ une application différentiable. Si $a, b \in U$ sont tels que $[\,a, b\,] \subset U$ et s'il existe $M > 0$ tel que, pour tout $t \in [\,0, 1\,]$, $\|Df(\gamma(t))\| \leq M$, alors $\|f(b) - f(a)\| \leq M\,\|b - a\|$.*

Preuve Soit $g = f \circ \gamma : [\,0,1\,] \to F$, c'est-à-dire $g(t) = f(a+t(b-a))$. Pour tout $t \in [\,0,1\,]$ on a
$g'(t) = D(f\circ\gamma)(t)\cdot 1 = Df(\gamma(t))\circ D\gamma(t)(1) = Df(\gamma(t))\cdot\gamma'(t) = Df(\gamma(t))\cdot(b-a)$
donc $\|g'(t)\| \leq M\,\|b-a\|$ et on applique la proposition 1.2 à g sur $[\,0,1\,]$ avec $\varphi(t) = M'\,\|b-a\|\,t$ où $M' > M$. On obtient

$$\|g(1) - g(0)\| \leq \varphi(1) - \varphi(0)$$

c'est-à-dire

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

pour tout $M' > M$ donc aussi $\|f(b) - f(a)\| \leq M\,\|b-a\|$. ∎

Rappelons qu'une partie A d'un espace vectoriel E est *convexe* si pour tous $a, b \in A$ le segment $[\,a,b\,] \subset A$. Les boules d'un espace normé sont convexes.

Corollaire 1.4 *Soit U un ouvert convexe de E et $f : U \to F$ une application différentiable. Alors, pour tout $x, y \in U$,*

$$\|f(x) - f(y)\| \leq \sup_{z\in U} \|Df(z)\|\;\|x-y\|$$

Preuve Ce résultat n'a d'intérêt que lorsque $\sup_{z\in U}\|Df(z)\|$ est fini. Dans ce cas, pour tout choix de deux points de U la convexité assure qu'on peut appliquer le corollaire précédent au segment défini par ces deux points. ∎

Le théorème des accroissements finis a la très importante conséquence suivante.

Proposition 1.5 *Soit U un ouvert* connexe *de E et $f : U \to F$ une application différentiable telle que $\forall x \in U$, $Df(x)$ est l'application nulle, alors f est constante.*

Preuve Fixons $a \in U$ et définissons $\Omega = \{x \in U \mid f(x) = f(a)\}$. Comme f est continue, Ω est un fermé de U. D'autre part si $x \in \Omega \subset U$, il existe $\rho > 0$ tel que $Bo(x,\rho) \subset U$. Une boule étant convexe, si $y \in Bo(x,\rho)$, on peut appliquer le corollaire 1.4 au segment $[\,x,y\,]$. Comme $\sup_{z\in Bo(x,\rho)}\|Df(z)\| = 0$, on en déduit $f(y) = f(x) = f(a)$ donc $y \in \Omega$. Ceci montre que Ω est aussi un ouvert. Comme U est connexe et $\Omega \neq \emptyset$ (il contient a), c'est que $\Omega = U$ et donc que f est constante sur U. ∎

2 Applications de classe $\mathcal{C}^1$

Soit U un ouvert de E et $f : U \to F$ une application différentiable dans U. On peut définir l'application $Df : U \to L(E,F)$ qui associe à x l'application linéaire continue $Df(x)$. Puisque $L(E,F)$ est aussi un espace normé, on se trouve dans la même situation qu'au point de départ et on peut parler de la continuité ou de la différentiabilité de Df en un point.

Définition 2.1 *Soit U un ouvert de E. On dit que $f : U \to F$ est de classe $\mathcal{C}^1$ (ou continûment différentiable) sur U si*
- *f est différentiable dans U,*
- *l'application Df est continue sur U.*

La deuxième propriété s'explicite en : pour tout $x \in U$ et tout $\varepsilon > 0$, il existe $\alpha > 0$ tel que pour tout $y \in U$ si $\|y-x\| < \alpha$, alors $\|Df(x)-Df(y)\| \leq \varepsilon$. Cette inégalité équivaut à la propriété : pour tout $h \in E$, $\|Df(x)\cdot h-Df(y)\cdot h\| \leq \varepsilon\|h\|$.

Remarque

Si E et F sont tous deux de dimension finie la continuité (en x) de Df équivaut à celle de l'application $x \mapsto Jf(x)$ ou encore à la continuité comme fonction de $x = (x_1, \cdots, x_m)$ de chacune des dérivées partielles $\dfrac{\partial f_i}{\partial x_j}(x)$.

Proposition 2.2 *Toute application de classe $\mathcal{C}^1$ sur un ouvert U est* localement lipshitzienne*, c'est-à-dire que, pour tout $a \in U$, il existe $\rho > 0$ et $k > 0$ tels que $\forall x, y \in Bo(a, \rho)$, on a $\|f(x) - f(y)\| \leq k\|x - y\|$.*

Preuve Noter que k et ρ dépendent de a. Soit $k > \|Df(a)\|$. Par continuité de Df, il existe une boule centrée en a sur laquelle $\|Df(x)\| \leq k$. On peut alors appliquer à cette boule, qui est convexe, le corollaire 2.1. ∎

Le plat de résistance de ce chapitre c'est le théorème suivant qui permet, en présence de continuité, de « remonter » de la différentiabilité partielle à la différentiabilité.

Théorème 2.3 *Soit U un ouvert de $E_1 \times E_2$ et $f : U \to F$. Alors f est $\mathcal{C}^1$ sur U* si et seulement si *f admet en tout point $x \in U$ une différentielle partielle dans la direction E_1 et une différentielle partielle dans la direction E_2 et, de plus, $D_1 f : U \to L(E_1, F)$ et $D_2 f : U \to L(E_2, F)$ sont continues.*

Preuve Dans le sens direct l'existence des différentielles partielles est connue. Montrons la continuité de $D_1 f$ sur U. Pour $x, a \in U$, on a

$$\begin{aligned}
\|D_1 f(x) - D_1 f(a)\| &= \sup_{h_1 \in E_1, \|h_1\|=1} \|D_1 f(x) \cdot h_1 - D_1 f(a) \cdot h_1\| \\
&= \sup_{h_1 \in E_1, \|h_1\|=1} \|Df(x) \cdot (h_1, 0) - Df(a) \cdot (h_1, 0)\| \\
&\leq \sup_{h \in E_1 \times E_2, \|h\|=1} \|Df(x) \cdot h - Df(a) \cdot h\| \\
&= \|Df(x) - Df(a)\|
\end{aligned}$$

et cette dernière quantité peut être rendue aussi petite que l'on veut en prenant $\|x - a\|$ assez petit.

Pour établir la réciproque, le plus difficile est de prouver la différentiabilité. Soit $a \in U$, Le « candidat » est connu : c'est l'application $u : E_1 \times E_2 \to F$ donnée pour $h = (h_1, h_2) \in E$ par

$$u \cdot h = D_1 f(a) \cdot h_1 + D_2 f(a) \cdot h_2$$

Clairement u est linéaire. Sa continuité est conséquence de la majoration

$$\begin{aligned}\|u\cdot h\| &\leq \|D_1f(a)\cdot h_1\| + \|D_2f(a)\cdot h_2\| \\ &\leq \|D_1f(a)\|\,\|h_1\| + \|D_2f(a)\|\,\|h_2\| \\ &\leq (\|D_1f(a)\| + \|D_2f(a)\|)\,\|h\|\end{aligned}$$

Il faut maintenant montrer que la fonction de $h=(h_1,h_2)$ définie au voisinage de 0 par

$$\varphi(h) = f(a+h) - f(a) - D_1f(a)\cdot h_1 - D_2f(a)\cdot h_2$$

vérifie $\lim\limits_{h\to 0}\dfrac{\varphi(h)}{\|h\|} = 0$. Supposons $a=(a_1,a_2)$ et posons $b=(a_1+h_1,a_2)$. On peut écrire

$$\varphi(h) = \varphi_1(h) + \varphi_2(h)$$

où

$$\varphi_1(h) = f(b) - f(a) - D_1f(a)\cdot h_1$$

et

$$\varphi_2(h) = f(a+h) - f(b) - D_2f(a)\cdot h_2.$$

Par définition de la différentielle partielle en a dans la direction E_1, on a $\lim\limits_{h_1\to 0}\dfrac{\varphi_1(h)}{\|h_1\|} = 0$. En fait φ_1 ne dépend de h que par l'intermédiaire de h_1 et $\|h_1\|\leq\|h\|$ donc, a fortiori, $\lim\limits_{h\to 0}\dfrac{\varphi_1(h)}{\|h\|} = 0$.
C'est pour étudier $\varphi_2(h)$ qu'on utilise le théorème des accroissements finis. On introduit l'application $x_2\mapsto g(x_2)$ avec

$$g(x_2) = f(a_1+h_1,x_2) - D_2f(a)\cdot x_2.$$

Elle est définie et différentiable dans un voisinage U_2 de a_2 dans E_2. De plus $Dg(x_2) = D_2f(a_1+h_1,x_2) - D_2f(a)$. La continuité de D_2f en a permet, pour tout $\varepsilon>0$, de déterminer $\eta>0$ tel que $\|x-a\|<\eta$ implique

$$\|D_2f(x) - D_2f(a)\| < \varepsilon.$$

Donc, si $x=(a_1+h_1,x_2)$ avec $\|h_1\|<\eta$ et $\|x_2-a_2\|<\eta$, on a $\|Dg(x_2)\|<\varepsilon$. En particulier, si $\|h_2\|<\eta$, cette majoration est valable pour tout $x_2\in[\,a_2,a_2+h_2\,]$ et le théorème des accroissements finis appliqué à g sur ce segment donne

$$\|g(a_2+h_2) - g(a_2)\| \leq \varepsilon\|h_2\|.$$

Or $g(a_2+h_2) = f(a+h) - D_2f(a)\cdot(a_2+h_2)$ et $g(a_2) = f(b) - D_2f(a)\cdot a_2$ de sorte que, utilisant la linéarité de $D_2f(a)$, ceci s'écrit

$$\|f(a+h) - f(b) - Df_2(a)\cdot h_2\| = \|\varphi_2(h)\| \leq \varepsilon\|h_2\| \leq \varepsilon\|h\|.$$

On a donc établi que f est différentiable dans U et que pour tout $x \in U$ et tout $h = (h_1, h_2) \in E$,

$$Df(x) \cdot h = D_1 f(x) \cdot h_1 + D_2 f(x) \cdot h_2.$$

Il reste à prouver la continuité de Df sur U. Si x et $a \in U$,

$$\begin{aligned} \|Df(x) - Df(a)\| &= \sup_{\|h\| \leq 1} \|Df(x) \cdot h - Df(a) \cdot h\| \\ &= \sup_{\|h_1\| \leq 1, \|h_2\| \leq 1} \|(D_1 f(x) - D_1 f(a)) \cdot h_1 + (D_2 f(x) - D_2 f(a)) \cdot h_2\| \\ &\leq \sup_{\|h_1\| \leq 1} \|(D_1 f(x) - D_1 f(a)) \cdot h_1\| + \sup_{\|h_2\| \leq 1} \|(D_2 f(x) - D_2 f(a)) \cdot h_2\| \\ &\leq \|D_1 f(x) - D_1 f(a)\| + \|D_2 f(x) - D_2 f(a)\| \end{aligned}$$

et cette dernière quantité tend vers 0 lorsque x tend vers a puisque les différentielles partielles sont continues en a. ∎

Voici un très utile corollaire.

Corollaire 2.4 *Soit E et F deux espaces vectoriels de dimension finie. Soit U un ouvert de E. L'application $f : U \to F$ est de classe $\mathcal{C}^1$ si et seulement si pour tout $i \in \{1, \cdots, \dim F\}$ et tout $j \in \{1, \cdots, \dim E\}$, $\dfrac{\partial f_i}{\partial x_j}$ est définie et continue dans U.*

Les points non démontrés de la proposition qui suit sont sans piège.

Proposition 2.5

1. *Toute combinaison linéaire d'applications de classe $\mathcal{C}^1$ sur un ouvert U est de classe $\mathcal{C}^1$ sur U.*
2. *La composée de deux applications de classe $\mathcal{C}^1$ est de classe $\mathcal{C}^1$.*
3. *Toute application affine continue, toute application (bi)linéaire continue est de classe $\mathcal{C}^1$.*
4. *Si f et $g : U \to \mathbb{R}$ sont de classe $\mathcal{C}^1$ sur l'ouvert U, leur produit est aussi de classe $\mathcal{C}^1$ sur U et il en est de même pour les autres produits étudiés au chapitre 1.*
5. *Si $f : U \to \mathbb{R}$ est $\mathcal{C}^1$ et si, pour tout $x \in U$, $f(x) \neq 0$, alors la fonction $\dfrac{1}{f}$ est de classe $\mathcal{C}^1$ sur U.*

Preuve

– On sait que si $f : U \to F$ est différentaible et si $g : V \to G$ est différentaible où V est un ouvert de F tel que $f(U) \subset V$, alors $\varphi = g \circ f$ est différentiable et, pour tout $x \in U$, $D\varphi(x) = Dg(f(x)) \circ Df(x)$. Cette formule peut se lire $D\varphi = B \circ \Phi$ où

$$\begin{array}{rcl} \Phi : U & \longrightarrow & L(F, G) \times L(E, F) \\ x & \mapsto & (Dg(f(x)), Df(x)) \end{array}$$

autrement dit $\Phi = (Dg \circ f, Df)$. Et B est l'application bilinéaire continue

$$\begin{array}{ccc} L(F,G) \times L(E,F) & \longrightarrow & L(E,G) \\ (u,v) & \mapsto & u \circ v \end{array}$$

Comme f est $\mathcal{C}^1$, Df est continue sur U. Comme g est $\mathcal{C}^1$, Dg est continue sur V et, puisque f est continue sur U et que $f(U) \subset V$, la composée $Dg \circ f$ est continue sur U. L'application Φ est continue sur U puisque ses deux composantes le sont. Enfin $D\varphi$ est continue comme composée de deux fonctions continues.

– On sait que si $f : U \to \mathbb{R}$ est différantiable et ne s'annule pas, la fonction $g = \dfrac{1}{f}$ est définie et différentiable en tout point de U, sa différentielle en $x \in U$ étant la forme linéaire continue sur E : $Dg(x) = -\dfrac{1}{f(x)^2} Df(x)$. Cette formule signifie que Dg est le produit de la fonction continue $-\dfrac{1}{f^2}$ et de l'application continue (puisque f est $\mathcal{C}^1$) Df, d'où la continuité de Dg sur U. ∎

Dans la suite de ce chapitre on suppose que les espaces sont des Banach.

Proposition 2.6 *Si E et F sont deux espaces de Banach, l'ensemble $GL(E,F)$ est un ouvert de $L(E,F)$.*

Preuve Le cas $GL(E,F) = \emptyset$ n'est pas exclu, mais c'est bien un ouvert. Sinon, soit $u \in GL(E,F)$ et $h \in L(E,F)$, on peut écrire, puisque u est inversible

$$u + h = u \circ (id_E + u^{-1} \circ h) = u \circ v$$

où $v = id_E + u^{-1} \circ h$ appartient à $L(E)$ (abréviation de $L(E,E)$) qui est un Banach puisque E en est un. Si $\|u^{-1} \circ h\| < 1$, ce qui a sûrement lieu si $\|h\| < r = \dfrac{1}{\|u^{-1}\|}$, la série de terme général $(-1)^n (u^{-1} \circ h)^n$ est normalement convergente donc convergente et sa somme est l'inverse de v qui appartient donc à $GL(E)$. Mais alors $u \circ v = u + h$ appartient à $GL(E,F)$. On vient de montrer que la boule ouverte de $L(E,F)$ de centre u de rayon r est incluse dans $GL(E,F)$. ∎

Remarque 2.7 *En particulier, lorsque $E = F$ est un Banach, la preuve précédente montre que la boule ouverte de centre id_E et de rayon 1 du Banach $L(E)$ est contenue dans $GL(E)$.*

Proposition 2.8 *Soit E et F deux espaces de Banach tels que $GL(E,F) \neq \emptyset$. L'application*

$$\mathcal{J} : \begin{array}{ccc} GL(E,F) & \longrightarrow & GL(F,E) \\ u & \mapsto & u^{-1} \end{array}$$

est différentiable et $D\mathcal{J}(u)$ est l'application (linéaire continue) de $L(E,F)$ dans $L(F,E)$ donnée, pour $h \in L(E,F)$, par

$$D\mathcal{J}(u)\cdot h = -u^{-1}\circ h\circ u^{-1}.$$

Preuve On reprend les notations précédentes. Pour $u \in GL(E,F)$ et $h \in L(E,F)$ avec $\|h\| < r$, on a

$$\begin{aligned}
\mathcal{J}(u+h)-\mathcal{J}(u) &= (u+h)^{-1}-u^{-1} = (u\circ v)^{-1}-u^{-1} = \\
v^{-1}\circ u^{-1}-u^{-1} &= (id_E+u^{-1}\circ h)^{-1}\circ u^{-1}-u^{-1} \\
&= \sum_{n=0}^{\infty}(-1)^n(u^{-1}\circ h)^n\circ u^{-1}-u^{-1} \\
&= \sum_{n=1}^{\infty}(-1)^n(u^{-1}\circ h)^n\circ u^{-1} \\
&= U(h)+\varphi(h)
\end{aligned}$$

où $U(h) = -u^{-1}\circ h\circ u^{-1}$ et $\varphi(h) = \sum_{n=2}^{\infty}(-1)^n(u^{-1}\circ h)^n\circ u^{-1}$, toutes ces égalités étant valables sous réserve de convergence des séries écrites.
L'application $U : L(E,F) \to L(F,E)$ ainsi définie est clairement linéaire et la majoration

$$\|U(h)\| \le \|u^{-1}\|^2\,\|h\|$$

montre qu'elle est continue (et de norme inférieure ou égale à $\|u^{-1}\|^2$).
Montrons que la série qui définit φ est absolument convergente. En effet, par récurrence sur n, on voit que $\|(u^{-1}\circ h)^n\circ u^{-1}\| \le \|u^{-1}\|^n\,\|h\|^n$, qui est le terme général d'une série (géométrique) convergente. Comme $L(E)$ est un Banach, la série définissant $\varphi(h)$ converge et les calculs précédents sont justifiés. De plus, en passant à la limite dans les sommes partielles, on a la majoration

$$\|\varphi(h)\| \le \sum_{n=2}^{\infty}\|u^{-1}\|^{n+1}\,\|h\|^n = \frac{\|h\|^2\,\|u^{-1}\|^3}{1-\|h\|\,\|u^{-1}\|}.$$

Cette dernière égalité montre que $\lim_{h\to 0}\dfrac{\varphi(h)}{\|h\|} = 0$. ∎

Corollaire 2.9 *Sous les mêmes hypothèses, l'application $\mathcal{J}$ est continue.*

Proposition 2.10 *Soit U un ouvert d'un espace de Banach E, V un ouvert d'un espace de Banach F et $f : U \to V$ un homéomorphisme de classe $\mathcal{C}^1$ tel que pour tout $x \in U$, $Df(x) \in GL(E,F)$, alors f^{-1} est $\mathcal{C}^1$ sur V.*

Preuve On sait déjà que f est différentiable en tout point de V et que, pour tout $y \in V$, $Df^{-1}(y) = Df(f^{-1}(y))^{-1}$. Cette formule peut se lire $Df^{-1} = \mathcal{J}\circ Df\circ f^{-1}$ et fait apparaître Df^{-1} comme composée de trois applications continues, d'où le résultat. ∎

Chapitre 3 : THÉORÈME D'INVERSION LOCALE ET THÉORÈME DES FONCTIONS IMPLICITES

19 février 2007

Dans tout ce chapitre les espaces considérés sont des Banach.

1 Théorème d'inversion locale

La définition suivante est naturelle.

Définition 1.1 *Soit U un ouvert de E, V un ouvert de F. Une application $f : U \to V$ est un* difféomorphisme *(resp. $\mathcal{C}^1$- difféomorphisme) si c'est une bijection telle que f et f^{-1} sont différentiables (resp. de classe $\mathcal{C}^1$).*

Remarque 1.2 *Clairement tout difféomorphisme est un homéomorphisme.*

S'il existe un difféomorphisme f d'un ouvert U de E sur un ouvert V de F, alors $GL(E,F)$ doit être non vide. En effet, puisque $f^{-1} \circ f = id_U$, en différentiant on a, si $x \in U$ et $y = f(x)$, $Df^{-1}(y) \circ Df(x) = id_E$. De même, puisque $f \circ f^{-1} = id_V$, on a $Df(x) \circ Df^{-1}(y) = id_F$. Ceci montre que $Df(x) \in GL(E,F)$. Si E est de dimension finie cela impose que F est aussi de dimension finie et que E et F sont de même dimension.

Le dernier résultat du chapitre précédent permet d'énoncer la proposition suivante.

Proposition 1.3 *Si $f : U \to V$ est un homéomorphisme différentiable (resp. de classe $\mathcal{C}^1$) sur U et tel que pour tout $x \in U$, $Df(x) \in GL(E,F)$ alors f est un difféomorphisme (resp. $\mathcal{C}^1$- difféomorphisme) de U sur V.*

Il est en général plus facile d'étudier l'inversibilité de l'application linéaire $Df(x)$ que celle de f : en dimension finie il suffit de calculer un déterminant (le jacobien de f en x).

En dimension 1 la condition $Df(x) \in GL(E,F)$ s'exprime par $f'(x) \neq 0$. C'est une condition « ouverte » : si f' est continue et si la condition est réalisée en un point, elle l'est au voisinage. L'analogue lorsque E et F sont des Banach,

en particulier en dimension finie, est le fait que $GL(E,F)$ est un *ouvert* de $L(E,F)$. Toujours en dimension 1, on peut même déduire de $f'(a) \neq 0$ et de la continuité de f' qu'au voisinage de a la fonction f est strictement monotone et réalise donc une *bijection* d'un voisinage de a sur son image. Autrement dit on peut *localement* « remonter » du fait que la différentielle est bijective au fait que la fonction l'est. Le théorème d'inversion locale permet d'obtenir cette propriété pour les Banach.

Théorème 1.4 *Soit E et F deux espaces de Banach et U un ouvert de E. Soit $f : U \to F$ une application de classe $\mathcal{C}^1$. Soit $a \in U$ tel que $Df(a) \in GL(E,F)$, alors il existe un voisinage ouvert U_1 de a et un voisinage ouvert V_1 de $f(a)$ tels que $U_1 \subset U$ et $f : U_1 \to V_1$ est un $\mathcal{C}^1$- difféomorphisme.*

Preuve Commençons par montrer qu'il suffit d'établir le théorème lorsque $E = F$, $a = f(a) = 0$ et $Df(0) = id_E$. En effet, posons

$$\tilde{f}(x) = Df(a)^{-1} \cdot [f(a+x) - f(a)].$$

On vérifie que

- l'application $\tilde{f}$ est définie sur $\widetilde{U} = \{x \in E \mid a+x \in U\}$ qui est un voisinage ouvert de 0 et $\tilde{f} : \widetilde{U} \to E$ est de classe $\mathcal{C}^1$,
- en 0, $\tilde{f}(0) = Df(a)^{-1} \cdot 0 = 0$,
- $D\tilde{f}(x) = Df(a)^{-1} \circ Df(a+x)$. En particulier $D\tilde{f}(0) = Df(a)^{-1} \circ Df(a) = id_E$.

Si on a établi le théorème pour $\tilde{f}$, on « revient » à f en remarquant que $f(a+x) = f(a) + Df(a) \cdot \tilde{f}(x)$ ou encore $f(x) = f(a) + Df(a) \circ \tilde{f}(x-a) = A \circ \tilde{f}(x-a)$ où $A : E \to F$ est la *bijection affine continue* $h \mapsto f(a) + Df(a) \cdot h$. Si donc $\tilde{f}$ est un $\mathcal{C}^1$- difféomorphisme de $\widetilde{U}_1$ (voisinage ouvert de 0) sur $\widetilde{V}_1$ (voisinage ouvert de 0), alors f est un $\mathcal{C}^1$- difféomorphisme de $U_1 = a + \widetilde{U}_1$ (voisinage ouvert de a) sur $V_1 = A(\widetilde{V}_1)$ qui est un voisinage ouvert de $A(0) = f(a)$ dans F.

On se place dans le cas particulier indiqué. Montrer que f est bijective de U_1 sur V_1 c'est montrer qu'on peut résoudre, pour y fixé dans V_1, l'équation $f(x) = y$ de façon unique si l'on impose que la solution trouvée x appartient à U_1. On le fait en montrant qu'on peut appliquer le théorème du point fixe à l'application $g_y(x) = x - f(x) + y$ dans une *boule fermée* convenable, puisqu'une boule fermée d'un Banach est un espace métrique complet.

