Calcul différentiel 1

Licence de Mathématiques

Table des matières

Avertissement

Dans tout ce qui suit, on considérera uniquement des espaces vectoriels *réels*. Autrement dit, l'expression "espace vectoriel" signifiera toujours "$\mathbb{R}$-espace vectoriel".

CHAPITRE 1

Préliminaires

1. Espaces vectoriels normés

1.1. Normes sur un espace vectoriel.

DÉFINITION 1.1. *Une **norme** sur un espace vectoriel E est une application de E dans $\mathbb{R}$, notée $u \mapsto \|u\|$ ou $\|\cdot\|$, vérifiant les propriétés suivantes :*

(1) $\|u\| \geq 0$ *pour tout* $u \in E$, *et* $\|u\| = 0$ *seulement pour* $u = 0$ *;*

(2) $\|\lambda u\| = |\lambda| \, \|u\|$ *pour tout* $u \in E$ *et pour tout* $\lambda \in \mathbb{R}$ *;*

(3) $\|u + v\| \leq \|u\| + \|v\|$ *pour tous* $u, v \in E$ *(**inégalité triangulaire**).*

*Un **espace vectoriel normé** est un espace vectoriel muni d'une norme. Pour abréger, on écrira en général **evn** au lieu de "espace vectoriel normé".*

Remarque. Il arrivera souvent qu'on considère plusieurs espaces vectoriels normés E, F, G ... en même temps. En général, on utilisera la même notation $\|\cdot\|$ pour la norme de chacun des espaces. S'il est absolument nécessaire d'être plus précis, on écrira $\|\cdot\|_E$, $\|\cdot\|_F$, $\|\cdot\|_G$, ...

Exemple 1. La valeur absolue est une norme sur $E = \mathbb{R}$, le module est une norme sur $E = \mathbb{C}$. On supposera toujours que $\mathbb{R}$ et $\mathbb{C}$ sont munis de ces normes "canoniques".

Exemple 2. (normes usuelles sur $\mathbb{R}^n$)
Soit $n \in \mathbb{N}^*$. On définit des normes $\|\cdot\|_\infty$, $\|\cdot\|_1$ et $\|\cdot\|_2$ sur $E = \mathbb{R}^n$ en posant, pour $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$:

$$\|\mathbf{x}\|_\infty = \max(|x_1|, \dots, |x_n|)\,,$$

$$\|\mathbf{x}\|_1 = \sum_{j=1}^{n} |x_j|\,,$$

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + \cdots + x_n^2}\,.$$

Dans la suite, on utilisera principalement la norme $\|\cdot\|_\infty$.

Remarque. Même si on aura tendance à privilégier la norme $\|\cdot\|_\infty$, il est bon d'ajouter quelques mots sur la norme $\|\cdot\|_2$. Cette norme s'appelle la **norme euclidienne**. Par définition, on a $\|\mathbf{x}\|_2 = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$, où $\langle\ ,\ \rangle$ est le produit scalaire usuel sur $\mathbb{R}^n$. La norme euclidienne et le produit scalaire sont également reliés par la très importante **inégalité de Cauchy-Schwarz** : si $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, alors

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\|_2 \, \|\mathbf{y}\|_2\,.$$

Il n'est pas évident que $\|\cdot\|_2$ est effectivement une norme : l'inégalité triangulaire $\|\mathbf{x}+\mathbf{y}\|_2 \leq \|\mathbf{x}\|_2 + \|\mathbf{y}\|_2$ n'est pas "immédiate". On peut la démontrer en "développant" $\|\mathbf{x} + \mathbf{y}\|_2^2 = \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle$ et en utilisant l'inégalité de Cauchy-Schwarz. Enfin, pour $n = 2$, la norme euclidienne n'est rien d'autre que la norme "module" sur $\mathbb{C} = \mathbb{R}^2$.

Exemple 3. Soit $[a,b]$ un intervalle fermé borné de $\mathbb{R}$, et soit $\mathcal{C}([a,b])$ l'espace vectoriel constitué par toutes les fonctions continues $u:[a,b]\to\mathbb{R}$. On définit une norme $\|\cdot\|_\infty$ sur $\mathcal{C}([a,b])$ en posant

$$\|u\|_\infty = \sup\{|u(t)|;\ t\in[a,b]\}\,.$$

Exemple 4. (espace produit)
Soient $E_1,\dots,E_n$ des evn, et soit $E=E_1\times\cdots\times E_n$. Écrivons chaque vecteur $u\in E$ sous la forme $u=(u(1),\dots,u(n))$, où $u(j)\in E_j$. Alors la formule

$$\|u\| = \max\Big(\|u(1)\|_{E_1},\dots,\|u(n)\|_{E_n}\Big)$$

définit une norme sur $E=E_1\times\cdots\times E_n$, qu'on appelle la **norme produit**. Chaque fois qu'on considérera un espace produit $E_1\times\cdots\times E_n$, on supposera implicitement qu'il est muni de la norme produit. (Ceci est en accord avec le choix de la norme $\|\cdot\|_\infty$ sur $\mathbb{R}^n=\mathbb{R}\times\cdots\times\mathbb{R}$).