Pour tout $y \in E$, l'application g_y est de classe $\mathcal{C}^1$ sur U et $Dg_y(x) = id_E - Df(x)$ est *indépendante de y* et nulle en $x = 0$. Donc par continuité, il existe $r > 0$ tel que, si x vérifie $\|x\| \leq r$, alors $x \in U$ et pour tout $y \in E$, $\|Dg_y(x)\| \leq \frac{1}{2}$. Le théorème des accroissements finis permet d'en déduire que pour tous $x, x' \in Bf(0,r)$ et tout $y \in E$,

$$(\star) \quad \|g_y(x) - g_y(x')\| \leq \frac{1}{2}\|x - x'\|.$$

Comme $g_y(0) = y$, on a en particulier $\|g_y(x) - y\| \leq \frac{1}{2}\|x\|$ pour tout $x \in Bf(0,r)$. Supposons que y vérifie $\|y\| < \frac{1}{2}r$. Alors, si $x \in Bf(0,r)$,

$$\|g_y(x)\| = \|g_y(x) - y\| + \|y\| \leq \frac{1}{2}\|x\| + \|y\| < r.$$

Donc g_y envoie la boule fermée $Bf(0,r)$ sur la boule ouverte $Bo(0,r)$ et, a fortiori, sur la boule fermée elle-même. L'inégalité $(\star)$ montre que g_y est contractante et a donc un unique point fixe. Ceci permet de *définir* une application $\varphi : V_1 \to E$ où $V_1 = Bo(0, \frac{1}{2}r)$ en associant à $y \in V_1$ l'unique point fixe $\varphi(y)$ de g_y dans $Bf(0,r)$. Comme ce point fixe est image par g_y d'un point de la boule fermée $Bf(0,r)$, il appartient en fait à la boule ouverte $Bo(0,r)$. On va montrer que $U_1 = \varphi(V_1)$ vérifie $U_1 = f^{-1}(V_1) \cap Bo(0,r)$ et est donc un voisinage ouvert de 0 et que les applications $f : U_1 \to V_1$ et $\varphi : V_1 \to U_1$ sont inverses l'une de l'autre. En effet :

- D'une part, si $y \in V_1$, $\varphi(y) \in Bo(0,r)$ et vérifie $g_y(\varphi(y)) = \varphi(y) = \varphi(y) - f(\varphi(y)) + y$. D'où $(*)$ $f(\varphi(y)) = y$. Si donc $x \in U_1$, de sorte que $x = \varphi(y)$ où $y \in V_1$, on a $x \in Bo(0,r)$ et $f(x) = y$, ce qui prouve l'inclusion $U_1 \subset f^{-1}(V_1) \cap Bo(0,r)$ et permet d'interpréter $(*)$ en $f \circ \varphi = id_{V_1}$.
- D'autre part, si $x \in f^{-1}(V_1) \cap Bo(0,r)$, alors $\|x\| < r$ et $f(x) \in V_1$. Mais $g_{f(x)}(x) = x - f(x) + f(x) = x$, ce qui montre que x est point fixe dans $Bf(0,r)$ de $g_{f(x)}$. Donc par unicité on a $(**)$ $x = \varphi(f(x))$. Ceci montre que $x \in U_1$, ce qui prouve l'inclusion $f^{-1}(V_1) \cap Bo(0,r) \subset U_1$ et permet d'interpréter $(**)$ en $\varphi \circ f = id_{U_1}$.

Pour prouver que φ est de classe $\mathcal{C}^1$ sur V_1, il suffit de montrer qu'elle est *continue*. En effet, pour $x \in U_1$, $\|id_E - Df(x)\| \leq \frac{1}{2} < 1$ et la remarque **2.7** du chapitre 2 montre que $Df(x) \in GL(E)$. On peut alors appliquer à l'homéomorphisme $f : U_1 \to V_1$ la proposition 1.3.
Pour établir la continuité de φ, nous allons montrer qu'elle est lipshitzienne de rapport 2 sur V_1. Si $y, y' \in V_1$, alors $x = \varphi(y)$ et $x' = \varphi(y')$ appartiennent à $Bo(0,r)$ et, pour $y_0 \in E$ fixé (quelconque), en utilisant $(\star)$, on a :

$$\|x - x' - f(x) + f(x')\| = \|g_{y_0}(x) - g_{y_0}(x')\| \leq \frac{1}{2}\|x - x'\|$$

d'où $\|x - x'\| \leq \|f(x) - f(x')\| + \frac{1}{2}\|x - x'\|$ et $\|x - x'\| \leq 2\,\|f(x) - f(x')\|$ ou encore

$$\|\varphi(y) - \varphi(y')\| \leq 2\,\|y - y'\|. \quad \blacksquare$$

On adopte parfois la terminologie suivante, utile lorsqu'on veut privilégier l'espace source.

Définition 1.5 *Soit $f : U \to F$ une application de classe $\mathcal{C}^1$ sur un ouvert d'un Banach E. On dit que f est un $\mathcal{C}^1$- difféomorphisme* local *au voisinage de $a \in U$ s'il existe un voisinage ouvert U_a de a et un ouvert V de F tel que la restriction de f à U_a soit un $\mathcal{C}^1$- difféomorphisme de U_a sur V.*

Le théorème d'inversion locale peut alors s'énoncer comme suit :

Théorème 1.6 *Soit E et F deux espaces de Banach et U un ouvert de E. Une application $f : U \to F$ de classe $\mathcal{C}^1$ est un $\mathcal{C}^1$- difféomorphisme local au voisinage de $a \in U$ si et seulement si $Df(a) \in GL(E,F)$.*

Lorsque $E = F$ est de dimension finie, un $\mathcal{C}^1$- difféomorphisme local s'appelle aussi un *changement de coordonnées.*
Le corollaire suivant explicite *en dimension finie* une généralisation locale $\mathcal{C}^1$ du théorème de la base incomplète de l'algèbre linéaire.

Corollaire 1.7 *Soit U un ouvert de $\mathbb{R}^n$, p un entier tel que $1 \leq p \leq n$ et $f_1, \cdots, f_p$ des fonctions numériques de classe $\mathcal{C}^1$. Soit $a \in U$, les deux propriétés suivantes sont équivalentes :*

- *les p formes linéaires $Df_1(a), \cdots, Df_p(a)$ sont linéairement indépendantes,*
- *il existe $n-p$ fonctions numériques $f_{p+1}, \cdots, f_n$ de classe $\mathcal{C}^1$ sur U telles que $f = (f_1, \cdots, f_n) : U \to \mathbb{R}^n$ soit un $\mathcal{C}^1$- difféomorphisme local au voisinage de a.*

Preuve Si $p = n$, c'est une simple reformulation du théorème d'inversion locale si on se souvient qu'une matrice est inversible si et seulement si les n formes linéaires définies par ses lignes sont indépendantes. Si $p < n$ et si les p vecteurs $Df_1(a), \cdots, Df_p(a)$ du dual de $\mathbb{R}^n$ sont indépendants, le théorème de la base incomplète montre qu'il existe $n-p$ formes linéaires sur $\mathbb{R}^n$, $u_{p+1}, \cdots, u_n$ telles que $(Df_1(a), \cdots, Df_p(a), u_{p+1}, \cdots, u_n)$ soit une base du dual de $\mathbb{R}^n$. Définisssons $f : U \to \mathbb{R}^n$ par $f = (f_1, \cdots, f_n)$ en posant, pour $j \geq p+1$, $f_j = u_{j_{|U}}$. La fonction numérique f_j $(p+1 \leq j \leq n)$ est de classe $\mathcal{C}^1$ et vérifie, pour tout $x \in U, Df_j(x) = u_j$. On est ainsi ramené au cas $n = p$.
La réciproque est claire. ∎

Remarques

1. Dans le théorème d'inversion locale, l'hypothèse de **continuité** de Df est indispensable. On pourrait croire qu'en demandant seulement à f d'être différentiable, on aurait le même résultat, sans la continuité de l'application inverse. L'exemple suivant prouve qu'il n'en est rien, même en dimension finie. Soit $f : \mathbb{R} \to \mathbb{R}$ définie par $f(x) = \frac{1}{2}x + x^2 \sin \frac{1}{x}$ lorsque $x \neq 0$ et $f(0) = 0$. Un dessin permet de se convaincre que f n'est injective sur aucun intervalle contenant 0. Pourtant f et dérivable en tout point et $f'(0) = \lim_{x \to 0} \frac{f(x)}{x} = \frac{1}{2} \neq 0$. Mais pour $x \neq 0$, $f'(x) = \frac{1}{2} + 2x \sin \frac{1}{x} - \cos \frac{1}{x}$, qui n'a pas de limite quand $x \to 0$.
2. On ne saurait trop insister sur le fait qu'il s'agit d'un résultat **local**. Même si f, de classe $\mathcal{C}^1$, vérifie $Df(x) \in GL(E, F)$ pour tout $x \in U$, on ne peut pas en déduire que f est une bijection de U sur son image. Voici un contre-exemple. L'application $f : U =]\,0, +\infty\,[\times \mathbb{R} \to \mathbb{R}^2$ définie par $f(\rho, \theta) = (\rho \cos\theta, \rho \sin\theta)$ est de classe $\mathcal{C}^1$ et en tout point de U le jacobien vaut $\det Jf(\rho, \theta) = \rho \neq 0$. Mais la périodicité en θ montre que f n'est pas injective sur U. Le théorème affirme que si (ρ_0, θ_0) est fixé dans U, il existe $U_1 \subset U$, $(U_1 \ni (\rho_0, \theta_0))$, tel que $f_{|U_1}$ soit injective. On peut en effet prendre $U_1 =]\,0, +\infty\,[\times]\,\theta_0 - \alpha, \theta_0 - \alpha + 2\pi\,[$, avec $0 < \alpha < 2\pi$. Un examen scrupuleux de la preuve montre que l'ouvert qui y est construit est plus petit que celui-ci.

Comme le suggère l'exemple qui vient d'être traité, c'est l'**injectivité** de f qui pose problème. Les deux résultats qui suivent sont d'une grande importance pratique. Le deuxième explicite comment passer du résultat local à un résultat global.

Proposition 1.8 *Si $f : U \to F$ de classe $\mathcal{C}^1$ est telle que pour tout $x \in U$, $Df(x) \in GL(E,F)$, alors $f(U)$ est un ouvert de F et, plus généralement, l'image par f de tout ouvert de E inclus dans U est un ouvert de F.*

Preuve Soit $y_0 \in f(U)$. Il existe $x_0 \in U$ tel que $y_0 = f(x_0)$. Le théorème appliqué au point x_0 donne un ouvert $U_1 \subset U$, $U_1 \ni x_0$, et un ouvert $V_1 \ni y_0$ tels que f de U_1 sur V_1 soit bijective. En particulier, pour tout $y \in V_1$, il existe $x \in U_1 \subset U$ tel que $y = f(x)$. Autrement dit $V_1 \subset f(U)$. Ceci montre que $f(U)$ est un ouvert de F. Bien sûr l'argument s'applique à tout ouvert de E inclus dans U. ∎

Proposition 1.9 *Soit $f : U \to F$ une application de classe $\mathcal{C}^1$. Pour que f soit un $\mathcal{C}^1$- difféomorphisme de U sur un ouvert de F il faut et suffit que*

1. *f soit injective*
2. *pour tout $x \in U$, $Df(x) \in GL(E,F)$*

Preuve Les conditions *1. et 2.* sont clairement nécessaires. Pour établir la réciproque, notons que la proposition 1.8 montre que l'image $V = f(U)$ est un ouvert de F et l'hypothèse *1.* permet d'affirmer que f est une bijection de U sur son image V. Il reste à montrer que l'application $f^{-1} : V \to U$ ainsi définie est continue. Les autres affirmations seront en effet cette fois encore conséquence de la proposition 1.3. Or la continuité de f^{-1} équivaut au fait que f est ouverte et la proposition 1.8 assure cette propriété. ∎

2 Le théorème des fonctions implicites

Le théorème d'inversion locale peut s'interpréter comme la possibilité de résoudre *localement* pour chaque valeur de y de façon unique l'équation en x : $f(x) - y = 0$ et d'affirmer la régularité de la *fonction* $x(y)$ ainsi définie. La condition à satisfaire, en plus d'une « régularité » convenable de f, est l'inversibilité de l'application linéaire que constitue la différentielle, ce que l'on peut interpréter comme la possibilité de résoudre *le même problème* dans le cas *linéaire*. Plus généralement le théorème des fonctions implicites va donner des conditions permettant de résoudre $f(x,y) = 0$ par une fonction $y(x)$ convenable (on a échangé les lettres x et y pour se conformer à l'habitude). Cette fois aussi l'idée directrice est que ce qu'on sait faire dans le cadre linéaire pourra être fait *localement* dans le cadre $\mathcal{C}^1$.

On sait que, si dans un système *linéaire* de n équations à $n+m$ inconnues un déterminant d'ordre n extrait de la matrice du système est non nul, on peut choisir convenablement n des inconnues et les exprimer comme *fonctions linéaires* des m autres inconnues. Si la matrice de ce système est la matrice

jacobienne en un point d'une application d'un ouvert de $\mathbb{R}^{n+m}$ ($n+m$ est le nombre de variables) dans $\mathbb{R}^n$ (n est le nombre d'équations), la condition précédente s'interprète comme la possibilité, pour un ordre convenable des variables, de décomposer $\mathbb{R}^{n+m}$ en $\mathbb{R}^m \times \mathbb{R}^n$ de sorte que la matrice jacobienne de la *différentielle partielle* dans la direction du facteur $\mathbb{R}^n$ (deuxième facteur) soit inversible. Le théorème d'inversion locale peut être traduit dans ce langage. Reprenons en effet les hypothèses du théorème 1.4 et introduisons l'application $\varphi : F \times U \to F$ définie par $\varphi(y,z) = f(z) - y$. On a, pour tout $y \in F$, $D_2\varphi(y,a) = Df(a)$. L'hypothèse peut donc se lire $D_2\varphi(b,a) \in GL(E,F)$ avec $(b,a) = (f(a),a)$, point où la relation $\varphi(b,a) = 0$ est réalisée. La conclusion du théorème peut se lire : il existe un voisinage ouvert U_1 de a, un voisinage ouvert V de b et une application de classe $\mathcal{C}^1$, $g : V \to U_1$ telle que pour tout $y \in V$, $z = g(y)$ est l'unique solution *dans* U_1 de $\varphi(y,z) = 0$. Autrement dit on peut résoudre *localement* l'équation $\varphi(y,z) = 0$ par une application de classe $\mathcal{C}^1$, $z = g(y)$. Le théorème suivant montre que cette situation est générale.

Théorème 2.1 *Soit E, F, G trois espaces de Banach, Ω un ouvert de $E \times F$ et $f : \Omega \to G$ une application de classe $\mathcal{C}^1$. Soit $(a,b) \in \Omega$ tel que $D_2f(a,b) \in GL(F,G)$. Alors il existe un ouvert U de E contenant a, un ouvert V de F contenant b et une application $g : U \to V$ tels que*

1. *$U \times V \subset \Omega$*
2. *pour tout $x \in U$, l'équation $f(x,y) = f(a,b)$ admet $y = g(x)$ pour unique solution dans V.*

De plus l'application g est de classe $\mathcal{C}^1$ dans U et, si U est assez petit, pour tout $x \in U$,

$$Dg(x) = -D_2f(x,g(x))^{-1} \circ D_1f(x,g(x)).$$

Remarques

1. L'hypothèse implique que F et G ont même dimension (finie ou non).
2. Il est important de bien comprendre l'affirmation *2.* qui *ne* dit *pas* que l'équation $f(x,y) = f(a,b)$ n'a qu'une solution : elle n'en a qu'une dans V (c'est-à-dire proche de b) mais peut en avoir ailleurs.
3. La propriété *2.* peut aussi s'exprimer par l'égalité :

$$\{(x,y) \in U \times V \mid f(x,y) = f(a,b)\} = \{(x,g(x)) \mid x \in U\} = \Gamma_g$$

en désignant par Γ_g le *graphe* de la fonction g.

Preuve Le préambule du théorème laisse deviner que sa preuve n'est pas très différente de celle du théorème d'inversion locale. En effet on va déduire ce résultat du précédent. On aurait pu aussi prouver d'abord ce théorème et en déduire, comme cas particulier, le théorème 1.4.

Définissons l'application $\Phi : \Omega \to E \times G$ par $\Phi(x,y) = (x, f(x,y))$ et montrons qu'on peut lui appliquer le théorème d'inversion locale au point (a,b). Elle

est clairement de classe $\mathcal{C}^1$ sur Ω et sa différentielle au point (a,b) est donnée, pour $(h,k) \in E \times F$, par

$$D\Phi(a,b)\cdot(h,k) = (h, Df(a,b)\cdot(h,k)) = (h, D_1f(a,b)\cdot h + D_2f(a,b)\cdot k).$$

On résout « à la main » l'équation $D\Phi(a,b)\cdot(h,k) = (h',k')$. On trouve $h = h'$ puis $D_2f(a,b)\cdot k = k' - D_1f(a,b)\cdot h'$ et donc

$$k = D_2f(a,b)^{-1}\cdot[k' - D_1f(a,b)\cdot h'].$$

Autrement dit l'application $D\Phi(a,b)$ est inversible, son inverse étant

$$D\Phi(a,b)^{-1}\cdot(h',k') = (h', D_2f(a,b)^{-1}\cdot[k' - D_1f(a,b)\cdot h']).$$

Il faut aussi montrer la continuité de l'application linéaire $D\Phi(a,b)^{-1}$. Examinons la deuxième composante. On a

$$D_2f(a,b)^{-1}\cdot[k' - D_1f(ab)\cdot h'] = D_2f(a,b)^{-1}\cdot k' - D_2f(a,b)^{-1}\circ D_1f(a,b)(h')$$

d'où

$$\begin{aligned}\|D_2f(a,b)^{-1}\cdot[k' - D_1f(ab)\cdot h']\| &\leq \|D_2f(a,b)^{-1}\|(\|k'\| + \|D_1f(a,b)\|\,\|h'\|) \\ &\leq \underbrace{\|D_2f(a,b)^{-1}\|(1+\|D_1f(a,b)\|)}_{M}\,\|(h',k')\|\end{aligned}$$

et donc $\|D\Phi(a,b)^{-1}\cdot(h',k')\| \leq \max(1,M)\|(h',k')\|$. On a ainsi montré que $D\Phi(a,b)$ appartient à $GL(E\times F, E\times G)$.

On a $\Phi(a,b) = (a, f(a,b))$ et le théorème d'inversion locale appliqué à Φ au point (a,b) donne un voisinage ouvert Ω_1 de (a,b), $\Omega_1 \subset \Omega$ et un voisinage ouvert W de $(a,f(a,b))$ tels que $\Phi : \Omega_1 \to W$ soit un $\mathcal{C}^1$- difféomorphisme. Quitte à diminuer de manière cohérente Ω_1 et W, on peut supposer que $\Omega_1 = U_1 \times V$ où U_1 est un voisinage ouvert de a dans E et V un voisinage ouvert de b dans F. L'application $\Phi^{-1} : W \to U_1 \times V$ a alors deux composantes : $\Phi^{-1} = (g_1, g_2)$ où $g_1 : W \to U_1$ et $g_2 : W \to V$ sont de classe $\mathcal{C}^1$. On a alors la chaîne d'équivalences :

$$\begin{array}{c}(x,y) \in U_1\times V \quad \text{et} \quad f(x,y) = f(a,b) \\ \Updownarrow \\ (x,y) \in U_1\times V \quad \text{et} \quad \Phi(x,y) = (x, f(a,b)) \\ \Updownarrow \\ (x,y) = (g_1(x,f(a,b)), g_2(x,f(a,b))) \quad \text{et} \quad (x,f(a,b)) \in W\end{array}$$

Considérons $U = \{x \in U_1 \mid (x,f(a,b)) \in W\}$. C'est un *ouvert* de U_1 (puisque W est ouvert), donc de E, contenant a et tel que $U \times V \subset U_1 \times V \subset \Omega$. De plus la chaîne d'équivalence indiquée montre que l'application $g : U \to V$ donnée par $g(x) = g_2(x, f(a,b))$ vérifie

$$\begin{array}{c}(x,y) \in U\times V \quad \text{et} \quad f(x,y) = f(a,b) \\ \Updownarrow \\ x \in U \quad \text{et} \quad y = g(x).\end{array}$$

La définition de g montre qu'elle est de classe $\mathcal{C}^1$ dans U.
Enfin de la relation $f(x, g(x)) = f(a, b)$ valable pour tout $x \in U$, on tire par différentiation que, pour tout $h \in E$,

$$D_1 f(x, g(x)) \cdot h + D_2 f(x, g(x)) \circ Dg(x)(h) = 0.$$

Si U est assez petit pour qu'en tout point $D_1 f(x, g(x))$ (qui dépend continûment de x) appartienne à l'ouvert $GL(F, G)$, on en déduit

$$Dg(x) = -D_2 f(x, g(x))^{-1} \circ D_1 f(x, g(x)). \quad \blacksquare$$

Insistons sur l'importance de se familiariser avec les diverses façons d'utiliser ce théorème. Il est souvent nécessaire de découvrir la bonne décomposition de E en produit de deux facteurs. Cette décomposition peut dépendre du point au voisinage duquel on souhaite faire l'étude. Par exemple prenons $U = E = \mathbb{R}^2$, $F = \mathbb{R}$ et $f(x, y) = x^2 + y^2 - R^2$, où $R > 0$, de sorte que $f(x, y) = 0$ est l'*équation cartésienne* du cercle (C) de centre 0 et de rayon R. Si $A = (a, b)$ est un point de ce cercle alors :

- si A n'est sur aucun des deux axes de coordonnées ($ab \neq 0$), on peut décrire les points de (C) voisins de A, soit en exprimant y comme fonction de x, soit en exprimant x comme fonction de y. L'écriture de cette fonction dépend de A. On a $y(x) = \sqrt{R^2 - x^2}$ si $b > 0$, $y(x) = -\sqrt{R^2 - x^2}$ si $b < 0$. On a aussi $x(y) = \sqrt{R^2 - y^2}$ si $a > 0$ et $x(y) = -\sqrt{R^2 - y^2}$ si $a < 0$. Dans tous les cas la fonction est de classe $\mathcal{C}^1$ sur l'ouvert $]-R, +R[$.
  On remarque que $\dfrac{\partial f}{\partial y}(a, b) = 2b \neq 0$ et que $\dfrac{\partial f}{\partial x}(a, b) = 2a \neq 0$ et qu'on peut voir l'un ou l'autre résultat comme une application du théorème 2.1.
- Si $A = (0, \pm R)$, seule l'écriture de y comme fonction de x est possible au voisinage de A, en conformité avec le fait que $\dfrac{\partial f}{\partial y}(A) \neq 0$ mais que $\dfrac{\partial f}{\partial x}(A) = 0$. Les conclusions sont analogues (en échangeant x et y) aux points $(\pm R, 0)$.

Remarquons que *tous* les points du cercle sont susceptibles d'être étudiés de l'une ou l'autre façon.

3 Applications géométriques en dimension finie

Nous donnons dans ce paragraphe des applications géométriques des théorèmes 1.4 et 2.1 en supposant que les espaces considérés sont de *dimension finie*. Toutes les applications seront supposées $\mathcal{C}^1$ et nous ne l'indiquerons pas systématiquement.

3.1 Sous – variétés de $\mathbb{R}^n$

Sans faire un cours de géométrie différentielle nous voulons illustrer l'usage que l'on peut faire des notions de calcul différentiel introduites pour généraliser en toute dimension les objets géométriques auxquels on est habitué en dimension 2 et 3.

Cette fois encore l'idée directrice est que la situation linéaire modélise *localement* la situation $\mathcal{C}^1$. Dans le modèle linéaire, un sous-espace vectoriel de dimension p ($1 \leq p \leq n-1$) de $\mathbb{R}^n$ c'est, à isomorphisme près, $\mathbb{R}^p \times \{0\}$ vu comme sous-espace de $\mathbb{R}^n$. (Bien sûr, 0 désigne ici le vecteur nul de $\mathbb{R}^{n-p}$ et on a decomposé $\mathbb{R}^n$ en $\mathbb{R}^p \times \mathbb{R}^{n-p}$). La définition suivante met en œuvre le principe rappelé.

Définition 3.1 *Soit p un entier tel que $1 \leq p \leq n-1$ et $M \subset \mathbb{R}^n$. On dit que M est une* sous-variété *de $\mathbb{R}^n$ de dimension p et de codimension $n-p$, si pour tout $x \in M$, il existe un voisinage ouvert U de x, un voisinage ouvert V de $0 \in \mathbb{R}^n$ et un $\mathcal{C}^1$- difféomorphisme $f : U \to V$ tels que*

$$f(M \cap U) = V \cap (\mathbb{R}^p \times \{0\}).$$

Dans cette définition, U, V et f dépendent (en général) de x.

Rappelons qu'un *arc de courbe* de $\mathbb{R}^2$ ou $\mathbb{R}^3$ est le graphe Γ_f d'une fonction $f : I \to \mathbb{R}$ ou $\mathbb{R}^2$, où I est un ouvert (souvent un intervalle) de $\mathbb{R}$ et qu'un *morceau de surface* de $\mathbb{R}^3$ est le graphe d'une fonction $f : U \to \mathbb{R}$ où U est un ouvert de $\mathbb{R}^2$. Plus généralement l'exemple le plus simple de sous-variétés est fourni par les graphes.

Proposition 3.2 (graphes) *Soit Ω un ouvert de $\mathbb{R}^p$ et $g : \Omega \to \mathbb{R}^{n-p}$ ($1 \leq p \leq n-1$) une application de classe $\mathcal{C}^1$. Le graphe de g, $\Gamma_g = \{(u, g(u)) \mid u \in \Omega\}$ est une sous-variété de dimension p de $\mathbb{R}^n$.*

Preuve Définissons

$$\begin{array}{rccc} f : & U = \Omega \times \mathbb{R}^{n-p} & \longrightarrow & \mathbb{R}^n \\ & x = (u, y) & \mapsto & (u, y - g(u)). \end{array}$$

C'est une application de classe $\mathcal{C}^1$(ses deux composantes le sont) et en tout point $x \in U$, la matrice

$$Jf(x) = \begin{pmatrix} I_p & 0 \\ -Jg(u) & I_{n-p} \end{pmatrix}$$

est inversible. D'autre part l'application f est injective puisque $(u, y - g(u)) = (u', y' - g(u'))$ impose $u' = u$, d'où $g(u) = g(u')$ et donc aussi $y = y'$. La proposition 1.9 montre que f est un $\mathcal{C}^1$- difféomorphisme de U sur l'*ouvert* de $\mathbb{R}^n$, $V = f(U)$. De plus, si $x \in U \cap \Gamma_g$, $x = (u, g(u))$ avec $u \in \Omega$. Donc $f(x) = (u, g(u) - g(u)) = (u, 0)$ appartient à $V \cap (\mathbb{R}^p \times \{0\})$. Réciproquement, si $z \in V \cap (\mathbb{R}^p \times \{0\})$, $z = (z', 0)$ et il existe $x = (u, y) \in U$ tel que $z = f(x) = (u, y - g(y))$. On en déduit $z' = u \in \Omega$ et $y = g(u)$ de sorte que $z = f(u, y)$ où $(u, y) \in U \cap \Gamma_g$. ∎

L'exemple du cercle (C) traité au paragraphe précédent n'entre pas dans cette catégorie, pourtant le théorème des fonctions implicites montre que *localement* au voisinage de tout point de (C), les points de (C) sont ceux d'un graphe. Le résultat qui suit montre que cette condition suffit à faire de (C) une sous-variété de dimension 1 de $\mathbb{R}^2$ et donne, plus généralement une condition sous laquelle une sous-variété de dimension p peut être donnée par $n - p$ équations.

Proposition 3.3 (équations cartésiennes) *Soit Ω un ouvert de $\mathbb{R}^n$ et $F : \Omega \to \mathbb{R}^{n-p}$ $(1 \leq p \leq n-1)$ une application de classe $\mathcal{C}^1$. On pose*

$$M = F^{-1}(0) = \{x \in \Omega \mid F(x) = 0\}.$$

Si pour tout $x \in M$, le rang de la matrice jacobienne en x, $JF(x)$, est $n - p$, M est une sous-variété de dimension p de $\mathbb{R}^n$.

Plus explicitement, si $F = (F_1, \cdots, F_{n-p})$, où $F_j : \Omega \to \mathbb{R}$, M est définie par les $n - p$ équations $F_j(x_1, \cdots, x_n) = 0$. Le nombre d'équations est donc la codimension.
La condition sur le rang de $JF(x)$ exprime que ce rang est *maximum*. Il est égal à la dimension de l'espace but et cela signifie donc que, pour tout $x \in M$, l'application linéaire $Df(x)$ est *surjective*. On dit aussi que F est une *submersion* en x.

Preuve Soit $a \in M$. L'hypothèse signifie que les $n-p$ formes linéaires $DF_1(a), \cdots, DF_{n-p}(a)$ sont linéairement indépendantes. Le corollaire 1.7 appliqué aux $n-p$ fonctions F_j permet de définir un voisinage U de a et p fonctions $G_1 \cdots, , G_p$ telles que $f = (G_1 \cdots, G_p, F_1, \cdots F_{n-p})$ soit un $\mathcal{C}^1$- difféomorphisme de U sur un ouvert V de $\mathbb{R}^n$. Quitte à remplacer G_j par $G_j - G_j(a)$, on peut supposer que V est un voisinage de 0. De plus, si $x \in U$, $f(x) \in V \cap (\mathbb{R}^p \times \{0\})$ équivaut à $F_j(x) = 0$ pour $1 \leq j \leq n - p$, ou encore à $x \in M$. ∎

Une autre manière de définir une sous-variété est d'utiliser un paramétrage.

Proposition 3.4 (sous-variété paramétrée) *Soit $1 \leq p \leq n-1$, Ω un ouvert de $\mathbb{R}^p$ et $\varphi : \Omega \to \mathbb{R}^n$ de classe $\mathcal{C}^1$. Si en $u_0 \in \Omega$, le rang de $J\varphi(u_0)$ est p, il existe un voisinage ouvert Ω_0 de u_0 tel que $\Omega_0 \subset \Omega$ et que $M = \varphi(\Omega_0)$ soit une sous-variété de dimension p de $\mathbb{R}^n$. On dit que φ est un paramétrage de M.*

Cette fois encore la condition sur le rang signifie qu'il est maximum. La condition est aussi équivalente au fait que $D\varphi(u_0)$ est injective. On dit alors que φ est une *immersion* en u_0.
Mais attention, même si φ est une immersion en tout point de Ω il n'est pas toujours vrai que son image est une sous-variété de $\mathbb{R}^n$.

Preuve L'hypothèse signifie que p des n lignes de la matrice $J\varphi(u_0)$ sont linéairement indépendantes. Un choix convenable de l'ordre des coordonnées de $\mathbb{R}^n$ permet de supposer que ce sont les p premières composantes, c'est-à-dire d'écrire $\varphi = (g, h)$, où $g : \Omega \to \mathbb{R}^p$ et $h : \Omega \to \mathbb{R}^{n-p}$ sont telles que

$$J\varphi(u_0) = \begin{pmatrix} Jg(u_0) \\ Jh(u_0) \end{pmatrix},$$

avec une matrice $Jg(u_0)$ inversible. Le théorème d'inversion locale appliqué à g au point u_0 donne un voisinage ouvert $\Omega_0 \subset \Omega$ de u_0 et un voisinage ouvert V (dans $\mathbb{R}^p$) de $a' = g(u_0)$ tels que $g : \Omega_0 \to V$ soit un $\mathcal{C}^1$- difféomorphisme. Si $x \in \mathbb{R}^n$, on l'écrit $x = (x', x'')$ où $x' \in \mathbb{R}^p$ et $x'' \in \mathbb{R}^{n-p}$ et on pose $f = h \circ g^{-1}$. C'est une application de classe $\mathcal{C}^1$ de V dans $\mathbb{R}^{n-p}$ et

$$\varphi(\Omega_0) = \{(g(u), h(u)) \mid u \in \Omega_0\} = \{(x', h(g^{-1}(x')) \mid x' \in V\}$$

$$= \{(x', f(x') \mid x' \in V\} = \Gamma_f.$$

La proposition 3.2 montre alors le résultat. ∎

On peut en fait montrer l'équivalence, en un sens à préciser, de ces diverses définitions d'une sous-variété de $\mathbb{R}^n$.

Exemples

Nous donnons quelques exemples et contre-exemples tous très classiques.

1. La sphère S^n est définie par

$$S^n = \{x = (x_0, \cdots, x_n) \in \mathbb{R}^{n+1} \mid x_0^2 + \cdots + x_n^2 - 1 = 0\}$$

est une sous-variété de dimension n de $\mathbb{R}^{n+1}$. En effet la fonction $F : \mathbb{R}^{n+1} \to \mathbb{R}$, $F(x) = x_0^2 + \cdots + x_n^2 - 1$ est $\mathcal{C}^1$ et est une submersion en tout point de S^n, puisque $JF(x) = (2x_0, \cdots, 2x_n)$ n'est jamais nulle sur S^n et est donc de rang 1 en tout point de S^n.

2. Le tore T^n est défini par

$$T^n = \{x \in \mathbb{R}^{2n} \mid x_1^2 + x_2^2 - 1 = x_3^2 + x_4^2 - 1 = \cdots = x_{2n-1}^2 + x_{2n}^2 - 1 = 0\}.$$

C'est une sous-variété de dimension n $(= 2n - n)$ de $\mathbb{R}^{2n}$. On peut le voir en appliquant la proposition 3.3 (exo.) ou en utilisant la paramétrisation $\varphi(t_1, \cdots, t_n) = (\cos t_1, \sin t_1, \cdots, \cos t_n, \sin t_n)$ et la proposition 3.4.

3. Le groupe orthogonal

$$O(n) = \{A \in \mathcal{M}_n(\mathbb{R}) \mid A^t A = I_n\}$$

où l'on a noté $\mathcal{M}_n(\mathbb{R})$ l'espace vectoriel, isomorphe à R^{n^2}, des matrices carrées $n \times n$ et A^t la matrice symétrique de la matrice A. Alors $O(n)$ est une sous-variété de $\mathcal{M}_n(\mathbb{R})$ de dimension $\frac{n(n-1)}{2}$. En effet appelons $\mathcal{S}_n(\mathbb{R})$ l'espace vectoriel des matrices $n \times n$ symétriques. Il est de dimension $\frac{n(n+1)}{2}$ et $O(n) = F^{-1}(0)$ où $F : \mathcal{M}_n(\mathbb{R}) \longrightarrow \mathcal{S}_n(\mathbb{R})$ est l'application $F(A) = A^t A - I_n$. Cette application est polynomiale donc $\mathcal{C}^1$ et c'est une submersion en chaque point $A \in O(n)$ puisque (exo.), pour tout $H \in \mathcal{M}_n(\mathbb{R})$, $DF(A) \cdot H = A^t H + H^t A$ et, si $K \in \mathcal{S}_n(\mathbb{R})$, la matrice $H = \frac{1}{2} AK$ vérifie $DF(A) \cdot H = K$, ce qui prouve la surjectivité de $DF(A)$. Il reste à remarquer que, de façon générale, si M est une sous-variété de $\mathbb{R}^n$ de dimension p et si on regarde M comme un sous-ensemble de $\mathbb{R}^m$ pour $m > n$ via une inclusion de $\mathbb{R}^n$ dans $\mathbb{R}^m$, alors M est aussi une sous-variété de dimension p de $\mathbb{R}^m$.

4. Le cône de révolution

$$\Gamma = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 - z^2 = 0\}$$

n'est pas une sous-variété de $\mathbb{R}^3$ de dimension 2. Il est facile de voir que le cône privé de son sommet $\Gamma \setminus \{0\}$ est une sous-variété de dimension 2 (surface) de $\mathbb{R}^3$. Pour voir que Γ n'en est pas une, il ne suffit pas de dire que les propositions énoncées ne s'appliquent pas... L'argument le plus simple (?) est de remarquer que, s'il existait deux voisinages ouverts U et V de $0 \in \mathbb{R}^3$ et un $\mathcal{C}^1$- difféomorphisme $f : U \to V$ tel que $f(U \cap \Gamma) = V \cap (\mathbb{R}^2 \times \{0\})$ et $f(0) = 0$, il en existerait aussi vérifiant la même propriété et connexes. Or $U \cap \Gamma \setminus \{0\}$ a (comme $\Gamma \setminus \{0\}$) deux composantes connexes alors qu'un voisinage de 0 dans $\mathbb{R}^2$ privé de 0 est connexe.

3.2 Espace tangent à une sous-variété en un point

Soit $M \subset \mathbb{R}^n$ une sous-variété de dimension p et $a \in M$.

Définition 3.5 *1. Un* arc de courbe tracé sur M passant par a *est l'image* $\gamma(I)$ *de* $\gamma : I \to \mathbb{R}^n$ *où* I *est un intervalle ouvert de* $\mathbb{R}$ *contenant* 0, γ *est de classe* $\mathcal{C}^1$, $\gamma(0) = a$ *et* $\gamma(I) \subset M$.

2. Le vecteur $\gamma'(0)$, *qui appartient à* $\mathbb{R}^n$, *est appelé* vecteur tangent en a à M.

La définition donnée ci-dessus d'un vecteur tangent est « raisonnable » car si J est aussi un intervalle de $\mathbb{R}$ contenant 0 et si $\varphi : J \to I$ est un $\mathcal{C}^1$-difféomorphisme tel que $\varphi(0) = 0$, de sorte que $\tilde{\gamma} = \gamma \circ \varphi$ définit la même courbe tracée sur M (avec un autre paramétrage), la relation $\tilde{\gamma}'(0) = \varphi'(0)\gamma'(0)$, où le réel $\varphi'(0)$ est non nul, montre que les vecteurs $\tilde{\gamma}'(0)$ et $\gamma'(0)$ sont nuls en même temps et, quand ils ne sont pas nuls, définissent la même direction.

Proposition 3.6 *L'ensemble des vecteurs tangents en* a *à une sous-variété* M *de dimension* p *est un sous-espace vectoriel de dimension* p *de* $\mathbb{R}^n$.

Preuve Soit U un voisinage de $a \in \mathbb{R}^n$, V un voisinage de $0 \in \mathbb{R}^n$ et $f : U \to V$ un $\mathcal{C}^1$- difféomorphisme tel que $f(U \cap M) = V \cap (\mathbb{R}^p \times \{0\})$. On peut supposer $f(a) = 0$. Notons $f = (f_1, \cdots, , f_n)$ les composantes de f et soit $\gamma : I \to \mathbb{R}^n = \mathbb{R}^p \times \mathbb{R}^{n-p}$ une fonction de classe $\mathcal{C}^1$ qui définit un arc de courbe tracé sur M passant par a. Comme U est un ouvert, si $|t|$ est assez petit $\gamma(t) \in U \cap M$ et donc $f_j(\gamma(t)) = 0$ pour $j = p+1, \cdots, n$. On en déduit $Df_j(a) \cdot \gamma'(0) = 0$ pour $j = p+1, \cdots, n$. Cela veut dire que tout vecteur tangent en a à M, $v = \gamma'(0)$ vérifie $Df(a) \cdot v \in \mathbb{R}^p \times \{0\}$ ou encore que l'ensemble des vecteurs tangents en a est inclus dans $Df(a)^{-1}(\mathbb{R}^p \times \{0\})$ qui est un sous-espace vectoriel de $\mathbb{R}^n$ de dimension p puisque $Df(a)^{-1}$ est un isomorphisme.
Inversement soit $v \in Df(a)^{-1}(\mathbb{R}^p \times \{0\})$, de sorte que $w = Df(a) \cdot v \in \mathbb{R}^p \times \{0\}$. Posons, si $|t|$ est assez petit pour assurer $tw \in V$ (c'est possible puisque V est ouvert et contient 0), $\gamma(t) = f^{-1}(tw)$. L'application γ définit une courbe tracée sur M, passant par a pour laquelle $\gamma'(0) = D(f^{-1})(0) \cdot w = Df(a)^{-1} \cdot w = v$. ∎

Définition 3.7 *Le sous-espace vectoriel des vecteurs tangents en* a *à* M *est noté* T_aM. *Le sous-espace affine de* $\mathbb{R}^n$ *:* $a + T_aM$ *est appelé* espace tangent en a à M.

Les trois propositions suivantes décrivent concrêtement T_aM lorsque M est un graphe ou définie par des équations ou un paramétrage.

Proposition 3.8 *Soit* Ω *un ouvert de* $\mathbb{R}^p$ *et* $g : \Omega \to \mathbb{R}^{n-p}$ *(*$1 \leq p \leq n-1$*) une application de classe* $\mathcal{C}^1$. *Si* $a = (a', g(a')) \in \Gamma_g$, $T_a\Gamma_g = \Gamma_{Dg(a')}$.

Un exercice indispensable consiste à expliciter ce résultat dans le cas d'une courbe plane, d'une courbe ou d'une surface de $\mathbb{R}^3$.

Preuve On commence par se souvenir que le graphe d'une application linéaire de $\mathbb{R}^p$ dans $\mathbb{R}^{n-p}$ est bien un sous-espace vectoriel de dimension p de $\mathbb{R}^n$. Si $\gamma : I \to \mathbb{R}^n$ est une courbe tracée sur Γ_g passant par a, on a, pour tout $t \in I$, $\gamma(t) = (\gamma_1(t), \gamma_2(t))$ où $\gamma_1(t) \in \Omega$ et $\gamma_2(t) = g(\gamma_1(t))$. Donc en utilisant le théorème de dérivation d'une fonction composée et en remarquant que $\gamma_1(0) = a'$, $\gamma_2'(0) = Dg(a') \cdot \gamma_1'(0)$. On en déduit que $v = (\gamma_1'(0), \gamma_2'(0))$ appartient au graphe de $Dg(a')$, ce qui donne l'inclusion $T_a\Gamma_g \subset \Gamma_{Dg(a')}$. L'égalité vient alors de ce que ces deux espaces vectoriels ont la même dimension. ∎

Proposition 3.9 *Soit $M = F^{-1}(0)$, où $F : U \to \mathbb{R}^{n-p}$ est de classe $\mathcal{C}^1$ sur un ouvert U de $\mathbb{R}^n$. Si F est une submersion en tout point de M, alors pour tout $a \in M$, $T_aM = \ker DF(a)$.*

Autrement dit, si $F = (F_1, \cdots, F_{n-p})$, $T_aM = \bigcap_{i=1}^{n-p} \ker DF_i(a)$.

Preuve Si $\gamma : I \to \mathbb{R}^n$ définit un arc de courbe tracé sur M et passant par a, pour tout $t \in I$, $F \circ \gamma(t) = 0$. Donc $DF(a) \cdot \gamma'(0) = 0$. C'est dire que $\gamma'(0) \in \ker DF(a)$ et que $T_aM \subset \ker DF(a)$. Mais puisque F est une submersion en a, $DF(a)$ est de rang $n-p$, donc $\ker DF(a)$ est de dimension $n - (n-p) = p$, qui est aussi celle de T_aM. Ceci établit l'égalité annoncée. ∎

L'interprétation en termes d'intersection des noyaux des formes linéaires que sont les différentielles des composantes est de l'algèbre linéaire classique. Cela veut aussi dire que T_aM est le sous-espace vectoriel de $\mathbb{R}^n$ défini par les $n-p$ équations $DF_i(a) \cdot h = 0$ qui sont les *linéarisées* en a des $n-p$ équations $F_i(x) = 0$ qui définissent S.

Proposition 3.10 *Sous les hypothèses et notations de la proposition 3.4, si $a = \varphi(u_0)$, $T_aM = D\varphi(u_0)(\mathbb{R}^p)$.*

Preuve Soit $w \in Im\, D\varphi(u_0)$ et $v \in \mathbb{R}^p$ tel que $w = D\varphi(u_0) \cdot v$. Puisque Ω est ouvert, il existe un intervalle ouvert $I \ni 0$ tel que pour tout $t \in I$, $u_0 + tv \in \Omega_0$. L'application $\gamma : I \to \mathbb{R}^n$ donnée par $\gamma(t) = \varphi(u_0 + tv)$ définit un arc de courbe tracé sur M passant par a pour lequel $\gamma'(0) = D\varphi(u_0) \cdot v = w$. Ceci montre que $w \in T_aM$ et donne l'inclusion $D\varphi(u_0)(\mathbb{R}^p) \subset T_aM$. L'égalité des dimensions permet de conclure. ∎

Exemples

1. Le plan tangent à la surface de $\mathbb{R}^3$ d'équation $f(x, y, z) = 0$ au point $a = (x_0, y_0, z_0)$ (vérifiant $f(x_0, y_0, z_0) = 0$) a pour équation

$$\frac{\partial f}{\partial x}(a)(x - x_0) + \frac{\partial f}{\partial y}(a)(y - y_0) + \frac{\partial f}{\partial z}(a)(z - z_0) = 0.$$

   Dire que f est une submersion en a c'est dire que les trois dérivées partielles ne sont pas simultanément nulles en a et l'équation écrite est bien celle d'un plan passant par a.

2. La tangente à la courbe de $\mathbb{R}^3$ donnée par $f(x,y,z) = g(x,y,z) = 0$ au point $a = (x_0, y_0, z_0)$ a pour direction la droite intersection des deux plans

$$\begin{cases} \frac{\partial f}{\partial x}(a)x + \frac{\partial f}{\partial y}(a)y + \frac{\partial f}{\partial z}(a)z = 0 \\ \frac{\partial g}{\partial x}(a)x + \frac{\partial g}{\partial y}(a)y + \frac{\partial g}{\partial y}(a)z = 0 \end{cases}$$

3. Le plan tangent en $a = \varphi(s_0, t_0)$ à la surface de $\mathbb{R}^3$ paramétrée par $\varphi : U \to \mathbb{R}^3$ où $\varphi(s,t) = (f_1(s,t), f_2(s,t), f_3(s,t))$ a pour direction le plan engendré par les deux vecteurs de $\mathbb{R}^3$:

$$\begin{cases} \frac{\partial \varphi}{\partial s}(s_0,t_0) = \left(\frac{\partial f_1}{\partial s}(s_0,t_0), \frac{\partial f_2}{\partial s}(s_0,t_0), \frac{\partial f_3}{\partial s}(s_0,t_0)\right) \\ \frac{\partial \varphi}{\partial t}(s_0,t_0) = \left(\frac{\partial f_1}{\partial t}(s_0,t_0), \frac{\partial f_2}{\partial t}(s_0,t_0), \frac{\partial f_3}{\partial t}(s_0,t_0)\right) \end{cases}$$

Une équation du plan tangent est donc

$$\begin{vmatrix} x - f_1(s_0,t_0) & \frac{\partial f_1}{\partial s}(s_0,t_0) & \frac{\partial f_1}{\partial t}(s_0,t_0) \\ y - f_2(s_0,t_0) & \frac{\partial f_2}{\partial t}(s_0,t_0) & \frac{\partial f_2}{\partial t}(s_0,t_0) \\ z - f_3(s_0,t_0) & \frac{\partial f_3}{\partial s}(s_0,t_0) & \frac{\partial f_3}{\partial t}(s_0,t_0) \end{vmatrix} = 0.$$

4. L'hyperplan (affine) de $\mathbb{R}^{n+1}$ tangent à la sphère S^n au point $a = (a_0, \cdots, a_n)$ a pour équation

$$a_0(x_0 - a_0) + \cdots + a_n(x_n - a_n) = 0.$$

5. Si $A \in O(n)$,

$$T_A O(n) = \{H \in \mathcal{M}_n(\mathbb{R}) \mid A^t H + H^t A = 0\} = \{H \mid A^t H = -(A^t H)^t\}.$$

En particulier $T_I O(n)$ est l'espace vectoriel des matrices antisymétriques. En remarquant que, pour $A \in O(n)$, $A^t = A^{-1}$, on voit que

$$T_A O(n) = A\, T_I O(n).$$

Chapitre 4 : DIFFÉRENTIELLES D'ORDRE SUPÉRIEUR

17 janvier 2007

1 Différentielle seconde

1.1 Définition

Soit E et F des espaces normés, U un ouvert de E et $f : U \to F$ une application de classe $\mathcal{C}^1$ sur U. L'application $Df : U \to L(E,F)$ définie par $x \mapsto Df(x)$ est continue. Si elle est différentiable en $a \in U$, on dit que f est *deux fois différentiable* en a. Dans ce cas $D(Df)(a)$ est une application linéaire continue de E dans $L(E,F)$, c'est-à-dire un élément de $L(E, L(E,F))$.

Lemme 1.1 *L'espace normé $L(E, L(E,F))$ s'identifie à l'espace normé des applications bilinéaires continues de $E \times E$ dans F, noté $L^2(E,F)$.*

Rappelons qu'une application bilinéaire $B : E \times E \to F$ est continue si et seulement s'il existe $C > 0$ tel que pour tous $h, k \in E$, $\|B(h,k)\| \leq C\|h\|\,\|k\|$. La *norme* de B est, par définition, le « meilleur » C ou encore :

$$\|B\| = \sup_{\|h\|=\|k\|=1} \|B(h,k)\|.$$

En dimension finie, il n'y a pas lieu de se préoccuper des normes et l'identification est la suivante. On définit l'application linéaire (à vérifier) :

$$\Phi : L(E, L(E,F)) \to L^2(E,F)$$

en associant à T l'application bilinéaire $B_T : E \times E \to F$ définie par

$$B_T(h,k) = T(h) \cdot k$$

On définit l'application linéaire (à vérifier)

$$\Psi : L^2(E,F) \to L(E, L(E,F))$$

en associant à B l'application $T_B : E \to L(E,F)$ qui envoie $h \in E$ sur l'application linéaire $T_B(h) : E \to F$ donnée par $T_B(h) \cdot k = B(h,k)$ (pour tout $k \in E$). On vérifie sans (trop de) peine que Φ et Ψ sont inverses l'une de l'autre. Si E ou F n'est plus de dimension finie il faut ajouter la vérification, fastidieuse mais sans piège, de la continuité de ces applications linéaires et la conservation de la norme (Φ et Ψ sont des isométries).

On fera systématiquement cette identification et on note $D^2f(a)$ l'*application bilinéaire continue* à laquelle s'identifie $D(Df)(a)$. On a donc pour $h, k \in E$,

$$D^2f(a)(h,k) = \underbrace{\overbrace{D(Df)(a)}^{\in L(E,L(E,F))}(h)}_{\in L(E,F)} \cdot k$$

Exemples

1. Soit $u : E \to F$ une application linéaire continue. Alors, pour tout $x \in E$, $Du(x) = u$ donc Du est une application *constante* et sa différentielle est nulle en tout point. Autrement dit u est deux fois différentiable en tout point de E et $D^2u(x) = 0$ pour tout $x \in E$.
2. Soit $B : E \times E \to F$ une application bilinéaire continue. On sait que B est différentiable en tout point $(x,y) \in E \times E$ et que, pour tout $(h,k) \in E \times E$,

   $$DB(x,y) \cdot (h,k) = B(x,k) + B(h,y).$$

   Donc l'application $DB : (x,y) \mapsto DB(x,y)$ est *linéaire* continue donc différentiable et, pour tout $(x,y) \in E \times E$,

   $$D(DB)(x,y) = DB.$$

   Cela s'écrit, pour (h,k) et (h',k') appartenant à $E \times E$,

   $$\begin{aligned} D^2B(x,y)((h,k),(h',k')) &= D(DB)(x,y)(h,k) \cdot (h',k') \\ &= DB(h,k) \cdot (h',k') \\ &= B(h,k') + B(h',k) \end{aligned}$$

   et l'application $D^2B : E \times E \to L^2(E \times E, F)$ obtenue est *constante*.

1.2 Lemme de Schwarz

Dans l'identification précédente les vecteurs h et k ne jouent pas a priori le même rôle. En fait il n'en est rien.