Exercice 1. Démontrer ce qui est dit dans les exemples 2, 3 et 4.

Exercice 2. Soit E un evn. Montrer que si $u\in E$ et $u\neq 0$ (donc $\|u\|\neq 0$), alors

$$\left\|\frac{u}{\|u\|}\right\| = 1\,.$$

Exercice 3. Soit E un evn. Montrer que si $u,v\in E$, alors :

$$\Big|\|u\|-\|v\|\Big| \le \|u-v\| \le \|u\|+\|v\|.$$

La première de ces inégalités s'appelle l'**inégalité triangulaire inverse**.

1.2. Normes équivalentes.

Définition 1.2. *Soient $\|\cdot\|$ et $\|\cdot\|'$ deux normes sur un espace vectoriel E. On dit que $\|\cdot\|$ et $\|\cdot\|'$ sont* ***équivalentes*** *si on peut trouver deux constantes $c>0$ et $C<\infty$ telles que*

$$\forall u\in E\ :\ c\,\|u\|'\le\|u\|\le C\,\|u\|'\,.$$

Remarque. La terminologie n'est pas délirante : on définit bien ainsi une *relation d'équivalence* sur l'ensemble de toutes les normes sur E. (*Exercice* : vérifier).

Exercice 1. Montrer que les normes $\|\cdot\|_\infty$, $\|\cdot\|_1$ et $\|\cdot\|_2$ sur $\mathbb{R}^n$ sont équivalentes.

Exercice 2. Pour $u\in\mathcal{C}([0,1])$, on pose $\|u\|_1=\int_0^1|u(t)|\,dt$. Montrer que $\|\cdot\|_1$ est une norme sur $\mathcal{C}([0,1])$, mais qu'elle n'est pas équivalente à $\|\cdot\|_\infty$.

Le théorème suivant signifie en gros que lorsqu'on est en dimension finie (et pour ce qui va nous intéresser), le choix de la norme n'a aucune importance. Il s'agit évidemment d'un résultat très utile, et donc très important. Il n'est pas nécessaire de savoir refaire la démonstration, qui est un peu compliquée. En revanche, il est *essentiel* de connaitre l'énoncé et de savoir l'utiliser.

Théorème 1.3. *Sur un espace vectoriel* de dimension finie*, toutes les normes sont équivalentes.*

Démonstration. Tout espace vectoriel de dimension finie étant isomorphe à $\mathbb{R}^n$ pour un certain $n\ge 1$, il s'agit de montrer que toutes les normes sur $\mathbb{R}^n$ sont équivalentes. Pour cela, il suffit de prouver que toute norme sur $\mathbb{R}^n$ est équivalente à $\|\cdot\|_\infty$.

Étape 1. Si $\|\cdot\|$ est une norme sur $\mathbb{R}^n$, il existe une constante $C < \infty$ telle que $\|\cdot\| \le C\|\cdot\|_\infty$.

Démonstration. Si $u = (u(1), \dots, u(n)) \in \mathbb{R}^n$ et si on note $(e_1, \dots, e_n)$ la base canonique de $\mathbb{R}^n$, alors

$$\begin{aligned}
\|u\| &= \left\|\sum_{j=1}^n u(j)e_j\right\| \\
&\le \sum_{j=1}^n \|u(j)e_j\| \\
&= \sum_{j=1}^n |u(j)|\,\|e_j\| \\
&\le \|u\|_\infty \times \sum_{j=1}^n \|e_j\|\,,
\end{aligned}$$

puisque $|u(j)| \le \|u\|_\infty$ pour tout j. On obtient donc le résultat souhaité avec $C := \sum_1^n \|e_j\|$ (qui est bien une constante indépendante de $u \in \mathbb{R}^n$). □

Étape 2. Si $\|\cdot\|$ est une norme sur $\mathbb{R}^n$, il existe une constante $c > 0$ telle que $\|\cdot\| \ge c\|\cdot\|_\infty$.

Démonstration. C'est la partie "difficile". On va démontrer le résultat par récurrence sur n.

Pour $n = 1$, ce n'est pas compliqué : si $\|\cdot\|$ est une norme sur $\mathbb{R}^1 = \mathbb{R}$, alors $\|u\| = \|u \times 1\| = |u| \times \|1\| = c\,\|u\|_\infty$ pour tout $u \in \mathbb{R}^1$, où $c = \|1\|$ est bien *strictement* positif car $1 \neq 0$.