Proposition 1.2 (Lemme de Schwarz) *Si $f : U \to F$ est deux fois différentiable en a, l'application bilinéaire $D^2f(a)$ est symétrique.*

Preuve Soit $r > 0$ assez petit pour que $Bo(a, 2r) \subset U$ et que Df y soit définie. Posons pour $\|h\| < r$ et $\|k\| < r$,

$$\varphi(h,k) = f(a+h+k) - f(a+h) - f(a+k) + f(a) - D^2f(a)(h,k)$$

et montrons que $\dfrac{\varphi(h,k)}{\|(h,k)\|^2} \to 0$ quand $\|(h,k)\| \to 0$. En effet la différentiabilité de Df en a s'exprime par le fait que, quel que soit $\varepsilon > 0$, il existe $\alpha > 0$ (on peut imposer $\alpha < r$) tel que si $\|h\| < 2\alpha$, alors

$$\|Df(a+h) - Df(a) - D(Df)(a)\cdot h\| < \varepsilon\|h\|$$

(il s'agit au premier membre de la norme d'un élément de $L(E,F)$).
Pour h fixé vérifiant $\|h\| < \alpha$, l'application $\varphi_h : k \mapsto \varphi(h,k)$ est définie dans la boule $Bo(0,\alpha)$, à valeurs dans F et différentiable en tout point de cette boule. De plus (dans $L(E,F)$) on a $D\varphi_h(k) = Df(a+h+k) - Df(a+k) - D(Df)(a)\cdot h = Df(a+h+k) - Df(a) - D(Df(a)\cdot(h+k) - [Df(a+k) - Df(a) - D(Df)(a)\cdot k]$. Et donc, puisque $\|h+k\| < 2\alpha$ et $\|k\| < 2\alpha$,

$$\|D\varphi_h(k)\| < \varepsilon(\|h+k\| + \|k\|) < 3\varepsilon\|(h,k)\|$$

La boule $Bo(0,\alpha)$ étant convexe, on peut appliquer le théorème des accroissements finis à φ_h sur le segment $[\,0,k\,]$. On obtient

$$\|\varphi_h(k) - \varphi_h(0)\| = \|\varphi(h,k)\| \le 3\varepsilon\|(h,k)\|\,\|k\| \le 3\varepsilon\|(h,k)\|^2$$

et le résultat annoncé suit.

Mais cela implique que l'application bilinéaire définie par

$$B(h,k) = D^2f(a)(h,k) - D^2f(a)(k,h)$$

vérifie $\dfrac{B(h,k)}{\|(h,k)\|^2} \to 0$ quand $\|(h,k)\| \to 0$, puisque $B(h,k) = \varphi(h,k) - \varphi(k,h)$. Or, la bilinéarité permet d'en déduire que $B = 0$. On raisonne comme dans le Lemme **1.1** du chapitre 1. : si $(h_0,k_0) \neq (0,0)$, pour tout réel $t \neq 0$, $\dfrac{B(th_0,tk_0)}{\|(th_0,tk_0)\|} = \dfrac{B(h_0,k_0)}{\|(h_0,k_0)\|}$. Mais, par hypothèse le terme de gauche tend vers 0 quand $t \to 0$. Cela oblige $B(h_0,k_0) = 0$. Or, dire que $B = 0$, c'est dire que $D^2f(a)$ est symétrique. ∎

1.3 Dérivées partielles secondes

Lorsque E et F sont de dimension finie, le choix de bases $\mathcal{V} = (v_1, \cdots, v_m)$ dans E et $\mathcal{W} = (w_1, \cdots, w_n)$ dans F permet de se donner l'application bilinéaire $D^2f(a)$ en se donnant les composantes dans $\mathcal{W}$ des m^2 vecteurs $D^2f(a)(v_i, v_j)$ $(i \leq j)$. Si f est donnée par ses composantes $f = (f_1, \cdots, f_n)$, cela revient à se donner pour chaque $\ell = 1, \cdots n$, les m^2 scalaires

$$D^2f_\ell(a)(v_i, v_j) = D(Df_\ell)(a)(v_i).v_j$$

Mais pour tout $x \in U$, on a par définition

$$Df_\ell(x)(v_i) = \frac{\partial f_\ell}{\partial x_i}(x)$$

et c'est donc la dérivée partielle par rapport à j de cette fonction de x que l'on calcule pour obtenir le scalaire

$$D^2f_\ell(a)(v_i, v_j) = \frac{\partial}{\partial x_j}\left(\frac{\partial f_\ell}{\partial x_i}\right)(a)$$

que l'on note $\boxed{\dfrac{\partial^2 f_\ell}{\partial x_j \partial x_i}(a)}$ et que l'on appelle dérivée partielle seconde de f_ℓ en a par rapport à x_j et x_i.

Le théorème de Schwarz s'énonce maintenant : pour tout $1 \leq i, j \leq m$ et tout $1 \leq \ell \leq n$, $\boxed{\dfrac{\partial^2 f_\ell}{\partial x_j \partial x_i}(a) = \dfrac{\partial^2 f_\ell}{\partial x_i \partial x_j}(a)}$

Lorsque $n = 1$, c'est-à-dire si $f : U \to \mathbb{R}$, $D^2f(a)$ est une *forme* bilinéaire symétrique dont la matrice est appelée matrice *hessienne* de f en a. C'est la matrice carrée symétrique constituée par les dérivées partielles secondes de f en a :

$$Hf(a) = \left(\frac{\partial^2 f}{\partial x_j \partial x_i}(a)\right)_{1 \leq i,j \leq m}.$$

Lorsque F est de dimension n, ceci s'applique à chacune des composantes f_ℓ de f.

1.4 Fonctions de classe $\mathcal{C}^2$

Définition 1.3 *On dit que f est de classe $\mathcal{C}^2$ dans U si f admet en tout point de U une différentielle seconde et si $D^2f : U \to L^2(E, F)$ est continue.*

Noter que l'*existence* de $D^2f(x)$ pour tout $x \in U$ impose à f d'être de classe $\mathcal{C}^1$ dans U.

Le résultat suivant s'obtient sans peine (?).

Proposition 1.4 *Si E et F sont de dimension finie, f est de classe $\mathcal{C}^2$ sur U si et seulement si chaque composante f_ℓ de f admet en tout point de U des dérivées partielles secondes et chacune des fonctions $x \mapsto \dfrac{\partial^2 f_\ell}{\partial x_j \partial x_i}(x)$ est continue dans U.*

2 Fonctions de classe $\mathcal{C}^p$

Rien n'empêche de continuer et d'adopter la définition inductive suivante.

Définition 2.1 *Soit E et F des espaces normés et U un ouvert de E.*

1. *On dit que $f : U \to F$ est de classe $\mathcal{C}^0$ si elle est continue.*
2. *On dit que $f : U \to F$ est de classe $\mathcal{C}^p$ ($p \geq 1$) si elle est différentiable dans U et si $Df : U \to L(E,F)$ est de classe $\mathcal{C}^{p-1}$.*
3. *On dit que $f : U \to F$ est de classe $\mathcal{C}^\infty$ si elle est de classe $\mathcal{C}^p$ pour tout $p \in \mathbb{N}$.*

Pour concrétiser les différentielles successives, il faut généraliser l'isomorphisme entre $L(E, L(E,F))$ et $L^2(E,F)$ en définissant $L^p(E,F) = \{$applications p-(multi)linéaires continues de $E^p = \underbrace{E \times \cdots \times E}_{p} \to F\}$. La continuité d'une application p-linéaire φ équivaut à l'existence de $C > 0$ tel que pour $(h_1, \cdots, h_p) \in E^p$, $\|\varphi(h_1, \cdots, h_p)\| \leq C\|h_1\|\,\|h_2\| \cdots \|h_p\|$. Ceci permet aussi de définir une norme sur $L^p(E,F)$. Le résultat clé est le fait (admis) que $L(E, L^{p-1}(E,F))$ s'identifie (avec sa norme) à $L^p(E,F)$.

On peut alors identifier la différentielle d'ordre p de f en a à une application p-linéaire continue de $E^p \to F$ notée $D^pf(a)$. L'analogue du lemme de Schwarz, est la

Proposition 2.2 *Si f est p fois différentiable en a, alors $D^pf(a)$ est symétrique, c'est-à-dire que pour toute permutation σ de $\{1, \cdots, p\}$, et tout $(h_1, \cdots, h_p) \in E^p$,*

$$D^pf(a)(h_{\sigma(1)}, h_{\sigma(2)}, \cdots, h_{\sigma(p)}) = D^pf(a)(h_1, \cdots, h_p).$$

A titre d'exemple, on peut retenir que toute application bilinéaire continue B est de classe $\mathcal{C}^\infty$ et que $D^pB = 0$ pour tout $p \geq 3$. Comment ce résultat se généralise-t-il ?

Un dernier résultat important.

Proposition 2.3 *Soient E, F et G des espaces normés, U un ouvert de E et V un ouvert de G. Si $f : U \to F$ et $g : V \to G$ sont de classe $\mathcal{C}^p$ et si $f(U) \subset V$, alors $g \circ f$ est de classe $\mathcal{C}^p$.*

Preuve Elle se fait par récurrence sur $p \geq 1$. Si $p = 1$, c'est un résultat connu et, pour tout $x \in U$, $D(g \circ f)(x) = Dg(f(x)) \circ Df(x)$. Cette formule s'interprête aussi en

$$D(g \circ f) = B \circ \Phi$$

où $\Phi = (Dg \circ f, Df)$ et où $B : L(F,G) \times L(E,F) \to L(E,G)$ est donnée par $B(u,v) = u \circ v$. Comme B est bilinéaire continue, elle est de classe $\mathcal{C}^\infty$. Supposons alors le résultat établi jusqu'à l'ordre $p-1$. Puisque Dg est de classe $\mathcal{C}^{p-1}$ et que f est de classe $\mathcal{C}^p$, a fortiori $\mathcal{C}^{p-1}$, l'hypothèse de récurrence montre déjà que les deux composantes de Φ, donc Φ, sont de classe $\mathcal{C}^{p-1}$. Un deuxième usage de l'hypothèse de récurrence montre alors que $B \circ \Phi$ est de classe $\mathcal{C}^{p-1}$. Mais cela veut dire que $D(g \circ f)$ est de classe $\mathcal{C}^{p-1}$ et donc que $g \circ f$ est de classe $\mathcal{C}^p$. ∎

La formule de Faa di Bruno permet d'exprimer les différentielles successives d'une fonction composée à l'aide de celles des facteurs. C'est vite très compliqué, mais on peut vérifier à titre d'exercice que

$$D^2(g\circ f)(a)(h,k) = D^2g(f(a))(Df(a)\cdot h, Df(a)\cdot k)+Dg(f(a))\cdot D^2f(a)(h,k)$$

formule qui généralise « visiblement » la formule

$$(g \circ f)''(x) = g''(f(x))f'(x)^2 + g'(f(x))f''(x).$$

3 Formule de Taylor

On connaît la formule de Taylor à une variable. Soit I un ouvert de $\mathbb{R}$ contenant le segment $[\,0,1\,]$ et $\varphi : I \to \mathbb{R}^n$ une fonction de classe $\mathcal{C}^p$ $(p \geq 1)$. Alors

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

Théorème 3.1 *Soit U un ouvert d'un espace normé E et $f : U \to \mathbb{R}^n$ une application de classe $\mathcal{C}^p$ ($p \geq 1$). Si a et b sont deux points de U tels que $[\,a, b\,] \subset U$, on pose $h = b - a$, $h^k = (\underbrace{h, \cdots, h}_{k})$ (pour $k \in \mathbb{N}^*$) et, pour $t \in [\,0, 1\,]$, $\gamma(t) = a + th$. Alors*

$$f(b) = f(a) + Df(a) \cdot h + \frac{1}{2} D^2 f(a)(h^2) + \cdots + \frac{1}{(p-1)!} D^{p-1} f(a)(h^{p-1}) + \\ + \int_0^1 \frac{(1-t)^{p-1}}{(p-1)!} D^p f(\gamma(t))(h^p)\, dt.$$

Preuve La fonction $\varphi(t) = f(a+th) = f \circ \gamma(t)$ est de classe $\mathcal{C}^p$ sur un ouvert de $\mathbb{R}$ contenant $[\,0, 1\,]$. On a $\gamma'(t) = h$ et pour tout $k \geq 2$, $\gamma^{(k)}(t) = 0$. On en déduit $\varphi'(t) = Df(\gamma(t)).h$, $\varphi''(t) = D^2 f(\gamma(t))(h^2)$ et, plus généralement, $\varphi^{(k)}(t) = D^k f(\gamma(t))(h^k)$. La formule est alors une simple retranscription de la formule rappelée en remarquant que $\varphi(0) = f(a)$ et $\varphi(1) = f(b)$. ∎

Corollaire 3.2 (Taylor-Young) *Soit $f : U \to \mathbb{R}^n$ une application de classe $\mathcal{C}^p$ et $a \in U$. Alors*

$$f(b) = f(a) + Df(a) \cdot h + \cdots + \frac{1}{p!} D^p f(a)(h^p) + R(h)$$

où

$$\lim_{h \to 0} \frac{R(h)}{\|h\|^p} = 0.$$

Preuve Soit $r > 0$ tel que $Bo(a, r) \subset U$. Si $\|h\| < r$, on peut appliquer le théorème précédent au segment $[\,a, a + h\,]$ et il suffit de montrer que

$$\int_0^1 \frac{(1-t)^{p-1}}{(p-1)!} D^p f(\gamma(t))(h^p)\, dt = \frac{1}{p!} D^p f(a)(h^p) + R(h)$$

avec $\lim\limits_{h \to 0} \dfrac{R(h)}{\|h\|^p} = 0$. Or, puisque $\displaystyle\int_0^1 \frac{(1-t)^{p-1}}{(p-1)!}\, dt = \frac{1}{p!}$, on a

$$R(h) = \int_0^1 \frac{(1-t)^{p-1}}{(p-1)!} [D^p f(\gamma(t))(h^p) - D^p f(a)(h^p)]\, dt.$$

Mais puisque $x \mapsto D^p f(x)$ est continue en a, pour tout $\varepsilon > 0$, il existe $\alpha > 0$ tel que si $\|x - a\| < \alpha$, $\|D^p f(x) - D^p f(a)\| < \varepsilon$. Au premier membre

de cette inégalité, il s'agit de la norme de l'application p-linéaire continue $D^p f(x) - D^p f(a)$. Si elle est plus petite que ε, alors, pour tout $v \in E$, on a $\|D^p f(x)(v^p) - D^p f(a)(v^p)\| \leq \varepsilon \|v\|^p$. Si on suppose que $\|h\| < \alpha$, alors $x = \gamma(t) = a + th$ vérifie $\|x - a\| = t\|h\| < \alpha$. On en déduit

$$\|R(h)\| \leq \varepsilon \|h\|^p \int_0^1 \frac{(1-t)^{p-1}}{(p-1)!}\, dt = \frac{\varepsilon}{p!}\|h\|^p$$

ce qui est le résultat cherché. ∎

Chapitre 5 : POINTS CRITIQUES ET EXTREMA

5 janvier 2006

Dans tout ce chapitre on étudie des fonctions à valeurs *réelles*. Si U est un ouvert d'un espace normé E et $f : U \to \mathbb{R}$ une fonction différentiable, alors pour tout $x \in U$, $Df(x)$ est une *forme* linéaire (continue) donc elle est *nulle ou surjective*.

Définition 0.1 *Si $Df(a) = 0$, on dit que a est un* point critique *de f et que $f(a)$ est une* valeur critique.

Définition 0.2 *Si $A \subset U$ (A* n'est pas *supposé ouvert...), on dit que $a \in A$ est un minimum (*resp. *maximum) local (*resp. *strict) relatif de f sur A s'il existe un voisinage ouvert V_a de a tel que, pour tout $x \in V_a \cap A$, on ait $f(x) \geq f(a)$ (resp. $f(x) \leq f(a)$) (resp. inégalités strictes si $x \neq a$). Si a est un minimum local ou un maximum local, on dit que c'est un extremum local.*
Si $A = U$, on supprime la partie soulignée et on dit que c'est le cas libre.

1 Etude du cas libre

Proposition 1.1 (Condition Nécessaire du 1er ordre) *Si la fonction différentiable $f : U \to \mathbb{R}$ admet un extremum local en $a \in U$, alors a est un point critique de f.*

Preuve Supposons que a est un minimum local de f. Il existe alors $\rho > 0$ tel que si $\|x - a\| < \rho$, alors $f(x) \geq f(a)$. Si $h \in E$, $h \neq 0$ et si $t \in \mathbb{R}$ vérifie $|t| < \frac{\rho}{\|h\|}$, alors $a + th \in Bo(a, \rho)$ et donc $f(a + th) \geq f(a)$. D'une part

$$Df(a) \cdot h = \lim_{t \to 0^+} \frac{f(a + th) - f(a)}{t} \geq 0$$

puisque numérateur et dénominateur sont des réels positifs. D'autre part

$$Df(a) \cdot h = \lim_{t \to 0^-} \frac{f(a+th) - f(a)}{t} \leq 0$$

puisque le numérateur est positif et le dénominateur négatif. La seule possibilité est donc $Df(a) \cdot h = 0$. ∎

Remarques

1. Si E est de dimension finie, a est un point critique si et seulement si les dérivées partielles de f sont toutes nulles en a.
2. La condition trouvée est nécessaire mais pas suffisante, comme le montre la fonction $x \to x^3$ qui est strictement croissante sur $\mathbb{R}$ bien que 0 en soit un point critique.

Si f est deux fois dérivable et si E est de *dimension finie*, ce que nous supposons jusqu'à la fin du chapitre, on peut obtenir une autre condition nécessaire en utilisant la formule de Taylor-Young à l'ordre 2. Rappelons que $D^2f(a)$ est une forme bilinéaire symétrique dont la forme quadratique associée, $h \mapsto D^2f(a)(h^2)$ est le terme d'ordre 2 figurant dans cette formule.

Quelques rappels sur les formes quadratiques réelles

On désigne par B une forme bilinéaire symétrique sur $E = \mathbb{R}^m$ et par Q la forme quadratique associée.

1. Le *noyau* de B (ou de Q) est le sous-espace vectoriel

$$\{h \in E \mid \forall k \in E, B(h,k) = 0\}.$$

2. On dit que B (ou Q) est *non dégénérée* si son noyau est le sous-espace 0.
3. On dit que B (ou Q) est *positive* (*resp.* négative) si pour tout $h \in E$, $Q(h) \geq 0$ (*resp.* $Q(h) \leq 0$).
4. On dit que B (ou Q) est *définie* si pour tout $h \neq 0$, $Q(h) \neq 0$. Une forme quadratique réelle définie est soit définie positive soit définie négative.
5. On dit que B (ou Q) est *indéfinie* s'il existe $h, k \in E$ tels que $Q(h) > 0$ et $Q(k) < 0$.

En termes de valeurs propres (toutes réelles) de la matrice M de B cela se traduit par :

– Q est non dégénérée si et seulement 0 n'est pas valeur propre ou encore $det\, M \neq 0$.
– Q est positive ssi toutes les valeurs propres de M sont ≥ 0,
– Q est définie positive ssi toutes les valeurs propres de M sont > 0,
– Q est indéfinie ssi M a des valeurs propres des deux signes.

Si p est le nombre de valeurs propres > 0, et q le nombre de valeurs propres < 0, le triplet $(p, q, m - q - p)$ est la *signature* de Q. Elle ne dépend pas de la base considérée (bien que les valeurs propres en dépendent).

Proposition 1.2 (Condition Nécessaire du 2e ordre) *Si $f : U \to \mathbb{R}$ de classe $\mathcal{C}^2$ a un minimum (resp. maximum) local en a, la hessienne de f en a est positive (resp. négative).*

Avant de donner la preuve on peut énoncer le

Corollaire 1.3 *Si a est un point critique et si la hessienne de f en a est indéfinie, a n'est pas un extremum local, c'est un* col *ou une* selle *(au choix).*

Preuve Supposons que a soit un minimum local et écrivons la formule de Taylor-Young à l'ordre 2. Si $a + h \in U$, on a

$$f(a + h) - f(a) = \frac{1}{2!} D^2 f(a)(h, h) + \varepsilon(h)$$

où $\frac{\varepsilon(h)}{\|h\|^2} \to 0$ quand $h \to 0$. On fixe $h_0 \in E$, $h_0 \neq 0$ et on écrit cette formule avec $h = th_0$, de sorte que $D^2 f(a)(h, h) = t^2 D^2 f(a)(h_0, h_0)$. Puisque $\|h\|^2 = t^2 \|h_0\|^2$, on a $\lim_{t \to 0} \dfrac{\varepsilon(th_0)}{t^2} = 0$ et donc, puisque

$$\frac{1}{2} D^2 f(a)(h_0, h_0) + \frac{\varepsilon(th_0)}{t^2} = \frac{f(a + th_0) - f(a)}{t^2} \geq 0$$

si $|t|$ est assez petit, à la limite $D^2 f(a)(h_0, h_0) \geq 0$. ∎

L'exemple de la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ donnée par $f(x, y) = x^2 - y^3$ montre que cette condition n'est pas suffisante. En effet $(0, 0)$ est un point critique en lequel la hessienne est positive mais dégénérée (c'est $(h, k) \mapsto h^2$) et f change de signe au voisinage de $(0, 0)$ qui n'est donc pas un extremum local.

Proposition 1.4 (Condition Suffisante) *Si $f : U \to \mathbb{R}$ est de classe $\mathcal{C}^2$ et si a est un point critique en lequel la hessienne est définie positive, alors a est un minimum local* strict.

Preuve Posons $S = \{h \in E \mid \|h\| = 1\}$. C'est une partie compacte de E (on est en dimension finie !) sur laquelle la forme quadratique $D^2f(a)(h,h)$ est bornée et atteint son minimum. Il existe donc h_0 tel que $\|h_0\| = 1$ et pour tout $h \in S$, $D^2f(a)(h,h) \geq D^2f(a)(h_0,h_0) = \rho$. Comme $h_0 \neq 0$ (il est de norme 1), et que $D^2f(a)$ est définie positive, $\rho > 0$. Pour tout $h \in E$, si $h \neq 0$, $\frac{h}{\|h\|} \in S$ et donc

$$D^2f(a)(h,h) = \|h\|^2 D^2f(a)(\frac{h}{\|h\|}, \frac{h}{\|h\|}) \geq \rho\|h\|^2$$

Mézalor (Taylor-Young encore)

$$f(a+h) - f(a) = \frac{1}{2}D^2f(a)(h,h) + \varepsilon(h) \geq \|h\|^2 \left[\frac{\rho}{2} + \frac{\varepsilon(h)}{\|h\|^2}\right]$$

où le crochet a une limite strictement positive si $h \to 0$ et est donc strictement positif si $\|h\|$ est assez petit. ∎

Le fait qu'on ait une propriété plus forte (extremum strict) suggère que cette condition n'est pas nécessaire. C'est bien le cas comme le montre l'exemple, toujours sur $\mathbb{R}^2$, de la fonction $f(x,y) = x^2 + y^4$ qui a $(0,0)$ pour minimum bien que sa hessienne en $(0,0)$ soit dégénérée.

2 Etude du cas lié

On étudie la situation suivante. On suppose que $f : U \to \mathbb{R}$ est définie sur un ouvert de $\mathbb{R}^n$ et que A est une sous-variété de $\mathbb{R}^n$ de dimension p.

Définition 2.1 *Si f est de classe $\mathcal{C}^1$, on dit que $a \in A \cap U$ est un* point critique *de f sur A si $T_aA \subset \ker Df(a)$.*

Cette définition est justifiée par la

Proposition 2.2 (Condition Nécessaire) *Si a est un extremum relatif de f sur A, alors a est un point critique de f sur A.*

Preuve Soit $v \in T_aA$. Il existe une fonction $\gamma : I \to \mathbb{R}^n$, de classe $\mathcal{C}^1$, où I est un intervalle ouvert contenant 0, telle que $\gamma(0) = a$, $\gamma'(0) = v$ et pour tout $t \in I$, $\gamma(t) \in A$. La fonction $\varphi = f \circ \gamma : I \to \mathbb{R}$ est une fonction d'une variable, dérivable et présentant en 0 un extremum local. Donc d'après l'étude du cas libre, $\varphi'(0) = 0$. Mais $\varphi'(0) = Df(a) \cdot \gamma'(0) = Df(a) \cdot v$, donc $v \in \ker Df(a)$. ∎

La proposition suivante traduit ce résultat lorsque A est « définie par équations », autrement dit lorsque $A = F^{-1}(0)$ où $F : V$ (ouvert de $\mathbb{R}^n$) $\to \mathbb{R}^{n-p}$ est une submersion en tout point de A.

Proposition 2.3 *Si $F = (F_1, \cdots, F_{n-p})$ et si a tel que $F_1(a) = F_2(a) = \cdots = F_{n-p}(a) = 0$ est un point critique de f sur A, alors il existe des scalaires $\lambda_1, \cdots, \lambda_{n-p}$, déterminés de manière unique, tels que*

$$Df(a) = \sum_{i=1}^{n-p} \lambda_i DF_i(a)$$

(il s'agit d'une égalité entre formes linéaires).

Preuve On sait que $T_aA = \bigcap_{i=1}^{n-p} \ker DF_i(a)$ et que, puisque T_aA est de dimension p, les $n-p$ formes linéaires $DF_i(a)$ sont linéairement indépendantes. La condition $\bigcap_{i=1}^{n-p} \ker DF_i(a) \subset \ker Df(a)$ signifie que les $n-p+1$ formes linéaires $Df(a)$ et $DF_i(a)$ $(i = 1, \cdots, n-p)$ sont linéairement dépendantes, donc que $Df(a)$ est une combinaison linéaire, déterminée de façon unique, des $DF_i(a)$. ∎

Les λ_i s'appellent *multiplicateurs de Lagrange* de f en a. Concrêtement les F_i s'interprètent comme des « contraintes » et f est une fonction à maximiser ou minimiser. On cherche les points critiques et les multiplicateurs de Lagrange en résolvant le système de $2n-p$ équations en les $2n-p$ inconnues $(\underbrace{a_1, \cdots, a_n}_{a}, \lambda_1, \cdots, \lambda_{n-p})$:

$$\begin{cases} F_j(a) = 0 & j = 1 \cdots n-p \\ \frac{\partial f}{\partial x_i}(a) = \sum_{j=1}^{n-p} \lambda_j \frac{\partial F_j}{\partial x_i}(a) & i = 1 \cdots n \end{cases}$$

Ce sytème est linéaire par rapport aux λ_j, ce qui, parfois, facilite sa résolution. ∎

On peut aussi énoncer des conditions (nécessaires ou suffisantes) au 2e ordre. Voici sans preuve un résultat. Si f et les F_i sont supposés de classe $\mathcal{C}^2$ sur U et si a est un point critique de f sur A de multiplicateurs de Lagrange $\lambda_1, \cdots, \lambda_{n-p}$, la restriction à l'espace vectoriel T_aA de la forme quadratique

$$D^2 f(a) - \sum_{i=1}^{n-p} \lambda_i D^2 F_i(a)$$

est appelée forme hessienne de f sur A en a.

1. Si a est un minimum local de f sur A, la hessienne de f sur A en a est positive.
2. Si la hessienne de f sur A en a est définie positive, a est un minimum local strict de f sur A.

Chapitre 6 : ÉQUATIONS DIFFÉRENTIELLES ORDINAIRES. THÉORÈMES FONDAMENTAUX

17 janvier 2007

Dans toute la suite les espaces vectoriels considérés sont de dimension finie.

1 Terminologie

1.1 E.D.O. du premier ordre. Solutions

Soit U un ouvert de $\mathbb{R} \times \mathbb{R}^n$ et $f : U \to \mathbb{R}^n$ une application. Elle définit l'« équation différentielle ordinaire du premier ordre »

$$(E) \quad y' = f(t, y)$$

parfois notée $\dot{y} = f(t, y)$.

Définition 1.1 *1. Une* solution *de (E) sur un intervalle J (d'intérieur non vide) est une fonction* dérivable $\varphi : J \to \mathbb{R}^n$ *telle que*

(a) pour tout $t \in J$, $(t, \varphi(t)) \in U$,

(b) pour tout $t \in J$, $\varphi'(t) = f(t, \varphi(t))$.

2. Le graphe d'une solution, $\Gamma_\varphi = \{(t, \varphi(t)) \mid t \in J\}$, est une courbe intégrale *de (E).*

3. Si $(t_0, y_0) \in U$, une solution de condition initiale (t_0, y_0) *est une solution φ telle que $J \ni t_0$ et $\varphi(t_0) = y_0$.*

Une « équation différentielle ordinaire » est donc une expression comportant une inconnue (représentée par y) qui est une *fonction* d'une variable notée t. Cette inconnue intervient aussi dans (E) par sa dérivée première

(d'où la mention *du premier ordre*). Résoudre l'équation c'est en trouver (ou caractériser) une (ou toutes les) solution(s).

Remarquons que J est un intervalle qui peut être ouvert, semi-ouvert, fermé, borné ou non. On a vu dans les chapitres précédents ce que signifie φ est dérivable sur J pour tout type d'intervalle : la notation φ' désigne la fonction de J dans $\mathbb{R}^n$ qui associe à t le vecteur dérivée en t. Si f est supposée *continue* toute solution de (E) est automatiquement $\mathcal{C}^1$. Elle est $\mathcal{C}^{k+1}$ (*resp.* $\mathcal{C}^\infty$) si on suppose que f est $\mathcal{C}^k$ (*resp.* $\mathcal{C}^\infty$).

Exemples

Evolution d'une population Si $p(t)$ représente la population d'une espèce donnée à l'instant t, le *taux de croissance* instantané est $\dfrac{p'(t)}{p(t)}$. L'évolution de la population est gouvernée par l'équation différentielle

$$\frac{p'}{p} = r(t,p) = g(t,p) - s(t,p)$$

où g et s sont deux fonctions, supposées connues, donnant le *taux de naissance* et le *taux de décès* de la population étudiée. Le cas le plus simple correspond à g et s constantes donc $r(t,p)$ est une constante α et les solutions sont de la forme $p(t) = ce^{\alpha t}$. Si $\alpha > 0$, ce modèle conduit à une population dont la croissance est illimitée, ce qui est peu réaliste. Un meilleur modèle est obtenu en supposant que $r = \alpha - \beta p$ où α et β sont des constantes positives. L'équation différentielle obtenue

$$p' = (\alpha - \beta p)p$$

est appelée *équation logistique*. Elle se résout (c'est une « équation à variables séparées »), mais il est intéressant de constater que les résultats de ce chapitre permettent d'établir les propriétés suivantes sans utiliser la forme explicite des solutions :

- si à l'instant t_0, $p(t_0) = p_0 = \frac{\alpha}{\beta}$, alors $p(t) = p_0$ pour tout $t > t_0$,
- si à l'instant t_0, $p(t_0) > p_0$, alors $p(t)$ est, après t_0, une fonction *décroissante* de t et $\lim_{t\to+\infty} p(t) = p_0$,
- si à l'instant t_0, $p(t_0) < p_0$, alors $p(t)$ est, après t_0, une fonction *croissante* de t et $\lim_{t\to+\infty} p(t) = p_0$.

Dans ces exemples la fonction $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ définissant l'équation (E) est $f(t,p) = \alpha p$ dans le premier cas et $f(t,p) = (\alpha - \beta p)p$ dans le second. Il s'agit de fonctions indépendantes de t. On dit alors que (E) est une équation *autonome*.

Système proie-prédateur Il s'agit de décrire l'évolution de deux espèces dont l'une (proie) est la nourriture de l'autre (prédateur). Il y a donc deux inconnues : la population des proies, $y_1(t)$ et la population des prédateurs $y_2(t)$. L'équation (E) est un *système différentiel* :

$$\left\{ \begin{array}{rcl} y_1' & = & r_1(t, y_1, y_2)y_1 \\ y_2' & = & r_2(t, y_1, y_2)y_2 \end{array} \right.$$

Une situation simple et pas trop irréaliste est $r_1(t, y_1, y_2) = \alpha - \beta y_2$ et $r_2(t, y_1, y_2) = -\gamma + \delta y_1$ où $\alpha, \beta, \gamma, \delta$ sont des constantes strictement positives. Ce choix s'interprête de la façon suivante : en l'absence de prédateurs le taux de croissance des proies est constant (α) et en présence de prédateurs ce taux diminue proportionnellement au nombre de prédateurs. De façon symétrique, en l'absence de proies les prédateurs s'éteignent à taux constant (γ) et en présence de proies le taux augmente proportionnellement au nombre de proies. Le système obtenu est un système de *Lotka – Volterra*

$$\left\{ \begin{array}{rcl} y_1' & = & \alpha y_1 - \beta y_1 y_2 \\ y_2' & = & -\gamma y_2 + \delta y_1 y_2 \end{array} \right.$$

qui n'est pas pas si facile à étudier...
Cette fois $f : \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2$ est donnée, si $y = (y_1, y_2)$, par $f = (f_1, f_2)$ où $f_1(t, y) = \alpha y_1 - \beta y_1 y_2$ et $f_2(t, y) = -\gamma y_2 + \delta y_1 y_2$. Il s'agit encore d'un système autonome, de dimension 2.

En dimension $n = 1$, on peut représenter les graphes des solutions dans le plan $(t, y) \in \mathbb{R}^2$. Même sans résoudre l'équation (E), on peut représenter le *champ des tangentes*. On peut en effet associer à tout point $(t_0, y_0) \in U$, la droite passant par M_0 de *pente* $f(t_0, y_0)$. Cette droite est la *tangente* en M_0 au graphe Γ_φ d'une solution de (E) telle que $\varphi(t_0) = y_0$. Inversement si Γ_φ est un graphe ayant la propriété qu'en tout point $M = (t, y) \in \Gamma_\varphi$, la tangente en M a pour pente $f(t, y)$, alors φ est une solution de (E) sur l'intervalle de définition de Γ_φ. Voici l'un des rares exemples où cette propriété permet de résoudre (E). Considérons l'équation différentielle

$$y' = -\frac{t}{y}$$

sur l'ouvert $U = \{(t, y) \in \mathbb{R} \times \mathbb{R} \mid y \neq 0\}$. Si $M = (t, y) \in U$, la droite OM a pour pente $\dfrac{y}{t}$ (elle est infinie si $t = 0$) et est donc *perpendiculaire* à

la tangente en M à une courbe intégrale passant par M (qui a pour pente $-\frac{t}{y}$). Or les courbes ayant la propriété qu'en chaque point la tangente est perpendiculaire au « rayon vecteur » sont les cercles de centre O. Un tel cercle a pour équation $t^2 + y^2 = R^2$, de sorte qu'on a bien $t + yy' = 0$.

Toujours en dimension 1, on a la classique définition suivante.

Définition 1.2 *Soit U un ouvert de $\mathbb{R} \times \mathbb{R}$ et $f : U \to \mathbb{R}$. Si $m \in \mathbb{R}$, la courbe C_m d'équation $f(t, y) = m$ est la* ligne isocline *de pente m de (E).*

Une courbe intégrale passant par $M_0 \in C_m$ a en M_0 une tangente de pente m. Le long de C_m les courbes intégrales, si elles existent, sont parallèles entre elles.

Exemple Les isoclines de l'équation (E) $y' = t - y^2$ sont les *paraboles* d'axe Ot, $t = y^2 + m$. L'isocline C_0 sépare le plan en deux régions : l'intérieur de la parabole $t = y^2$, région dans laquelle toute courbe intégrale est *croissante* et l'extérieur de la parabole, dans laquelle toute courbe intégrale est décroissante. Les points de C_0 ont vocation à être des extrema locaux des courbes intégrales.

Toujours sans chercher à résoudre (E), on peut déterminer les points en lesquels la dérivée seconde d'une solution s'annule (probables points d'inflexion). Rappelons que dans cet exemple toute solution est $\mathcal{C}^\infty$. Si $\varphi(t)$ est une solution on a $\varphi'(t) = t - \varphi(t)^2$ et, en dérivant, $\varphi''(t) = 1 - 2\varphi(t)\varphi'(t) = 1 - 2\varphi(t)(t - \varphi(t)^2)$. Donc $\varphi''(t) = 0$ si et seulement si $(t, \varphi(t)$ est un point de la courbe d'équation $2y(t - y^2) = 1$.

Si $n > 1$, il est plus difficile de faire des dessins. Cependant, si $n = 2$ et si $\varphi = (\varphi_1, \varphi_2)$ est une solution de (E) sur J, on peut représenter dans le plan $(y_1, y_2) \in \mathbb{R}^2$, la courbe paramétrée $\left\{ \begin{array}{rcl} y_1 & = & \varphi_1(t) \\ y_2 & = & \varphi_2(t) \end{array} \right.$, $t \in J$. Lorsque la fonction $f = (f_1, f_2)$ est indépendante de t (cas autonome), on peut sans résoudre représenter dans le plan (y_1, y_2) le *champ des tangentes* en traçant pour chaque point $M_0 = (y_1^0, y_2^0) \in U$, un segment de droite passant par M_0 de direction donnée par le vecteur $f(M_0) = (f_1(M_0), f_2(M_0))$. Ce segment est la tangente en M_0 à une (éventuelle) solution de (E) passant par ce point. A titre d'exercice on représentera ce champ pour le système de Lotka – Volterra. La représentation paramétrée des courbes intégrales s'appelle le *portrait de phases* de (E).

1.2 Généralisations

1.2.1 Equation différentielle implicite

C'est une équation $(\widetilde{E})$ $F(t, y, y') = 0$ où $F : \Omega \to \mathbb{R}^n$, Ω ouvert de $\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n$. Si on suppose que F est de classe $\mathcal{C}^1$ et qu'en $M_0 = (t_0, y_0, y_1) \in \Omega$ tel que $F(M_0) = 0$ on a $D_3F(M_0) \in GL(\mathbb{R}^n)$, on peut appliquer le théorème des fonctions implicites qui permet de se ramener (au moins théoriquement) au voisinage de M_0 à la forme « explicite » $y' = G(t, y)$ traitée précédemment.

1.2.2 E.D.O. d'ordre > 1

En Mécanique le mouvement d'un point matériel est régi par une équation différentielle du type $y'' = F(t, y, y')$ qui est du second ordre, donnée par une fonction $F : \Omega \to \mathbb{R}^n$ où Ω est un ouvert de $\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n$, de la forme $(t, y_1, y_2) \mapsto F(t, y_1, y_2)$. L'astuce suivante permet de faire entrer ce type d'équations dans le cadre précédent en augmentant la dimension. On pose $Y = (y_1, y_2)$ et on définit $f : \Omega \to \mathbb{R}^{2n}$ par $f(t, Y) = (y_2, F(t, y_1, y_2))$. L'équation différentielle (E) $Y' = f(t, Y)$ est identique au système

$$\left\{ \begin{array}{rcl} y_1' & = & y_2 \\ y_2' & = & F(t, y_1, y_2) \end{array} \right.$$

Une solution de (E) est une fonction $\varphi : J \to \mathbb{R}^{2n}$ de la forme $\varphi = (\varphi_1, \varphi_2)$ où φ_1 et φ_2 sont des fonctions de J dans $\mathbb{R}^n$ vérifiant

$$\left\{ \begin{array}{rcl} \varphi_1'(t) & = & \varphi_2(t) \\ \varphi_2'(t) & = & F(t, \varphi_1(t), \varphi_2(t)) \end{array} \right.$$

donc aussi $\varphi_1''(t) = F(t, \varphi_1(t), \varphi_1'(t))$. Inversement si $\varphi_1 : J \to \mathbb{R}^n$ est une fonction deux fois dérivable et telle que $\varphi_1''(t) = F(t, \varphi_1(t), \varphi_1'(t))$, en posant $\varphi_2(t) = \varphi_1'(t)$, la fonction $\varphi = (\varphi_1, \varphi_2)$ est solution de (E). Autrement dit φ_1 est solution de l'équation du second ordre si et seulement (φ_1, φ_1') est solution du système (E) du premier ordre de dimension $2n$ associé.

Pour un système d'ordre 2, une condition initiale est la donnée d'un point $(t_0, y_0, y_1) \in \Omega$. Une solution vérifiant cette condition initiale est une solution φ_1 telle que $\varphi_1(t_0) = y_0$ et $\varphi_1'(t_0) = y_1$. Si t est le temps, y_0 est donc la *position* et y_1 la *vitesse* à l'instant t_0.

Ceci se généralise à l'ordre p : un système de dimension n et d'ordre p, donnant la dérivée p-ième en fonction de t, y et des dérivées d'ordre $< p$, se transforme en un système d'ordre 1 de dimension np.

2 Théorie locale

2.1 Lemme de Gronwall

Proposition 2.1 *Soit $a < b$ deux réels, $f : [\,a, b\,] \to [\,0, +\infty\,[$ une fonction continue et $c \in [\,a, b\,]$. S'il existe $A \geq 0$ et $B \geq 0$ tels que pour tout $t \in [\,a, b\,]$,*

$$f(t) \leq A + B \left| \int_c^t f(s)\, ds \right|$$

alors pour tout $t \in [\,a, b\,]$,

$$f(t) \leq A\, e^{B|t-c|}.$$

Preuve Supposons d'abord $t \geq c$. Posons $F(t) = A + B \displaystyle\int_c^t f(s)\, ds$. C'est une fonction de classe $\mathcal{C}^1$ et $F'(t) = Bf(t)$. Par hypothèse, pour tout $t \in [\,c, b\,]$, $f(t) \leq F(t)$, de sorte que $\dfrac{d}{dt}\left[e^{-Bt}F(t)\right] = e^{-Bt}[Bf(t) - BF(t)] \leq 0$. On en déduit que la fonction $e^{-Bt}F(t)$ est décroissante sur $[\,c, b\,]$ et donc, pour tout $t \in [\,c, b\,]$,

$$e^{-Bt}F(t) \leq e^{-Bc}F(c) = A\, e^{-Bc}.$$

D'où, pour tout $t \in [\,c, b\,]$, $F(t) \leq Ae^{B(t-c)}$ et, a fortiori,

$$f(t) \leq A\, e^{B(t-c)}.$$

Si maintenant $t \leq c$, on pose $G(t) = A + B \displaystyle\int_t^c f(s)\, ds$ et l'hypothèse est $f(t) \leq G(t)$ pour tout $t \in [\,a, c\,]$. On a $G'(t) = -Bf(t) \geq -BG(t)$ et on en déduit, comme plus haut, la croissance de $e^{Bt}G(t)$ sur $[\,a, c\,]$, puis la majoration $e^{Bt}G(t) \leq e^{Bc}G(c) = Ae^{Bc}$ d'où l'on déduit enfin, pour tout $t \in [\,a, c\,]$,

$$f(t) \leq Ae^{B(c-t)} = Ae^{B|t-c|}. \quad \blacksquare$$

2.2 Fonctions lipschitziennes par rapport à y

Définition 2.2 *Soit U un ouvert de $\mathbb{R} \times \mathbb{R}^n$, $f : U \to \mathbb{R}^n$, $(t, y) \mapsto f(t, y)$ et $A \subset U$.*

1. *f est* uniformément sur *A lipschitzienne par rapport à y s'il existe $k > 0$ tel que pour (t, y_1) et $(t, y_2) \in A$,*

$$\|f(t, y_1) - f(t, y_2)\| \leq k\|y_1 - y_2\|.$$

2. *f est l*ocalement *lipschitzienne par rapport à y si pour tout $(t_0, y_0) \in U$, il existe un voisinage de ce point sur lequel f est uniformément lipschitzienne par rapport à y.*

Remarques

1. Une fonction lipschitzienne par rapport à y n'est pas automatiquement continue.
2. Si f admet une différentielle partielle dans la direction de y et si $D_2 f$ est continue sur U, alors f est localement lipschitzienne par rapport à y. Soit en effet $(t_0, y_0) \in U$. La continuité de $D_2 f$ implique l'existence d'un ouvert U_0 qu'on peut supposer de la forme $I \times Bo(y_0, \rho)$ avec I ouvert de $\mathbb{R}$ contenant t_0, tel que, si $(t, y) \in U_0$,

$$\|D_2 f(t, y) - D_2 f(t_0, y_0)\| \leq 1$$

Donc aussi $\|D_2 f(t, y)\| \leq m = 1 + \|D_2 f(t_0, y_0)\|$. On peut alors appliquer la formule des accroissements finis à la fonction $y \mapsto f(t, y)$ sur le convexe $Bo(y_0, \rho)$. Pour tout $t \in I$, si (t, y_1) et $(t, y_2) \in U_0$ on a la majoration $\|f(t, y_1) - f(t, y_2)\| \leq m\|y_1 - y_2\|$.
Cela s'applique bien sûr dans le cas, souvent réalisé en pratique, où f est de classe $\mathcal{C}^1$.

Proposition 2.3 *Si f est continue et localement lipschitzienne par rapport à y, alors pour tout K compact tel que $K \subset U$, f est uniformément sur K lipschitzienne par rapport à y.*

Preuve Par l'absurde. Sinon, pour tout entier non nul n, il existe (t_n, y_n) et (t_n, y'_n) dans K tels que $\|f(t_n, y_n) - f(t_n, y'_n)\| > n\|y_n - y'_n\|$. Grâce à la compacité de K, on peut extraire de la suite $((t_n, y_n))$ une sous-suite $((t_{\varphi(n)}, y_{\varphi(n)}))$ convergeant vers $(t_0, y_0) \in K$. Ensuite on extrait de la suite $((t_{\varphi(n)}, y'_{\varphi(n)}))$ une sous-suite $((t_{\varphi\circ\psi(n)}, y'_{\varphi\circ\psi(n)}))$ convergeant vers un point de K qui est nécessairement de la forme (t_0, y'_0), puisque la suite $(t_{\varphi\circ\psi(n)})$ est extraite de la suite $(t_{\varphi(n))}$ qui converge vers t_0. En changeant les notations pour alléger l'écriture, on peut donc supposer que $(t_n, y_n) \to (t_0, y_0)$, $(t_n, y'_n) \to (t_0, y'_0)$ et (puisque $\varphi \circ \psi(n) \geq n$) que pour tout $n > 0$,

$$\|f(t_n, y_n) - f(t_n, y'_n)\| > n\|y_n - y'_n\|.$$

Puisque f est continue elle est bornée (par M) sur le compact K et on a donc $\|y_n - y'_n\| \leq \dfrac{2M}{n}$. On en déduit que $y'_0 = y_0$.

L'hypothèse faite sur f permet alors de trouver un voisinage U_0 de (t_0, y_0) et un $k > 0$ tels que si (t, y) et $(t, y') \in U_0$, alors $\|f(t, y) - f(t, y')\| \leq k\|y - y'\|$. Or il existe un entier N tel que si $n > N$, (t_n, y_n) et (t_n, y'_n) appartiennent à U_0. Les deux inégalités $\|f(t_n, y_n) - f(t_n, y'_n)\| > n\|y_n - y'_n\|$ et $\|f(t_n, y_n) - f(t_n, y'_n)\| \leq k\,\|y_n - y'_n\|$ sont alors vérifiées pour $n > N$. Elles sont contradictoires dès que $n > k$. ∎

2.3 Théorie de Cauchy locale

Théorème 2.4 (Cauchy – Lipschitz local précisé) *Si f est une fonction continue et localement lipschitzienne par rapport à y sur un ouvert U de $\mathbb{R} \times \mathbb{R}^n$, à valeurs dans $\mathbb{R}^n$. Soit $[\,t_0 - a, t_0 + a\,]$ un segment de $\mathbb{R}$ et $Bf(y_0, r)$ une boule compacte de $\mathbb{R}^n$ tels que $K = [\,t_0 - a, t_0 + a\,] \times Bf(y_0, r) \subset U$. Soit M tel que pour tout $(t, y) \in K$, $\|f(t, y)\| \leq M$, et $T = \min(a, \frac{r}{M})$. Il existe une unique fonction φ dérivable sur $J = [\,t_0 - T, t_0 + T\,]$ à valeurs dans $Bf(y_0, r)$ telle que $\varphi(t_0) = y_0$ et, pour tout $t \in J$, $\varphi'(t) = f(t, \varphi(t))$.*

Preuve Unicité. Si φ et $\tilde{\varphi}$ vérifient les assertions de l'énoncé, pour tout $t \in J$, $\varphi(t) - \tilde{\varphi}(t) = \varphi(t) - \varphi(t_0) - [\tilde{\varphi}(t) - \tilde{\varphi}(t_0)] = \displaystyle\int_{t_0}^{t} \varphi'(s)\,ds - \int_{t_0}^{t} \tilde{\varphi}'(s)\,ds = \int_{t_0}^{t} [f(s, \varphi(s) - f(s, \tilde{\varphi}(s))]\,ds$.

La proposition 2.3 montre qu'il existe $k > 0$ tel que si $t \in [\,t_0 - a, t_0 + a\,]$ et $y_1, y_2 \in Bf(y_0, r)$ alors $\|f(t, y_1) - f(t, y_2)\| \leq k\|y_1 - y_2\|$. On en déduit pour tout $t \in J$,

$$\|\varphi(t) - \tilde{\varphi}(t)\| \leq k \left| \int_{t_0}^{t} \|\varphi(s) - \tilde{\varphi}(s)\|\,ds \right|.$$

On applique le lemme de Gronwall à la fonction continue $\|\varphi(t) - \tilde{\varphi}(t)\|$ sur le segment J avec $A = 0$, $B = k$ et $c = t_0$. On en déduit $\|\varphi(t) - \tilde{\varphi}(t)\| \leq 0$ donc $\varphi(t) = \tilde{\varphi}(t)$ pour tout $t \in J$.

Existence. Montrons qu'on peut définir par récurrence une suite $\varphi_n(t)$ de fonctions continues sur J à valeurs dans $Bf(y_0, r)$ en partant de $\varphi_0(t) = y_0$ (fonction constante) et en posant, pour $n > 1$ et $t \in J$,

$$\varphi_n(t) = y_0 + \int_{t_0}^{t} f(s, \varphi_{n-1}(s))\,ds.$$

En effet, si φ_{n-1} est définie, continue sur J et prend ses valeurs dans $Bf(y_0, r)$, alors φ_n est clairement définie et continue sur J et la majora-

tion $(*)_n$ suivante : pour tout $t \in J$,

$$\|\varphi_n(t) - y_0\| = \left\| \int_{t_0}^{t} f(s, \varphi_{n-1}(s))\, ds \right\| \leq M\,|t - t_0| \leq MT \leq r$$

montre que φ_n est à valeurs dans $Bf(y_0, r)$.

Etudions la convergence de la suite (φ_n). D'abord, pour tout $t \in J$, on a

$$(\star)_1 \quad \|\varphi_1(t) - \varphi_0(t)\| = \left\| \int_{t_0}^{t} f(s, y_0)\, ds \right\| \leq M\,|t - t_0|.$$

Ensuite, pour $n \geq 2$ et $t \in J$, on a la majoration $(\star)_n$ suivante :

$$\begin{aligned} \|\varphi_n(t) - \varphi_{n-1}(t)\| &= \left\| \int_{t_0}^{t} [f(s, \varphi_{n-1}(s)) - f(s, \varphi_{n-2}(s))]\, ds \right\| \\ &\leq \left| \int_{t_0}^{t} \|f(s, \varphi_{n-1}(s)) - f(s, \varphi_{n-2}(s))\|\, ds \right| \\ &\leq k \left| \int_{t_0}^{t} \|\varphi_{n-1}(s) - \varphi_{n-2}(s)\|\, ds \right| \end{aligned}$$

Si $n = 2$ et en utilisant $(\star)_1$ on en déduit, pour tout $t \in J$,

$$\|\varphi_2(t) - \varphi_1(t)\| \leq kM \frac{|t - t_0|^2}{2}.$$

En utilisant cette majoration dans $(\star)_3$, on voit qu'il est raisonnable d'espérer prouver par récurrence que, pour tout $t \in J$ et tout entier n,

$$\|\varphi_n(t) - \varphi_{n-1}(t)\| \leq k^{n-1} M \frac{|t - t_0|^n}{n!}.$$

Cette formule est vraie pour $n = 1$ et, en la supposant vraie au rang $n - 1$, on déduit de $(\star)_n$:

$$\|\varphi_n(t) - \varphi_{n-1}(t)\| \leq k \left| \int_{t_0}^{t} k^{n-2} M \frac{|s - t_0|^{n-1}}{(n-1)!}\, ds \right| = k^{n-1} M \frac{|t - t_0|^n}{n!}$$

ce qui est la formule cherchée au rang n. En prenant le sup sur J, on en déduit $\|\varphi_n - \varphi_{n-1}\|_\infty \leq \dfrac{k^{n-1} M T^n}{n!}$, terme général d'une série convergente. Ceci établit la convergence normale (donc uniforme) de la suite de fonctions (φ_n) sur J vers une fonction φ qui est donc continue sur J.

En faisant tendre n vers l'infini dans $(*)_n$, on voit que, pour tout $t \in J$, $\|\varphi(t) - y_0\| \leq r$, ce qui montre que φ est à valeurs dans $Bf(y_0, r)$. On peut alors écrire, pour tout $t \in J$,

$$\|f(t, \varphi_n(t)) - f(t, \varphi(t))\| \leq k\|\varphi_n(t) - \varphi(t)\| \leq k\|\varphi_n - \varphi\|_\infty$$

ce qui prouve que la suite de fonctions $(f(t, \varphi_n(t)))$ converge uniformément sur J vers $f(t, \varphi(t))$. On en déduit, pour tout $t \in J$,

$$\begin{aligned} \varphi(t) &= \lim_{n\to\infty} \varphi_n(t) = \lim_{n\to\infty} [y_0 + \int_{t_0}^{t} f(t, \varphi_{n-1}(t))\, dt] \\ &= y_0 + \int_{t_0}^{t} \lim_{n\to\infty} f(t, \varphi_{n-1}(t))\, dt = y_0 + \int_{t_0}^{t} f(t, \varphi(t))\, dt \end{aligned}$$

Cette formule montre, d'une part, que $\varphi(t_0) = y_0$ et, d'autre part, puisqu'une primitive de fonction continue est dérivable, que φ est dérivable dans J et que $\varphi'(t) = f(t, \varphi(t))$. ∎

On peut remarquer que la longueur $(2T)$ de l'intervalle J dépend de a, r, M et de l'existence mais pas de la valeur d'une constante de Lipschitz k par rapport à y.

Corollaire 2.5 (Cauchy – Lipschitz local) *Soit U un ouvert de $\mathbb{R} \times \mathbb{R}^n$ et $f : U \to \mathbb{R}^n$ une fonction continue et localement lipschitzienne par rapport à y. Alors, pour tout $(t_0, y_0) \in U$, il existe une unique solution φ de (E) définie dans un voisinage de t_0 et telle que $\varphi(t_0) = y_0$.*

Preuve Soit U_0 un voisinage de (t_0, y_0) sur lequel f est uniformément lipschitzienne par rapport à y. Soit $a, r > 0$ tels que $[\, t_0 - a, t_0 + a\,] \times Bf(y_0, r) \subset U_0$. Le théorème 2.4 fournit une solution de (E) définie dans un voisinage compact $[\, t_0 - T, t_0 + T\,]$ de t_0 et telle que $\varphi(t_0) = y_0$. L'unicité veut dire que, si $\tilde{\varphi}$ est une solution de (E) sur un intervalle J telle que $t_0 \in \overset{\circ}{J}$ et $\tilde{\varphi}(t_0) = y_0$, alors il existe un intervalle ouvert contenant t_0 tel que $\tilde{J} \subset J \cap [\, t_0 - T, t_0 + T\,]$ sur lequel $\tilde{\varphi}(t) = \varphi(t)$. Mais la continuité de $\tilde{\varphi}$ implique, puisque $\varphi(t_0) = y_0$ l'existence de $\tilde{J} \subset J \cap [\, t_0 - T, t_0 + T\,]$ tel que, pour $t \in \tilde{J}$, $\|\tilde{\varphi}(t) - y_0\| < r$, ce qui montre que sur cet intervalle $\tilde{\varphi}$ est à valeurs dans $Bf(y_0, r)$ et l'unicité découle aussi de 2.4. ∎

3 Théorie globale

3.1 Unicité globale

Proposition 3.1 *Soit $f : U \to \mathbb{R}^n$ continue et localement lipschitzienne par rapport à y. Si φ_1 et φ_2 sont deux solutions de (E) sur un même intervalle J et s'il existe $t_0 \in J$ en lequel $\varphi_1(t_0) = \varphi_2(t_0)$ alors $\varphi_1 = \varphi_2$.*

Interprétation géométrique Cette proposition signifie que deux courbes intégrales de (E) ne peuvent se couper en un point en lequel 2.5 s'applique.

Preuve C'est la *connexité* de J qui assure cette propriété. En effet soit $J_1 = \{t \in J \mid \varphi_1(t) = \varphi_2(t)\}$. C'est un fermé de J puisque φ_1 et φ_2 sont continues. Il est non vide par hypothèse. Montrons que c'est aussi un ouvert de J. Soit $t_1 \in J_1$. Posons $y_0 = \varphi_1(t_1) = \varphi_2(t_1)$ et appliquons 2.5 au point $(t_1, y_0) \in U$. Il existe une unique solution φ de (E) définie dans un voisinage J' de t_1 et telle que $\varphi(t_1) = y_0$. Sur $\overset{\circ}{J'} \cap J$, qui est un ouvert de J, φ_1 et φ_2 doivent toutes deux coïncider avec φ. Cela implique que $\overset{\circ}{J'} \cap J \subset J_1$ et montre que J_1 est un ouvert de J. La connexité de J exige donc $J_1 = J$. ∎

3.2 Solutions maximales

Définition 3.2 *Soit (E) une équation différentielle donnée par une fonction $f : U \to \mathbb{R}^n$.*

1. *Soit φ une solution de (E) sur un intervalle J. On dit que $\tilde{\varphi}$ est un* prolongement *de φ si*
   (a) *$\tilde{\varphi}$ est une solution de (E) sur un intervalle $\tilde{J}$ et*
   (b) *$J \subset \tilde{J}$ et pour tout $t \in J$, $\varphi(t) = \tilde{\varphi}(t)$.*
2. *Une solution φ de (E) sur un intervalle J est une* solution maximale *de (E) si elle n'admet aucun prolongement défini sur un intervalle $\tilde{J}$ avec $\tilde{J} \neq J$.*

Proposition 3.3 (critère de prolongement) *Supposons $f : U \to \mathbb{R}^n$ continue et localement lipschitzienne par rapport à y. Soit φ une solution de (E) sur un intervalle J. Si $\beta = \sup J$ est fini et si $\lim_{t \to \beta^-} \varphi(t) = \ell$ existe et vérifie $(\beta, \ell) \in U$, alors φ se prolonge à un intervalle $\tilde{J}$ pour lequel $\sup \tilde{J} > \beta$.*

Preuve Le théorème 2.5 s'applique en (β, ℓ) et fournit $T > 0$ et une unique solution φ_0 sur $[\,\beta - T, \beta + T\,]$ telle que $\varphi_0(\beta) = \ell$. Si $\beta \in J$, par continuité, on a $\varphi(\beta) = \ell$ donc φ_0 et φ, qui coïncident en β, coïncident par 3.1 sur $J \cap [\,\beta - T, \beta\,]$ et en posant $\tilde{\varphi}(t) = \begin{cases} \varphi(t) & \text{si} \quad t \in J \\ \varphi_0(t) & \text{si} \quad t \in [\,\beta - T, \beta + T\,] \end{cases}$ on définit le prolongement cherché.

Si $\beta \notin J$, montrons que l'hypothèse permet de prolonger φ en une solution définie sur $J \cup \{\beta\}$. On est alors ramené au cas précédent. On pose $\bar{\varphi}(t) = \begin{cases} \varphi(t) & \text{si} \quad t \in J \\ \ell & \text{si} \quad t = \beta \end{cases}$ et il faut montrer que $\bar{\varphi}$ est dérivable à gauche en β et que cette dérivée (à gauche) est $f(\beta, \ell)$. Or, par continuité de f et de φ,

$$f(\beta, \ell) = \lim_{t \to \beta^-} f(t, \varphi(t)) = \lim_{t \to \beta^-} \varphi'(t) = \lim_{t \to \beta^-} \bar{\varphi}'(t)$$

et on raisonne de la façon suivante (classique et à connaître). Pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que si $\beta - \eta < t < \beta$, $\|\bar{\varphi}'(t) - f(\beta, \ell)\| < \varepsilon$.
Si $\beta - \eta < t < t' < \beta$, on a

$$\bar{\varphi}(t') - \bar{\varphi}(t) = \int_t^{t'} \bar{\varphi}'(s)\, ds = \int_t^{t'} [\bar{\varphi}'(s) - f(\beta, \ell)]\, ds + (t' - t) f(\beta, \ell)$$

d'où en faisant tendre t' vers β,

$$(*) \quad \ell - \bar{\varphi}(t) = \lim_{t' \to \beta^-} \int_t^{t'} \bar{\varphi}'(s)\, ds = (\beta - t) f(\beta, \ell) + \psi(t)$$

où

$$\psi(t) = \lim_{t' \to \beta^-} \int_t^{t'} [\bar{\varphi}'(s) - f(\beta, \ell)]\, ds.$$

Mais

$$\left\| \int_t^{t'} [\bar{\varphi}'(s) - f(\beta, \ell)]\, ds \right\| \leq \varepsilon(t' - t) \leq \varepsilon(\beta - t)$$

et donc $\|\psi(t)\| \leq \varepsilon(\beta - t)$. En divisant par $\beta - t$ (qui est positif) $(*)$ s'écrit, pour $\beta - \eta < t < \beta$,

$$\left\| \frac{\bar{\varphi}(t) - \ell}{t - \beta} - f(\beta, \ell) \right\| \leq \varepsilon$$

et cela montre que $\bar{\varphi}$ admet en β une dérivée à gauche égale à $f(\beta, \ell)$. ∎

Théorème 3.4 (Cauchy – Lipschitz global) *Soit U un ouvert de $\mathbb{R} \times \mathbb{R}^n$ et $f : U \to \mathbb{R}^n$ une application continue et localement lipschitzienne par rapport à y. Pour tout $(t_0, y_0) \in U$, il existe une* unique solution maximale *φ telle que $\varphi(t_0) = y_0$. Cette solution est définie sur un intervalle* ouvert.

Preuve Soit $\mathcal{J} = \{(J, \varphi)\}$ où J est un intervalle contenant t_0 et $\varphi : J \to \mathbb{R}^n$ une solution telle que $\varphi(t_0) = y_0$. Le théorème 2.5 montre que $\mathcal{J} \neq \emptyset$. Posons

$$J_{max} = \bigcup_{(J,\varphi)\in\mathcal{J}} J$$

C'est une réunion de connexes (de $\mathbb{R}$) dont l'intersection est non vide (elle contient t_0). C'est donc un connexe de $\mathbb{R}$, c'est-à-dire un intervalle.
Définissons $\varphi_{max} : J_{max} \to \mathbb{R}^n$ de la façon suivante. Si $t \in J_{max}$, il existe par définition un couple (J, φ) tel que $J \ni t$ et φ est une solution de (E) sur J. On pose $\varphi_{max}(t) = \varphi(t)$. Bien sûr, il n'y a pas unicité du couple (J, φ) mais le théorème d'unicité assure que tout choix conduit au même résultat et permet donc de définir *la* solution maximale passant par (t_0, y_0). Pour montrer que J_{max} est ouvert, supposons que $\beta = \sup J_{max} \in J_{max}$. Par construction de J_{max} et définition d'une solution cela implique que $(\beta, \varphi_{max}(\beta)) \in U$, mais alors le critère de prolongement 3.3 contredit la maximalité de J_{max}. Le même argument s'applique à la borne inférieure et montre donc que l'intervalle J_{max} est ouvert. ∎

3.3 Solutions globales

Dans tout ce paragraphe plaçons-nous dans le cas particulier où $U = I \times V$ avec I intervalle ouvert de $\mathbb{R}$ et V ouvert de $\mathbb{R}^n$. On peut dans ce cas préciser le critère de prolongement 3.3.

Proposition 3.5 *Soit $f :]\,a, b\,[\times V \to \mathbb{R}^n$ une application continue et localement lipschitzienne par rapport à y. Soit $\varphi :]\,\alpha, \beta\,[\to V$ une solution de (E). On suppose que $\beta < b$. S'il existe $\delta > 0$ et K compact de V tels que pour tout $t \in]\,\beta - \delta, \beta\,[$, $\varphi(t) \in K$, alors φ se prolonge en une solution définie sur un intervalle $\tilde{J}$ tel que* $\sup \tilde{J} > \beta$. *Il y a une propriété analogue en α.*

Preuve Soit $K_0 = [\,\beta - \delta, \beta\,] \times K$. C'est un compact inclus dans U. Si $M_0 = \sup_{(t,y)\in K_0} \|f(t, y)\|$, alors φ est M_0 – lipschitzienne sur $[\,\beta - \delta, \beta\,[$ puisque si t et t' appartiennent à cet intervalle,

$$\|\varphi(t) - \varphi(t')\| = \left\| \int_{t'}^{t} f(s, \varphi(s))\, ds \right\| \leq M_0 |t - t'|.$$

Cela implique que φ est *uniformément* continue sur $[\,\beta - \delta, \beta\,[$ et admet donc un prolongement continu à l'adhérence de $[\,\beta - \delta, \beta\,[$, c'est-à-dire à $[\,\beta - \delta, \beta\,]$.

Cela veut dire que $\ell = \lim_{t\to\beta^-} \varphi(t)$ existe et appartient à K. Donc $(\beta, \ell) \in U$ et on peut appliquer le critère 3.3 pour conclure. ∎

Corollaire 3.6 (« Théorème des bouts ») *Soit φ une solution maximale de (E) définie sur $]\,T_*, T^*\,[$. Alors*
- *ou bien $T^* = b$,*
- *ou bien $T^* < b$ et pour tout K compact de V, il existe $t \in]\,T_*, T^*\,[$ tel que $\varphi(t) \notin K$. (Autrement dit « une solution maximale sort de tout compact »).*

Il y a un énoncé analogue en T_*.

Si on suppose de plus que $V = \mathbb{R}^n$, l'alternative s'écrit
- ou bien $T^* = b$,
- ou bien $T^* < b$ et $\limsup_{t\to\beta^-} \|\varphi(t)\| = +\infty$.

On peut montrer (nous l'admettrons) qu'en fait $\lim_{t\to\beta^-} \|\varphi(t)\| = +\infty$.

La définition suivante est naturelle.

Définition 3.7 *Une solution $\varphi : J \to V$ de (E) est* globale *si $J =]\,a, b\,[$.*

Clairement toute solution globale est maximale. La réciproque est fausse comme l'illustre l'exemple de l'équation (E) $y' = y^2$. Elle admet la solution globale $\varphi(t) = 0$. On en déduit qu'aucune autre solution ne s'annule sur son intervalle de définition. Si alors φ est une solution non nulle sur J, on a $\dfrac{\varphi'(t)}{\varphi(t)^2} = 1 = \dfrac{d}{dt}\left(-\dfrac{1}{\varphi(t)}\right)$ et donc $-\dfrac{1}{\varphi(t)} = t + c$, $t \in J$ et $c \in \mathbb{R}$. Pour tout $c \in \mathbb{R}$, la formule $\varphi(t) = -\dfrac{1}{t+c}$ définit *deux* solutions différentes de (E) : $\varphi_1(t) :]-\infty, -c\,[\to \mathbb{R}$ et $\varphi_2(t) :]-c, +\infty\,[\to \mathbb{R}$. Ces deux solutions sont maximales mais pas globales. On a $\lim_{t\to -c^-} \varphi_1(t) = +\infty$ et $\lim_{t\to -c^+} \varphi_2(t) = -\infty$, ce qui illustre le « théorème des bouts ».

Lorsque $V = \mathbb{R}^n$, voici un critère bien utile pour affirmer que les solutions maximales sont globales.

Proposition 3.8 *Soit $f : U = I \times \mathbb{R}^n \to \mathbb{R}^n$ une application continue. S'il existe $k : I \to [\,0, +\infty\,[$ continue telle que pour tout $t \in I$, l'application $y \mapsto f(t, y)$ soit $k(t)$ – lipschitzienne, alors pour tout $(t_0, y_0) \in U$, il existe une unique solution* globale *φ de (E) telle que $\varphi(t_0) = y_0$.*

Preuve Remarquons d'abord que l'hypothèse faite sur f entraîne qu'elle est localement lipschitzienne en y. En effet si (t_0, y_0) est un point de U et si $\alpha > 0$ est tel que $[\, t_0 - \alpha, t_0 + \alpha \,] \subset I$, en posant $\lambda = \max_{t \in [\, t_0 - \alpha, t_0 + \alpha \,]} k(t)$, on a, pour $y, y' \in \mathbb{R}^n$ et $t \in [\, t_0 - \alpha, t_0 + \alpha \,]$,

$$\|f(t, y) - f(t, y')\| \leq k(t)\|y - y'\| \leq \lambda \|y - y'\|.$$

On peut alors appliquer le théorème 3.4 et obtenir l'existence d'une unique solution maximale φ définie sur $] T_*, T^* [$ telle que $\varphi(t_0) = y_0$.
Supposons $T^* < b = \sup I$. L'intervalle $[\, t_0, T^* \,]$ est un compact de I. Posons $C = \max_{t \in [\, t_0, T^* \,]} k(t)$ et $M = \sup_{t \in [\, t_0, T^* \,]} \|f(t, y_0)\|$. Pour tout $t \in [\, t_0, T^* \,[$, on a

$$\|\varphi(t) - \varphi(t_0)\| = \left\| \int_{t_0}^{t} \varphi'(s)\, ds \right\| = \left\| \int_{t_0}^{t} f(s, \varphi(s))\, ds \right\|.$$

Mais pour tout $y \in \mathbb{R}^n$ et tout $s \in [\, t_0, T^* \,]$,

$$\|f(s, y)\| \leq \|f(s, y_0)\| + C\|y - y_0\| \leq C\|y\| + D$$

où $D = M + C\|y_0\|$. D'où pour $t \in [\, t_0, T^* \,[$,

$$\|\varphi(t)\| \leq \|y_0\| + D(t - t_0) + C \int_{t_0}^{t} \|\varphi(s)\|\, ds \leq A + C \int_{t_0}^{t} \|\varphi(s)\|\, ds$$

si $A = \|y_0\| + D(T^* - t_0)$. Le lemmme de Gronwall permet d'en déduire pour tout $t \in [\, t_0, T^* \,[$, la majoration

$$\|\varphi(t)\| \leq A e^{C(t - t_0)} \leq A e^{C(T^* - t_0)}$$

qui est en contradiction avec le théorème des bouts puisqu'on a supposé $T^* < b$. Donc $T^* = b$ et une preuve analogue fournit $T_* = a$. ∎

4 Un théorème de régularité

Voici, sans preuve (pour le moment ?) un théorème décrivant la dépendance de la solution par rapport aux conditions initiales où à d'éventuels paramètres.

Théorème 4.1 *Soit $W = I \times V \times \Lambda$ un ouvert de $\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^p$ (ce dernier facteur est l'espace des paramètres) et $f : W \to \mathbb{R}^n$, $(t, y, \lambda) \mapsto f(t, y, \lambda)$ une application de classe $\mathcal{C}^k$ ($k \geq 0$) supposée de plus localement lipschitzienne*

par rapport à y si $k = 0$. Pour tout $(t_0, y_0, \lambda_0) \in W$, on note $\varphi(t, y_0, \lambda_0)$ l'unique solution maximale de l'équation

$$y' = f(t, y, \lambda_0)$$

telle que $\varphi(t_0, y_0, \lambda_0) = y_0$. On note $J_{(y_0,\lambda_0)}$ l'intervalle ouvert sur lequel cette solution est définie. On pose

$$\Omega = \{(t, y_0, \lambda_0) \in W \mid t \in J_{(y_0,\lambda_0)}\}.$$

Alors

1. *Ω est un ouvert de $\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^p$,*
2. *l'application $\varphi : \Omega \to \mathbb{R}^n$ donnée par $(t, y_0, \lambda_0) \mapsto \varphi(t, y_0, \lambda_0)$ est de classe $\mathcal{C}^k$, et de classe $\mathcal{C}^{k+1}$ par rapport à t.*

C'est le nom donné traditionnellement aux équations différentielles définies par une fonction $f(t, y)$ qui dépend de façon *affine* de y. En identifiant, comme d'habitude, une application linéaire à sa matrice dans la base canonique, on adopte dans ce chapitre les notations suivantes. On note $\mathcal{M}(n, \mathbb{R})$ l'espace vectoriel des matrices $n \times n$ à coefficients réels. Si on a fixé une norme sur $\mathbb{R}^n$, on munit $\mathcal{M}(n, \mathbb{R})$ de la norme d'application linéaire correspondante.
Si I est un intervalle ouvert de $\mathbb{R}$ et si $A : I \to \mathcal{M}(n, \mathbb{R})$ et $b : I \to R^n$ sont deux applications continues, on considère l'équation différentielle

$$(E) \quad y' = A(t)y + b(t)$$

où le produit de la matrice $A(t)$ par le vecteur y de $\mathbb{R}^n$ s'effectue comme d'habitude et fournit un vecteur de $\mathbb{R}^n$. Le choix des normes permet d'écrire, pour tout $t \in I$, $\|A(t)y\| \leq \|A(t)\| \, \|y\|$. L'équation différentielle (E) est donnée par l'application $f : I \times \mathbb{R}^n \to \mathbb{R}^n$, $(t, y) \mapsto A(t)y + b(t)$.

L'équation (E) est dire *homogène* si $b = 0$, *autonome* (ou « à coefficients constants ») si, de plus, A est une application constante.

1 Équations non autonomes

La linéarité se traduit par le **très important** résultat général suivant.

Proposition 1.1 *Par tout point $(t_0, y_0) \in I \times \mathbb{R}^n$ passe une unique solution* globale.

Preuve L'application $k : I \to [\,0, +\infty\,[$ donnée par $k(t) = \|A(t)\|$ est continue par hypothèse et, pour $y_1, y_2 \in \mathbb{R}^n$, on a

$$\|f(t, y_1) - f(t, y_2)\| = \|A(t)(y_1 - y_2)\| \leq \|A(t)\| \, \|y_1 - y_2\| = k(t)\|y_1 - y_2\|$$

On peut donc appliquer la proposition 3.8 du chapitre 6 qui établit le caractère global des solutions maximales. ∎

Puisque toutes les solutions de (E) sont globales, il est naturel de noter $\mathcal{S}$ l'ensemble qu'elles forment, c'est-à-dire le sous-ensemble de $\mathcal{C}^1(I, \mathbb{R}^n)$ constitué par les fonctions dérivables $\varphi : I \to \mathbb{R}^n$ telles que, pour tout $t \in I, \varphi'(t) = A(t)\varphi(t) + b(t)$. La proposition précédente a le corollaire suivant.

Corollaire 1.2 *Pour tout $t_0 \in I$, on définit une* bijection F_{t_0} *de $\mathbb{R}^n$ dans $\mathcal{S}$ en associant à $y \in \mathbb{R}^n$ l'unique solution globale de (E) qui prend en t_0 la*

valeur y. Autrement dit, pour $y \in \mathbb{R}^n$, $F_{t_0}(y)$ est une fonction dérivable sur I vérifiant $F_{t_0}(y)(t_0) = y$ et, pour tout $t \in I$, $(F_{t_0}(y))'(t) = A(t)F_{t_0}(y)(t) + b(t)$.

Remarquons que l'inverse de F_{t_0} est l'application *évaluation en* t_0, $\varepsilon_{t_0} : \mathcal{S} \to \mathbb{R}^n$ définie par $\varphi \mapsto \varphi(t_0)$.

La structure algébrique de $\mathcal{S}$ sera précisée dans les paragraphes qui suivent.

1.1 Équations homogènes

Dans ce paragraphe la fonction $b(t)$ est nulle. On peut alors préciser la proposition 1.1.

Proposition 1.3 *La solution de condition initiale (t_0, y_0) vérifie, pour tout $t \in I$,*

$$\|F_{t_0}(y_0)(t)\| \leq \|y_0\| e^{\left|\int_{t_0}^t \|A(s)\|\, ds\right|}.$$

Preuve La preuve s'inspire de celle du lemme de Gronwall (et pourrait se déduire d'une version plus forte de celui-ci).

Pour tout $t \in I$, on a $\|F_{t_0}(y_0)(t) - y_0\| = \left\| \int_{t_0}^t A(s)F_{t_0}(y_0)(s)\, ds \right\|$.

En posant $g(t) = \|F_{t_0}(y_0)(t)\|$, on définit une fonction continue sur I à valeurs dans $[\,0, +\infty\,[$, qui vérifie l'inégalité :

$$g(t) \leq \|y_0\| + \left| \int_{t_0}^t \|A(s)\|\, g(s)\, ds \right|$$

Supposons $t > t_0$ (le cas $t < t_0$ est laissé en exercice).

Définisssons $G(t) = \|y_0\| + \displaystyle\int_{t_0}^t \|A(s)\|\, g(s)\, ds$.

C'est une fonction dérivable sur I et $G'(t) = \|A(t)\|\, g(t) \leq \|A(t)\|\, G(t)$ par hypothèse.

De sorte que la fonction $\Phi(t) = e^{-\int_{t_0}^t \|A(s)\|\, ds} G(t)$ a pour dérivée

$$\Phi'(t) = e^{-\int_{t_0}^t \|A(s)\|\, ds} \left[-\|A(t)\|\, G(t) + G'(t) \right] \leq 0.$$

C'est donc une fonction décroissante pour $t \geq t_0$ et

$$\Phi(t) \leq \Phi(t_0) = G(t_0) = \|y_0\|.$$

On en déduit

$$\|F_{t_0}(y_0)(t)\| = g(t) \leq G(t) \leq \|y_0\| e^{\int_{t_0}^{t} \|A(s)\|\, ds}$$

ce qui est le résultat voulu. ∎

Par ailleurs le caractère vraiment linéaire de (E) se traduit par la proposition suivante.

Proposition 1.4 *L'ensemble $\mathcal{S}$ est un sous-espace vectoriel de dimension n de $\mathcal{C}^1(I, \mathbb{R}^n)$. Plus précisément, pour tout $t_0 \in I$, F_{t_0} est un isomorphisme de $\mathbb{R}^n$ sur $\mathcal{S}$.*

Preuve Il est immédiat de vérifier que, comme $b(t) = 0$, toute combinaison linéaire de solutions est solution. Les applications ε_{t_0} et F_{t_0} sont dans ce cas clairement linéaires. Ce sont donc des isomorphismes et $\dim \mathcal{S} = \dim \mathbb{R}^n = n$. ∎

Remarques

1. Pour toute $\varphi \in \mathcal{S}$, si φ n'est pas la fonction nulle, alors, pour tout $t \in I$, $\varphi(t) \neq 0$.
2. Si $(v_1, \cdots, v_n)$ est une base de $\mathbb{R}^n$ et $t_0 \in I$, $(F_{t_0}(v_1), \cdots, F_{t_0}(v_n))$ est une base de $\mathcal{S}$. On parle plutôt de *système fondamental de solutions* de (E).

Définition 1.5 (matrice de solutions) *Une matrice carrée $n \times n$ dont chaque colonne est une solution de (E) est appelée matrice de solutions de (E).*

Si $X(t)$ est une matrice de solutions de (E), alors pour tout $t \in I$, on a $X'(t) = A(t)X(t)$ (produit des deux matrices). Ici $X'(t)$ désigne la matrice $n \times n$ obtenue en dérivant chaque coefficient de $X(t)$. On peut donc regarder $X(t)$ comme une solution de l'équation

$$(\widetilde{E}) \quad X' = A(t)X$$

qui est une équation linéaire homogène sur l'espace $\mathbb{R}^{n^2}$. Inversement toute solution de cette équation est une matrice dont les colonnes sont des solutions de (E).

Définition 1.6 (Wronskien) *Soit $X(t)$ une matrice de solutions de (E). Le déterminant de $X(t)$ est appelé* wronskien *de $X(t)$. On le note $W(t) = \det X(t)$.*

Théorème 1.7 (Théorème de Liouville) *Soit $X(t)$ une matrice de solutions de (E) et $W(t)$ son wronskien. Alors*

1. *$W(t)$ vérifie l'équation différentielle scalaire du premier ordre*

$$W'(t) = tr\, A(t) W(t)$$

où $tr\, A(t)$ désigne la trace de la matrice $A(t)$.

2. *Pour $t, t_0 \in I$,*

$$W(t) = W(t_0) e^{\int_{t_0}^{t} tr\, A(s)\, ds}.$$

Preuve La deuxième affirmation se déduit immédiatement de la première et de la formule bien connue donnant les solutions d'une équation linéaire scalaire du premier ordre.

Pour prouver la première affirmation, on sait que le déterminant de n vecteurs colonnes est une forme n-linéaire alternée sur $\mathbb{R}^n$ et on en déduit que si $C_1(t), \cdots, C_n(t)$ sont n fonctions dérivables de t (à valeurs dans $\mathbb{R}^n$), la fonction $U(t) = \det(C_1(t), \cdots, C_n(t))$ est dérivable et

$$U'(t) = \sum_{j=1}^{n} \det(C_1(t), \cdots, C_{j-1}(t), C_j'(t), C_{j+1}(t) \cdots, C_n(t))$$

Si chaque $C_j(t)$ est une solution $\varphi_j(t)$ de (E), $U(t)$ est le wronskien et la formule s'écrit

$$W'(t) = \sum_{j=1}^{n} \det(\varphi_1(t), \cdots, \varphi_{j-1}(t), A(t)\varphi_j(t), \varphi_{j+1}(t), \cdots, \varphi_n(t))$$

Un résultat « bien connu » d'algèbre linéaire fournit alors le résultat. En effet, on s'assure facilement que, pour t fixé, l'application $(\mathbb{R}^n)^n \to \mathbb{R}^n$ qui, à $(U_1, \cdots, U_n)$, associe $\sum_{j=1}^{n} \det(U_1, \cdots, U_{j-1}, A(t)U_j, U_{j+1}, \cdots, U_n)$, est n- linéaire alternée et coïncide donc à constante multiplicative près avec $\det(U_1, \cdots, , U_n)$. La constante se calcule en choisissant pour $(U_1, \cdots, U_n)$ la base canonique et en remarquant que $\det(e_1, \cdots, e_{j-1}, A(t)e_j, e_{j+1}, \cdots, e_n)$ est le $j$ème élèment de la diagonale de la matrice $A(t)$. La somme de ces n termes est alors la trace de $A(t)$ et donc $W'(t) = \mathrm{tr}\, A(t) W(t)$. ∎

Corollaire 1.8 *Le wronskien d'une matrice de solutions de (E) soit est identiquement nul soit ne s'annule jamais dans I. Une matrice de solutions est un système fondamental de solutions si et seulement si son wronskien est non nul en un point (donc en tout point) de I.*

On reprend les notations introduites ci-dessus et on note $X_{t_0}(t)$ la matrice de solutions dont la $j$ème colonne est $F_{t_0}(e_j)$ où $(e_1, \cdots, e_n)$ est la base canonique de $\mathbb{R}^n$. C'est un système fondamental de solutions de (E). On peut aussi voir X_{t_0} comme la solution de $(\widetilde{E})$ vérifiant la condition initiale $X_{t_0}(t_0) = I_n$ (matrice identité de dimension n).

Proposition 1.9 *Pour tout $y \in \mathbb{R}^n$ et tout $t \in I$,*

$$F_{t_0}(y)(t) = X_{t_0}(t)y$$

(produit de la matrice $X_{t_0}(t)$ par le vecteur y).

Preuve Si on pose $\varphi(t) = X_{t_0}(t)y$, on a $\varphi'(t) = X'_{t_0}(t)y = A(t)X_{t_0}(t)y = A(t)\varphi(t)$. Ceci montre que $\varphi \in \mathcal{S}$. D'autre part $\varphi(t_0) = I_n y = y$, donc φ est la solution de (E) de condition initiale (t_0, y), c'est donc $F_{t_0}(y)$. ∎

Corollaire 1.10 *Si $X(t)$ est une matrice de solutions de (E) et si $t_0 \in I$,*

$$X(t) = X_{t_0}(t)X(t_0).$$

Preuve Il suffit d'appliquer la proposition en prenant pour y la $j$ème colonne de $X(t_0)$ et d'utiliser la définition du produit de deux matrices. ∎

Proposition 1.11 (Définition de la Résolvante) *Pour $s, t \in I$ la matrice $X(t)X(s)^{-1}$ est indépendante du système fondamental $X(t)$ de solutions de (E) choisi. On la note $\mathcal{R}(t, s)$ et on l'appelle résolvante de (E) (ou aussi opérateur d'évolution).*

Preuve Soit $X(t)$ et $Y(t)$ deux systèmes fondamentaux de solutions de (E). Fixons $t_0 \in I$. On a pour tout $t \in I$, $X(t) = X_{t_0}(t)X(t_0)$ et $Y(t) = X_{t_0}(t)Y(t_0)$. Donc $X(t)X(s)^{-1} = X_{t_0}(t)X_{t_0}(s)^{-1} = Y(t)Y(s)^{-1}$. ∎

Proposition 1.12

1. *La matrice $\mathcal{R}(t, s)$ est celle de l'application linéaire $\varepsilon_t \circ F_s : \mathbb{R}^n \to \mathbb{R}^n$ dans la base canonique de $\mathbb{R}^n$. Elle est inversible et, pour tout $y \in \mathbb{R}^n$, l'application $t \mapsto \mathcal{R}(t, s)y$ est la solution de (E) qui vaut y en $t = s$. Autrement dit $\mathcal{R}(t, s)y = F_s(y)(t)$.*
2. *Pour $s, t, u \in I$, $\mathcal{R}(t, u)\mathcal{R}(u, s) = \mathcal{R}(t, s)$ et $\mathcal{R}(s, s) = I_n$. En particulier $\mathcal{R}(t, s)^{-1} = \mathcal{R}(s, t)$.*
3. *Pour tout s fixé dans I, l'application $t \mapsto \mathcal{R}(t, s)$ est le système fondamental de solutions de (E) qui prend la valeur I_n en $t = s$. Son wronskien vaut $e^{\int_s^t tr\, A(u)\, du}$.*

Preuve

1. Par définition, si $(e_1, \cdots, e_n)$ est la base canonique de $\mathbb{R}^n$, $\varepsilon_t \circ F_s(e_j)$ est la valeur en t de la solution de (E) qui prend en s la valeur e_j, mais cette solution constitue, toujours par définition, la j-ième colonne de $X_s(t)$. Donc $X_s(t)$ est la matrice de $\varepsilon_t \circ F_s$ dans la base canonique. Mais $X_s(t) = X_s(t)X_s(s)^{-1} = \mathcal{R}(t, s)$.
2. Les diverses affirmations se déduisent immédiatement des définitions.
3. La première affirmation se déduit de $\mathcal{R}(t, s) = X_s(t)$ et la deuxième s'obtient en appliquant le théorème de Liouville avec $t_0 = s$ en remarquant que le wronskien de $X_s(t)$ vaut 1 en $t = s$. ∎

1.2 Équations non homogènes

Maintenant (E) désigne l'équation $y' = A(t)y + b(t)$.

Proposition 1.13 *L'ensemble $\mathcal{S}$ des solutions de (E) est un sous-espace affine de $\mathcal{C}^1(I, \mathbb{R}^n)$, de la forme $\varphi + V$ où V désigne l'espace vectoriel des solutions de l'équation homogène (H) $y' = A(t)y$, dite « associée » à (E).*

Preuve Il suffit de remarquer que deux fonctions ψ et φ sont solutions de (E) si et seulement si la fonction φ est solution de (E) et la fonction $\psi - \varphi$ est solution de (H), c'est-à-dire appartient à V. ∎

L'utilisation de la résolvante de (H) permet de donner une formule exprimant la solution de (E) prenant en $t_0 \in I$ la valeur $y_0 \in \mathbb{R}^n$. C'est la « méthode de variation de la constante », appelée parfois aussi méthode de d'Alembert.

Théorème 1.14 *Si $\mathcal{R}(t, s)$ désigne la résolvante de (H) et si $(t_0, y_0) \in I \times \mathbb{R}^n$, la solution de (E) de condition initiale (t_0, y_0) est :*

$$F_{t_0}(y_0)(t) = \mathcal{R}(t, t_0)y_0 + \int_{t_0}^{t} \mathcal{R}(t, s)b(s)\, ds.$$

Preuve Une vérification directe est facile. La présentation « habituelle » qui suit est plus instructive et mérite d'être connue voire retenue. On sait que la solution « générale » de (H) est donnée par $\varphi(t) = \mathcal{R}(t, t_0)y$ où $y \in \mathbb{R}^n$ est arbitraire (c'est la « constante »). On fait « varier » y, c'est-à-dire qu'on cherche à déterminer une fonction $\psi : I \to \mathbb{R}^n$ telle que $\varphi(t) = \mathcal{R}(t, t_0)\psi(t)$

soit la solution de (E) de condition initiale (t_0, y_0). Ce dernier point demande $\psi(t_0) = y_0$. Pour satisfaire la première exigence on dérive :

$$\varphi'(t) = A(t)\mathcal{R}(t,t_0)\psi(t) + \mathcal{R}(t,t_0)\psi'(t) = A(t)\varphi(t) + \mathcal{R}(t,t_0)\psi'(t).$$

On voit alors que φ est solution de (E) si et seulement si $\mathcal{R}(t,t_0)\psi'(t) = b(t)$ ou encore $\psi'(t) = \mathcal{R}(t,t_0)^{-1}b(t) = \mathcal{R}(t_0,t)b(t)$. On en déduit

$$\psi(t) = y_0 + \int_{t_0}^{t} \psi'(s)\,ds = y_0 + \int_{t_0}^{t} \mathcal{R}(t_0,s)b(s)\,ds$$

puis

$$\begin{aligned} F_{t_0}(y_0)(t) &= \mathcal{R}(t,t_0)\psi(t) \\ &= \mathcal{R}(t,t_0)y_0 + \mathcal{R}(t,t_0)\int_{t_0}^{t} \mathcal{R}(t_0,s)b(s)\,ds \\ &= \mathcal{R}(t,t_0)y_0 + \int_{t_0}^{t} \mathcal{R}(t,s)b(s)\,ds \end{aligned}$$

Dans cette dernière formule, le premier terme appartient à V et le second est une « solution particulière » de (E). ∎

2 Équations autonomes

On rappelle maintenant comment résoudre explicitement une équation à coefficients constants $y' = Ay$ où $A \in \mathcal{M}(n,\mathbb{R})$. On peut alors expliciter la résolvante d'un tel système et, en appliquant le théorème 1.14, résoudre explicitement toute équation $y' = Ay + b(t)$ si $A \in \mathcal{M}(n,\mathbb{R})$ et $b \in \mathcal{C}^0(I,\mathbb{R}^n)$.

2.1 La fonction exponentielle

Vous connaissez probablement le résultat suivant.

Proposition 2.1 *Soit $A \in \mathcal{M}(n,\mathbb{R})$. La série de terme général $\dfrac{A^p}{p!}$ converge dans $\mathcal{M}(n,\mathbb{R})$. Sa somme se note e^A. On a $\left\|e^A\right\| \leq e^{\|A\|}$.*

Preuve L'espace $\mathcal{M}(n,\mathbb{R})$ est complet (il est de dimension finie) et la majoration $\dfrac{\|A^p\|}{p!} \leq \dfrac{\|A\|^p}{p!}$, due au choix de la norme indiqué en début de chapitre, montre que la série considérée est normalement convergente. Elle est donc convergente. L'inégalité s'obtient par passage à la limite dans l'inégalité correspondante pour les sommes partielles. ∎

Proposition 2.2 *1. Si $A, B \in \mathcal{M}(n,\mathbb{R})$ commutent, alors*

$$e^{A+B} = e^A\, e^B.$$

De plus, dans ce cas, e^A commute avec B et avec e^B.

2. Si $P \in GL(n,\mathbb{R})$ et $A \in \mathcal{M}(n,\mathbb{R})$, alors

$$e^{P^{-1}AP} = P^{-1}e^A P.$$

Preuve Puisque $AB = BA$, on peut utiliser la formule du binôme de Newton : pour tout entier p,

$$(A+B)^p = \sum_{k=0}^{p} \frac{p!}{k!(p-k)!} A^k B^{p-k}.$$

On en déduit, en utilisant la convergence normale pour justifier le calcul,

$$\begin{aligned} \sum_{p=0}^{\infty} \frac{(A+B)^p}{p!} &= \sum_{p=0}^{\infty}\sum_{k=0}^{p} \frac{A^k}{k!}\frac{B^{p-k}}{(p-k)!} \\ &= \sum_{k=0}^{\infty} \frac{A^k}{k!} \sum_{p=k}^{\infty} \frac{B^{p-k}}{(p-k)!} = e^A\, e^B. \end{aligned}$$

Si A et B commutent, A commute avec tout polynôme en B donc, par passage à la limite dans les sommes partielles, avec e^B. On peut en déduire la commutativité de e^A et e^B, mais cela se déduit de la formule établie puisque le membre de gauche garde la même valeur lorsqu'on échange A et B. Enfin le deuxième point se montre en utilisant la relation $(P^{-1}AP)^n = P^{-1}A^nP$ et en passant à la limite dans les sommes partielles. ∎

On en déduit immédiatement le corollaire suivant.

Corollaire 2.3 *Pour tout $A \in \mathcal{M}(n,\mathbb{R})$, $e^A \in GL(n,\mathbb{R})$ et $\left(e^A\right)^{-1} = e^{-A}$.*

L'exemple suivant montre que $e^A e^B$ peut être différent de e^{A+B} lorsque $AB \neq BA$. Prenons $A = \begin{pmatrix} 0 & 0 \\ \theta & 0 \end{pmatrix}$ et $B = \begin{pmatrix} 0 & -\theta \\ 0 & 0 \end{pmatrix}$. On a $A^2 = B^2 = 0$ donc $e^A = I_2 + A = \begin{pmatrix} 1 & 0 \\ \theta & 1 \end{pmatrix}$ et $e^B = I_2 + B = \begin{pmatrix} 1 & -\theta \\ 0 & 1 \end{pmatrix}$ d'où

$$e^A e^B = \begin{pmatrix} 1 & -\theta \\ \theta & 1-\theta^2 \end{pmatrix}.$$

D'autre part $A+B=\begin{pmatrix}0&-\theta\\ \theta&0\end{pmatrix}=\theta J_2$ où $J_2=\begin{pmatrix}0&-1\\ 1&0\end{pmatrix}$. On établit, par récurrence, que $(A+B)^{2n}=(-1)^n\theta^{2n}I_2$ et $(A+B)^{2n+1}=(-1)^n\theta^{2n+1}J_2$ et on en déduit que

$$e^{A+B}=\begin{pmatrix}\cos\theta&-\sin\theta\\ \sin\theta&\cos\theta\end{pmatrix}$$

qui est manifestement distinct de e^Ae^B. L'explication est :

$$AB=\begin{pmatrix}0&0\\ 0&-\theta^2\end{pmatrix}\neq BA=\begin{pmatrix}-\theta^2&0\\ 0&0\end{pmatrix}$$

La proposition suivante explique pourquoi on peut utiliser l'exponentielle d'une matrice pour résoudre un système linéaire autonome.

Proposition 2.4 *Si $A\in\mathcal{M}(n,\mathbb{R})$, la fonction $U:\mathbb{R}\to\mathcal{M}(n,\mathbb{R})$ définie par $U(t)=e^{tA}$ est de classe $\mathcal{C}^\infty$ et réalise un homomorphisme du groupe additif $(\mathbb{R},+)$ dans le groupe multiplicatif $(GL(n,\mathbb{R}),\cdot)$. (Ceci signifie que pour $s,t\in\mathbb{R}$, $U(s)$ et $U(t)$ sont inversibles et $U(s+t)=U(s)U(t)$).*

De plus, pour tout $t\in\mathbb{R}$, $U'(t)=AU(t)=U(t)A$.

Preuve La proposition 2.2, jointe à la remarque que pour $s,t\in\mathbb{R}$, les matrices sA et tA commutent, montre la partie « algébrique » de la proposition. Etablissons la dérivabilité de U. Si t_0 est choisi dans $\mathbb{R}$, pour tout $h\in\mathbb{R}$, on a (en utilisant 2.2)

$$\begin{aligned}\frac{U(t_0+h)-U(t_0)}{h}&=e^{t_0A}\frac{e^{Ah}-I_n}{h}=e^{t_0A}\sum_{p=1}^{\infty}\frac{h^{p-1}A^p}{p!}\\ &=e^{t_0A}A+h\,e^{t_0A}A^2\sum_{p=0}^{\infty}\frac{h^pA^p}{(p+2)!}\end{aligned}$$

et donc, en supposant $|h|\leq 1$,

$$\left\|\frac{U(t_0+h)-U(t_0)}{h}-e^{t_0A}A\right\|\leq\|A\|^2e^{(|t_0|+1)\|A\|}|h|$$

quantité qui tend vers 0 avec h. Ceci montre que $U(t)$ est dérivable en t_0 et que $U'(t_0)=e^{t_0A}A=Ae^{t_0A}$ puisque A et t_0A commutent. La valeur obtenue pour $U'(t)$ montre que c'est une fonction dérivable et, par récurrence, on voit que U est $\mathcal{C}^\infty$ et aussi que $U^{(p)}(t)=A^pe^{tA}$. ∎

2.2 Calcul d'une base de V

Proposition 2.5 *1. Pour $t_0 \in \mathbb{R}$ et $y_0 \in \mathbb{R}^n$, la solution de $y' = Ay$ de condition initiale (t_0, y_0) est*

$$\varphi(t, y_0) = e^{(t-t_0)A} y_0.$$

2. Si $(v_1, \cdots, v_n)$ est une base de $\mathbb{R}^n$, alors $(e^{tA}v_1, \cdots, e^{tA}v_n)$ est une base de V, ou encore un système fondamental de solutions de (E).

3. La résolvante de (E) est

$$\mathcal{R}(t, t_0) = e^{(t-t_0)A}.$$

Preuve La fonction $\varphi(t) = e^{(t-t_0)A}y_0$ vérifie $\varphi(t_0) = e^{0A}y_0 = I_n y_0 = y_0$ et s'écrit $\varphi(t) = U(t)c$ où $c = e^{-t_0 A}y_0$ est un vecteur de $\mathbb{R}^n$ indépendant de t. On déduit de la proposition 2.4 que φ est dérivable et $\varphi'(t) = U'(t)c = AU(t)c = A\varphi(t)$.
Avec les notations du premier paragraphe, $(e^{tA}v_1, \cdots, e^{tA}v_n)$ est l'image par F_{t_0} de la base choisie dans $\mathbb{R}^n$. C'est donc une base de V.
Le troisième point se montre aussi en faisant le lien avec le paragraphe précédent. La matrice de solutions associée au système fondamental de solutions du 2. s'écrit $X(t) = e^{tA}C$ où C est la matrice $n \times n$ dont la j-ième colonne est formée des coordonnées de v_j dans la base canonique. La matrice C est donc inversible et, par définition, $\mathcal{R}(t, t_0) = X(t)X(t_0)^{-1} = e^{tA}CC^{-1}e^{-t_0 A} = e^{(t-t_0)A}$. (Dans ce calcul, attention à l'ordre des termes : les matrices C et A ne commutent pas !). ∎

Cette proposition est une méthode pratique de calcul d'une base de V lorsque e^{tA} est (assez) facile à calculer... ce qui ne se produit pas très souvent... Citons cependant le cas $n = 2$ (toutes les méthodes y donnent des calculs « faciles ») et surtout, en dimension quelconque, le cas où A est diagonalisable et le cas, très particulier, $A = \lambda I_n + N$ où N est une matrice nilpotente. Expliquons ces deux situations.

- Si $A = P^{-1}DP$, alors $tA = P^{-1}(tD)P$ et (proposition 2.2) $e^{tA} = P^{-1}e^{tD}P$. Mais, si $D = \operatorname{diag}(\lambda_1, \cdots, \lambda_n)$, la définition montre que $e^{tD} = \operatorname{diag}(e^{t\lambda_1}, \cdots, e^{t\lambda_n})$. Bien sûr, provisoirement, on a supposé que c'est sur $\mathbb{R}$ que A est diagonalisable.
- Dans le deuxième cas, par définition, il existe un entier positif p tel que $N^p = 0$ (on sait d'ailleurs que $p \leq n$). Comme N et λI_n commutent, on peut écrire :

$$e^{tA} = e^{t(\lambda I_n + N)} = e^{\lambda t I_n} e^{tN} = e^{\lambda t} I_n e^{tN} = e^{\lambda t} e^{tN}$$

Or $e^{tN} = \sum_{k=0}^{p-1} \frac{t^k N^k}{k!}$ puisqu'à partir de p les puissances de N sont nulles. On obtient finalement

$$e^{t(\lambda I_n + N)} = e^{\lambda t} \sum_{k=0}^{p-1} \frac{t^k N^k}{k!}$$

formule qui montre qu'il suffit de calculer les puissances de N jusqu'à l'ordre $p-1$.

Inspiré par ces exemples on peut penser à établir le lemme suivant.

Lemme 2.6 *1. Si v est un vecteur propre de A pour la valeur propre λ, alors $e^{tA}v = e^{\lambda t}v$.*

2. Si $r \in \mathbb{N}^*$ et $w \in \ker(A - \lambda I_n)^r$, alors $e^{tA}w = e^{\lambda t}P(t)$ où $P(t)$ est un polynôme de degré inférieur ou égal à $r-1$, à coefficients dans $\mathbb{R}^n$.

Preuve Puisque $Av = \lambda v$, par récurrence, pour tout entier p, $A^p v = \lambda^p v$ et donc

$$e^{tA}v = \sum_{p=0}^{\infty} \frac{t^p \lambda^p}{p!} v = e^{\lambda t}v.$$

On remarque qu'un vecteur propre étant un élément de $\ker(A - \lambda I_n)$, le deuxième point est établi pour $r = 1$: le vecteur v peut être vu comme un polynôme de degré 0.
En écrivant $A = \lambda I_n + (A - \lambda I_n)$ et en calculant comme plus haut, on a

$$e^{tA}w = e^{\lambda t}e^{(tA - \lambda I_n)}w = e^{\lambda t} \sum_{p=0}^{\infty} \frac{t^p}{p!}(A - \lambda I_n)^p w$$

Mais, si $p \geq r$, $(A - \lambda I_n)^p w = 0$, ce qui donne le résultat. ∎

Malheureusement, une matrice réelle n'a pas toujours de valeur propre réelle (ou pas toujours « suffisamment ») . On s'en sort en « complexifiant », c'est-à-dire en faisant agir A sur $\mathbb{C}^n$. On cherche des fonctions $\varphi_{\mathbb{C}} : \mathbb{R} \to \mathbb{C}^n$ (c'est bien un espace vectoriel sur $\mathbb{R}$), dérivables et telles que $\varphi'_{\mathbb{C}}(t) = A\varphi_{\mathbb{C}}(t)$. Si $z = (z_1, \cdots, z_n) \in \mathbb{C}^n$, on écrit $z_j = x_j + iy_j$ où $x_j, y_j \in \mathbb{R}$ et on note $\Re z = (x_1, \cdots, x_n) \in \mathbb{R}^n$ et $\Im z = (y_1, \cdots, y_n) \in \mathbb{R}^n$. On peut alors écrire $z = \Re z + i\Im z$. Avec ces notations, comme A est une matrice réelle, on voit que $\Re(A\varphi_{\mathbb{C}}(t)) = A\Re\varphi_{\mathbb{C}}(t)$. D'autre part, par définition de la dérivée d'une fonction (d'une variable réelle) à valeurs complexes, on a $\Re\varphi'_{\mathbb{C}}(t) =$

$(\Re\varphi_{\mathbb{C}})'(t)$. Le même discours étant valable pour la partie imaginaire, on voit que la connaissance d'une solution à valeurs complexes fournit, en prenant la partie réelle et la partie imaginaire, deux solutions réelles.

En recopiant les points correspondants de l'étude précédente, on montre que $V_{\mathbb{C}} = \{\varphi : \mathbb{R} \to \mathbb{C}^n \mid \varphi'(t) = A\varphi(t)\}$ est un $\mathbb{C}$-espace vectoriel de dimension n et que si $(v_1, \cdots, v_n)$ est une base de $\mathbb{C}^n$, alors $(e^{tA}v_1, \cdots, e^{tA}v_n)$ est une base de $V_{\mathbb{C}}$. Mais cette fois l'algèbre linéaire nous dit que, si le polynôme caractéristique de A est $P(\lambda) = \prod_{j=1}^{s}(\lambda - \lambda_j)^{r_j}$ (et donc $\sum_{j=1}^{s} r_j = n$), on a $\mathbb{C}^n = \bigoplus_{j=1}^{s} \ker(A - \lambda_j I_n)^{r_j}$. Il est donc possible de déterminer une base de $\mathbb{C}^n$ formée de vecteurs auxquels on peut appliquer le lemme 2.6. On obtient de cette façon une base de $V_{\mathbb{C}}$ constituée de fonctions de la forme $e^{\lambda_j t}p(t)$ où λ_j est une valeur propre de A et p un polynôme à coefficients dans $\mathbb{C}^n$ de degré $\leq r_j - 1$. Chaque fois que λ_j est réel p l'est aussi et la solution correspondante appartient à V. Si $\lambda_j = \alpha_j + i\beta_j$ (avec $\beta_j \neq 0$), la partie réelle et la partie imaginaire sont des éléments de V de la forme $e^{\alpha_j t}[q(t)\cos\beta_j t + \tilde{q}(t)\sin\beta_j t]$ où q et $\tilde{q}$ sont des polynômes de degré $\leq r_j - 1$. Bien sûr $\bar{\lambda}_j$ (qui est aussi valeur propres de A) fournit les mêmes éléments de V.

2.3 Cas d'une équation scalaire d'ordre n

Il s'agit dans ce paragraphe d'étudier une équation scalaire d'ordre n de la forme

$$(e) \quad a_n y^{(n)} + a_{n-1} y^{(n-1)} + \cdots + a_0 y = f(t)$$

où les $a_i \in \mathbb{R}$, avec $a_n \neq 0$, et $f : I \to \mathbb{R}$ est une fonction continue. On sait transformer cette équation en le système d'ordre 1 et de dimension n, $Y' = AY + b(t)$ où $A = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & 1 & \cdots \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \\ c_1 & c_2 & \cdots & c_n \end{pmatrix}$ avec $c_j = -\dfrac{a_{j-1}}{a_n}$ et $b(t) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \frac{1}{a_n} f(t) \end{pmatrix}$. La matrice A est appelée « matrice compagnon », pour indiquer son origine et sa forme particulière.

Il est cependant possible et instructif d'étudier directement les solutions (à valeurs complexes d'abord) de l'équation sans passer par le système associé. Pour cela, on introduit l'*opérateur différentiel* :

$$P(\frac{d}{dt}) = \sum_{j=1}^{n} a_j \frac{d^j}{dt^j}.$$

Un calcul aisé montre que si $\lambda \in \mathbb{C}$,

$$(1) \quad P(\frac{d}{dt})(e^{\lambda t}) = P(\lambda) e^{\lambda t}.$$

On en déduit que $e^{\lambda t} \in V_{\mathbb{C}}$, espace vectoriel (de dimension n sur $\mathbb{C}$) des solutions de l'équation homogène (h) associée à (e), si et seulement si $P(\lambda) = 0$. On constate facilement que P, appelé polynôme caractéristique de (e), n'est autre que le polynôme caractéristique de A.
En dérivant (1) q-fois par rapport à λ et en utilisant le lemme de Schwarz pour justifier l'échange de l'ordre des dérivations par rapport à λ et par rapport à t, on a

$$P(\frac{d}{dt})(t^q e^{\lambda t}) = P(\frac{d}{dt}) \frac{d^q}{d\lambda^q}(e^{\lambda t}) = \frac{d^q}{d\lambda^q} \left(P(\lambda) e^{\lambda t} \right)$$

A l'aide de la formule de Leibnitz, on en déduit

$$P(\frac{d}{dt})(t^q e^{\lambda t}) = \sum_{k=0}^{q} \left(\begin{array}{c} q \\ k \end{array} \right) P^{(k)}(\lambda) t^{q-k} e^{\lambda t}.$$

Si λ_0 est racine d'ordre r de P, on a $P(\lambda_0) = P'(\lambda_0) = \cdots = P^{(r-1)}(\lambda_0) = 0$ et $P^{(r)}(\lambda_0) \neq 0$. On en déduit que les fonctions $t^q e^{\lambda_0 t}$ appartiennent à $V_{\mathbb{C}}$ si et seulement si $0 \leq q \leq r-1$. Il est alors facile, en prouvant l'indépendance linéaire sur $\mathbb{C}$ des n solutions obtenues, détablir le théorème suivant.

Théorème 2.7 *1. Si $\lambda_1, \cdots, \lambda_s$ sont les racines complexes distinctes du polynôme caractéristique de (e) et si $r_1, \cdots, r_s$ sont les multiplicités respectives, alors une base de $V_{\mathbb{C}}$ est constituée des n fonctions $t^q e^{\lambda_j t}$ pour $0 \leq q \leq r_j - 1$ et $1 \leq j \leq s$.*

2. Si $\lambda_j = \alpha_j + i\beta_j$ avec $\alpha_j, \beta_j \in \mathbb{R}$, une base de l'espace vectoriel V des solutions réelles de (h) est consituée des fonctions $t^q e^{\lambda_j t}$ pour $0 \leq q \leq r_j - 1$ si $\beta_j = 0$ et des fonctions $t^q e^{\alpha_j t} \cos \beta_j t$ et $t^q e^{\alpha_j t} \sin \beta_j t$ pour $0 \leq q \leq r_j - 1$ si $\beta_j \neq 0$.

Enfin il est utile de connaître la forme que prend dans ce cas la « méthode de variation de la constante ». Supposons que $(\varphi_1(t), \cdots, \varphi_n(t))$ soit une base de V. Pour $j = 1, \cdots, n$, le vecteur $Y_j(t) = \begin{pmatrix} \varphi_j(t) \\ \varphi_j'(t) \\ \vdots \\ \varphi_j^{(n-1)}(t) \end{pmatrix}$ est une solution du système $Y' = AY$ et la matrice $n \times n$ dont la j-ième colonne est $Y_j(t)$ est un système fondamental de solutions de ce système. Son wronskien $W(t)$ est appelé wronskien des n solutions $\varphi_1(t), \cdots, \varphi_n(t)$. Il est non nul pour tout $t \in \mathbb{R}$. Cherchons alors à déterminer n fonctions $\mu_j(t)$ $(j = 1, \cdots, n)$ de sorte que le vecteur $Y(t) = \sum_{j=1}^{n} \mu_j(t) Y_j(t)$ vérifie $Y'(t) = AY(t) + b(t)$. On doit avoir

$$\begin{aligned} AY(t) + b(t) &= \sum_{j=1}^{n} \mu_j'(t) Y_j(t) + \mu_j(t) Y_j'(t) = \sum_{j=1}^{n} \mu_j'(t) Y_j(t) + \mu_j(t) A Y_j(t) \\ &= AY(t) + \sum_{j=1}^{n} \mu_j'(t) Y_j(t) \end{aligned}$$

et ceci est réalisé si et seulement si

$$\sum_{j=1}^{n} \mu_j'(t) Y_j(t) = b(t)$$

Cette condition demande aux μ_j' d'être solutions d'un système linéaire dont le déterminant est $W(t)$, donc non nul. On peut alors déterminer de façon unique les μ_j', puis (par intégation) les μ_j et obtenir une solution particulière de (e) donnée par la formule

$$(p) \qquad \varphi(t) = \sum_{j=1}^{n} \mu_j(t)\varphi_j(t).$$

Signalons la forme explicite du système à résoudre :

$$\begin{cases} \mu_1'\varphi_1 + \mu_2'\varphi_2 + \cdots + \mu_n'\varphi_n = 0 \\ \mu_1'\varphi_1' + \mu_2'\varphi_2' + \cdots + \mu_n'\varphi_n' = 0 \\ \cdots \\ \mu_1'\varphi_1^{(n-2)} + \mu_2'\varphi_2^{(n-2)} + \cdots + \mu_n'\varphi_n^{(n-2)} = 0 \\ \mu_1'\varphi_1^{(n-1)} + \mu_1'\varphi_2^{(n-1)} + \cdots + \mu_n'\varphi_n^{(n-1)} = \frac{1}{a_n} f(t) \end{cases}$$

On peut retenir cette méthode sous la forme suivante. On part de la formule (p) que l'on dérive n fois en imposant, dans les $n-1$ premiers calculs que les dérivées des μ_j n'apparaissent pas. Cela donne les $n-1$ premières équations. La dernière est obtenue en remplaçant chaque $\varphi_j^{(n)}$ par sa valeur tirée de (e).

3 Équation aux variations

On reprend la situation du chapitre 6 et on donne une preuve partielle d'une version améliorée d'une partie du dernier théorème (!). Soit $U = I \times V$ un ouvert de $\mathbb{R} \times \mathbb{R}^n$ et $f : U \to \mathbb{R}^n$ une application qu'on suppose de classe $\mathcal{C}^1$. On appelle (E) l'équation différentielle associée à f. Dans tout ce paragraphe $t_0 \in I$ est supposé fixé et on adopte les notations suivantes. Pour $y \in V$, J_y désigne l'intervalle (ouvert) sur lequel est définie la solution maximale de (E) de condition initiale (t_0, y), notée $\varphi(t, y)$. Donc pour tout $y \in V$, $\varphi(t_0, y) = y$.

On pose

$$\Omega = \{(t, y) \in I \times V \mid t \in J_y\}$$

et on définit $\varphi : \Omega \to \mathbb{R}^n$ par $(t, y) \mapsto \varphi(t, y)$. L'ensemble Ω est donc l'« ensemble de définition » de φ et, par construction, φ admet en tout point de Ω, une différentielle partielle par rapport à t qui est l'application linéaire $\mathbb{R} \to \mathbb{R}^n$ donnée par $u \mapsto u\dfrac{\partial \varphi}{\partial t}(t, y)$ où, pour $t \in J_y$,

$$(1) \quad \frac{\partial \varphi}{\partial t}(t, y) = f(t, \varphi(t, y)).$$

On se propose d'étudier la régularité de φ, en particulier sa dépendance pa rapport à y. Pour ce dernier point, remarquons, que, en supposant que cela a un sens, on obtient en dérivant (1) par rapport à y,

$$D_2\left(\frac{\partial\varphi}{\partial t}\right)(t,y) = D_2f(t,\varphi(t,y)) \circ D_2\varphi(t,y).$$

En utilisant le théorème de Schwarz ceci s'écrit aussi, pour tout $y \in V$,

$$\frac{\partial}{\partial t}(D_2\varphi(t,y)) = D_2f(t,\varphi(t,y)).D_2\varphi(t,y) \qquad (2)$$

que l'on peut encore interprêter de la façon suivante. Définissons les fonctions $J_y \to \mathcal{M}(n,\mathbb{R})$ par $X(t) = D_2\varphi(t,y)$ et $A(t) = D_2f(t,\varphi(t,y))$. Puisque, pour tout $y \in V$, $\varphi(t_0,y) = y$, on a $D_2\varphi(t_0,y) = Id_{\mathbb{R}^n}$ et donc $X(t_0) = I_n$. Jointe à la formule (2) cette propriété traduit le fait que $X(t)$ est sur J_y la résolvante $\mathcal{R}(t,t_0)$ de l'équation *linéaire* $X' = A(t)X$. Cette équation gère « la variation » de la fonction φ dans la direction y en fonction de celle de f dans la direction y mais « le long de la solution $\varphi(t,y)$ ». Elle porte le nom d'*équation aux variations* (ou parfois variationnelle) le long de la solution $\varphi(t,y)$.

Théorème 3.1 *Avec les notations et hypothèses ci-dessus,*

1. *Ω est un ouvert de U,*
2. *φ est de classe $\mathcal{C}^1$,*
3. *pour tout $y_0 \in V$, et tout $t \in J_{y_0}$, $D_2\varphi(t,y_0)$ est la valeur en t de la solution de condition initiale (t_0,I_n) de l'équation aux variations $X' = A(t)X$ où $A(t) = D_2f(t,\varphi(t,y_0))$.*

Preuve Remarquons d'abord que le théorème de Cauchy-Lipschitz global s'applique en tout point de U.

Commençons par traiter le cas particulier suivant. On dit que f vérifie l'hypothèse (H) en y_0 sur I_0, intervalle compact dont t_0 est l'une des bornes, si $y_0 \in V$, $I_0 \subset I$ et, pour tout $t \in I_0$, $f(t,y_0) = 0$ et $D_2f(t,y_0) = 0$.
Dans ce cas la fonction constante égale à y_0 sur I_0 est clairement une solution de (E). Donc $J_{y_0} \supset I_0$ et $\varphi(t,y_0) = y_0$ sur I_0.
L'hypothèse (H) jointe à la *compacité* de I_0 permettent, pour tout $\varepsilon > 0$, de déterminer $\delta > 0$ tel que pour tout $(t,y) \in I_0 \times Bf(y_0,\delta)$, on ait $\|D_2f(t,y)\| \leq \varepsilon$.
Le théorème des accroissements finis (sur le *convexe* $Bf(y_0,\delta)$) permet d'en déduire, pour tout $t \in I_0$ et pour $y_1, y_2 \in Bf(y_0,\delta)$,

$$\|f(t,y_1) - f(t,y_2)\| \leq \varepsilon\|y_1 - y_2\|.$$

En particulier, pour tout $(t,y) \in I_0 \times Bf(y_0,\delta)$, on a $\|f(t,y)\| \leq \varepsilon \|y-y_0\|$. Posons $\mathcal{E} = \mathcal{C}^0(I_0, Bf(y_0,\delta))$ muni de la distance d_∞. C'est un espace métrique (mais pas vectoriel !) *complet*.
Pour $y \in Bf(y_0, \frac{\delta}{2})$, définissons une application Φ_y sur $\mathcal{E}$ en posant pour $g \in \mathcal{E}$,

$$\Phi_y(g)(t) = y + \int_{t_0}^{t} f(s, g(s))\, ds.$$

La fonction $\Phi_y(g)$ est clairement définie sur I_0 et, puisque $\|f(s,g(s))\| \leq \varepsilon \|g(s)-y_0\| \leq \varepsilon \frac{\delta}{2}$, elle vérifie, pour tout $t \in I_0$,

$$\|\Phi_y(g)(t) - y_0\| \leq \|y-y_0\| + \| \int_{t_0}^{t} f(s,g(s))\, ds\| \leq \frac{\delta}{2} + \varepsilon \ell \frac{\delta}{2}$$

si ℓ désigne la longueur du segment I_0. Supposons désormais ε assez petit pour que $\ell\varepsilon < 1$. Alors pour tout $t \in I_0$, $\Phi_y(g)(t) \in Bf(y_0,\delta)$ et, comme $\Phi_y(g)$ est clairement une fonction continue (et même $\mathcal{C}^1$) de t, on en déduit que $\Phi_y : \mathcal{E} \to \mathcal{E}$.
De plus, si $g_1, g_2 \in \mathcal{E}$ et $t \in I_0$,

$$\begin{aligned} \|\Phi_y(g_1)(t) - \Phi_y(g_2)(t)\| &= \left\| \int_{t_0}^{t} [f(s,g_1(s)) - f(s,g_2(s))]\, ds \right\| \\ &\leq \left| \int_{t_0}^{t} \varepsilon \|g_1(s) - g_2(s)\|\, ds \right| \leq \varepsilon \ell d_\infty(g_1,g_2) \end{aligned}$$

et donc

$$d_\infty(\Phi_y(g_1), \Phi_y(g_2)) \leq \varepsilon \ell d_\infty(g_1,g_2).$$

Comme $\ell\varepsilon < 1$, cela montre que Φ_y est contractante et a donc un unique point fixe $\tilde{g}_y \in \mathcal{E}$. La fonction continue $\tilde{g}_y$ vérifie donc, pour tout $t \in I_0$,

$$\tilde{g}_y(t) = y + \int_{t_0}^{t} f(s, \tilde{g}_y(s))\, ds$$

ce qui montre d'une part que $\tilde{g}_y(t_0) = y$ et d'autre part que $\tilde{g}_y$ est dérivable dans I_0, sa dérivée vérfiant dans I_0, $\tilde{g}_y'(t) = f(t, \tilde{g}_y(t))$.
Le théorème de Cauchy-Lipschitz et la définition de φ imposent alors $J_y \supset I_0$ et $\boxed{\tilde{g}_y(t) = \varphi(t,y)_{|_{I_0}}}$. Puisque y est arbitraire dans $Bf(y_0, \frac{\delta}{2})$, la définition de Ω montre alors que $\boxed{\Omega \supset I_0 \times Bf(y_0, \frac{\delta}{2})}$.

De plus, toujours d'après le théorème du point fixe, si on définit $g_n \in \mathcal{E}$ par récurrence en posant $g_0(t) = y_0$ (fonction constante) et, pour $n \geq 1$,

$g_n = \Phi_y(g_{n-1})$, alors pour tout $n \geq 1$,

$$(\star)_n \quad d_\infty(\tilde{g}_y, g_n) \leq \frac{(\varepsilon\ell)^n}{1-\varepsilon\ell}\, d_\infty(g_1, g_0).$$

On a $g_1(t) = y + \displaystyle\int_{t_0}^{t} f(s, y_0)\, ds = y$ (ne pas oublier les hypothèses de ce cas particulier...) et g_1 est donc aussi une fonction constante. La relation $(\star)_1$ s'écrit alors

$$d_\infty(\tilde{g}_y, g_1) = \sup_{t\in I_0} \|\varphi(t,y) - y\| \leq \frac{\varepsilon\ell}{1-\varepsilon\ell}\, \|y - y_0\|$$

et on en déduit (en se souvenant que $\varphi(t, y_0) = y_0$) que, pour tout $t \in I_0$,

$$\frac{\|\varphi(t,y) - \varphi(t,y_0) - (y - y_0)\|}{\|y - y_0\|} \to 0$$

quand $y \to y_0$. Autrement dit, pour tout $t \in I_0$, la fonction $y \mapsto \varphi(t, y)$ est différentiable en y_0 et sa différentielle en y_0 est l'application identité. A titre d'exercice, on pourra s'assurer que c'est là ce qu'on voulait démontrer dans ce cas particulier. Remarquons encore que la preuve a utilisé seulement le fait que f est continue dans U et admet une différentielle partielle par rapport à y qui est continue (dans U).

Passons maintenant au cas général. Bien sûr on cherche à se ramener au cas précédent. Soit $y_0 \in V$. Sur J_{y_0} on considère le système différentiel linéaire $X' = A(t)X$ de matrice $A(t) = D_2 f(t, \varphi(t, y_0))$ et on note $\mathcal{R}(t, t_0)$ sa résolvante, qui est définie sur J_{y_0}. A toute fonction (dérivable) φ définie sur un intervalle inclus dans J_{y_0}, on associe la fonction ψ donnée sur le même intervalle par

$$(*) \quad \psi(t) = \mathcal{R}(t_0, t).(\varphi(t) - \varphi(t, y_0))$$

Puisque $\mathcal{R}(t_0, t)$ est inversible, on peut tout aussi bien partir de ψ et définir φ à partir de $(*)$. Si on suppose que φ est solution de (E) alors ψ sera solution sur le même intervalle d'une équation $(\tilde{E})$ que l'on va déterminer. Puisque $\varphi(t) = \varphi(t, y_0) + \mathcal{R}(t, t_0).\psi(t)$, on a d'une part

$$\varphi'(t) = f(t, \varphi(t, y_0)) + A(t)\mathcal{R}(t, t_0)\psi(t) + \mathcal{R}(t, t_0)\psi'(t)$$

D'autre part $f(t, \varphi(t)) = f(t, \varphi(t, y_0) + \mathcal{R}(t, t_0).\psi(t))$ et donc φ est solution de (E) si et seulement si

$$\psi'(t) = \mathcal{R}(t_0, t).[f(t, \varphi(t, y_0) + \mathcal{R}(t, t_0).\psi(t)) - f(t, \varphi(t, y_0)) - A(t)\mathcal{R}(t, t_0)\psi(t)]$$

c'est-à-dire si ψ est solution de l'équation $(\tilde{E})$ associée à la fonction $\tilde{f}$ donnée par :

$$\tilde{f}(t,z) = \mathcal{R}(t_0,t).[f(t,\varphi(t,y_0) + \mathcal{R}(t,t_0).z) - f(t,\varphi(t,y_0)) - A(t)\mathcal{R}(t,t_0)z)]$$

Cette fonction est définie si $t \in J_{y_0}$ et si $\varphi(t,y_0) + \mathcal{R}(t,t_0).z \in V$.
Fixons alors $t_1 \in J_{y_0}$ et appelons I_0 le segment d'extrêmités t_0 et t_1. Il existe un intervalle ouvert $\tilde{J} \supset I_0$ et un réel $\rho > 0$ tels que $\tilde{f}$ soit définie sur $\tilde{U} = \tilde{J} \times Bo(0,\rho)$. Cela découle des trois remarques : 1) I_0 est compact et pour tout $t \in I_0$, $\tilde{f}$ est définie en $(t,0)$ puisque $\varphi(t,y_0) + \mathcal{R}(t,t_0).0 = \varphi(t,y_0) \in V$, 2) V est ouvert et 3) toutes les fonctions en jeu sont continues.
Sur $\tilde{U}$ la fonction $\tilde{f}$ est continue, admet une différentielle partielle continue par rapport à z et vérifie l'hypothèse (H) en 0 sur I_0. En effet, pour tout $t \in \tilde{J}$ (a fortiori dans I_0),

- $\tilde{f}(t,0) = \mathcal{R}(t_0,t).[f(t,\varphi(t,y_0)) - f(t,\varphi(t,y_0))] = 0$
- $D_2\tilde{f}(t,0) = \mathcal{R}(t_0,t).[D_2f(t,\varphi(t,y_0)).\mathcal{R}(t,t_0) - A(t)\mathcal{R}(t,t_0))] = 0.$

En notant $\tilde{\Omega}$ et $\tilde{\varphi}$ les analogues de Ω et φ pour $(\tilde{E})$, l'étude du cas particulier montre ce qui suit. Il existe $\delta > 0$ (avec $\delta < \rho$) tel que $\tilde{\Omega} \supset I_0 \times Bo(0,\delta)$ et pour tout $t \in I_0$, l'application $z \mapsto \tilde{\varphi}(t,z)$ est différentiable en 0, sa différentielle étant l'application identité.
La formule $\varphi(t) = \varphi(t,y_0) + \mathcal{R}(t,t_0).\tilde{\varphi}(t,z)$ définit alors une solution de (E) sur I_0 qui vérifie $\varphi(t_0) = y_0 + \mathcal{R}(t_0,t_0).z = y_0 + z$ et donc, pour tout $t \in I_0$,

$$\varphi(t) = \varphi(t,y_0+z) = \varphi(t,y_0) + \mathcal{R}(t,t_0).\tilde{\varphi}(t,z).$$

On en déduit les deux affirmations :

- $\Omega \supset I_0 \times Bo(y_0,\delta)$ et est donc ouvert,
- si $t \in I_0$, la fonction $\varphi(t,y)$ admet en (t,y_0) une différentielle partielle par rapport à y et $D_2\varphi(t,y_0) = \mathcal{R}(t,t_0).Id_{\mathbb{R}^n} = \mathcal{R}(t,t_0)$.

Il n'y a plus qu'à remarquer que I_0 contient t_1 arbitrairement choisi dans J_{y_0} et à se souvenir de la définition de la résolvante. Bien sûr il faudrait encore montrer que les dérivées partielles $D_1\varphi(t,y)$ et $D_2\varphi(t,y)$ sont continues dans Ω mais l'auteure déclare forfait...