Supposons le résultat vrai pour $n \ge 1$, et démontrons le pour $n+1$. Soit $\|\cdot\|$ une norme sur $\mathbb{R}^{n+1}$. Par l'absurde, on suppose qu'il n'existe pas de constante $c > 0$ telle que $\|\cdot\| \ge c\|\cdot\|_\infty$. Cela signifie que pour tout $c > 0$, on peut trouver un vecteur $u^c \in \mathbb{R}^{n+1}$ tel que $\|u^c\| < c\,\|u^c\|_\infty$. On a $\|u^c\|_\infty \neq 0$ (d'après l'inégalité stricte), donc on peut définir $v^c = \frac{u^c}{\|u^c\|_\infty}\cdot$ Alors $\|v^c\|_\infty = 1$ et

$$\|v^c\| = \frac{1}{\|u^c\|_\infty} \times \|u^c\| < c\,.$$

En prenant $c = 1/k$ pour tout $k \in \mathbb{N}^*$ et en posant $v_k = v^{1/k}$, on obtient ainsi une suite $(v_k) \subset \mathbb{R}^{n+1}$ telle que $\|v_k\|_\infty = 1$ pour tout k et $\|v_k\| \to 0$.

Écrivons $v_k = (v_k(1), \dots, v_k(n+1))$. Par définition de $\|\cdot\|_\infty$, on peut choisir pour tout k un indice $j_k \in \{1, \dots, n\}$ tel que $|v_k(j_k)| = \|v_k\|_\infty = 1$. Comme il y a une infinité d'entiers k et un nombre fini d'indices $j \in \{1, \dots, n\}$, il existe au moins un $j \in \{1, \dots, n\}$ tel que $j_k = j$ pour une infinité de k. On peut par exemple supposer que $j = n+1$, et on obtient ainsi une sous-suite (v'_k) de (v_k) telle que $|v'_k(n+1)| = 1$ pour tout k, autrement dit $v'_k(n+1) = \pm 1$. De même, il y a une infinité de k tels que $v'_k(n+1) = 1$, ou bien une infinité de k tels que $v'_k(n+1) = -1$. On peut par exemple supposer qu'on est dans le premier cas, ce qui donne une sous-suite (v''_k) de (v'_k) (et donc une sous-suite de (v_k)) telle que $v''_k(n+1) = 1$ pour tout k. Pour simplifier les notations, on change le nom de v''_k qu'on appelle à nouveau v_k. Au total, on a ainsi trouvé une suite $(v_k) \subset \mathbb{R}^{n+1}$ telle que $v_k(n+1) = 1 = \|v_k\|_\infty$ pour tout k et $\|v_k\| \to 0$.

Le point clé est maintenant le suivant : si $p, q \in \mathbb{N}^*$, alors $(v_q - v_p)(n+1) = 0$. *Donc*, on peut considérer $v_q - v_p$ comme un vecteur de $\mathbb{R}^n$, en identifiant $\mathbb{R}^n$ à $\mathbb{R}^n \times \{0\} \subset \mathbb{R}^{n+1}$.

Par hypothèse de récurrence, il existe une constante $c' > 0$ telle que $\|u\| \geq c' \|u\|_\infty$ pour tout $u \in \mathbb{R}^n = \mathbb{R}^n \times \{0\}$. En prenant $u := v_q - v_p$, on en déduit qu'on a

$$|v_q(j) - v_p(j)| \leq (1/c') \|v_q - v_p\| \tag{1.1}$$

pour tous $p, q \in \mathbb{N}^*$ et pour tout $j \in \{1, \ldots, n\}$. D'autre part, on a également

$$\|v_q - v_p\| \leq \|v_q\| + \|v_p\|$$

d'après l'inégalité triangulaire, et donc $\|v_q - v_p\| \to 0$ quand $p, q \to \infty$. D'après (1.1), cela montre que pour tout $j \in \{1, \ldots, n\}$, la suite $(v_k(j))_{k \in \mathbb{N}^*}$ est une *suite de Cauchy* dans $\mathbb{R}$, et est donc convergente (critère de Cauchy). Pour chaque $j \in \{1, \ldots, n\}$, il existe donc un nombre réel $v(j)$ tel que $v_k(j) \to v(j)$ quand $k \to \infty$. Si maintenant on pose $v = (v(1), \ldots, v(n), 1) \in \mathbb{R}^{n+1}$, alors $v_k(j) \to v(j)$ pour tout $j \in \{1, \ldots, n+1\}$ puisque $v_k(n+1) \equiv 1$. Par définition de la norme $\|\cdot\|_\infty$, cela entraine que $\|v_k - v\|_\infty \to 0$ (*Exercice*). Et comme $\|\cdot\| \leq C \|\cdot\|_\infty$ pour une certaine constante C (d'après l'étape 1), on en déduit que $\|v_k - v\| \to 0$.

D'après l'inégalité triangulaire, on a

$$\|v\| \leq \|v - v_k\| + \|v_k\|$$

pour tout $k \in \mathbb{N}^*$. Comme $\|v_k\| \to 0$ et $\|v_k - v\| \to 0$, on en déduit $\|v\| \leq 0$ en faisant tendre k vers l'infini, et donc $\|v\| = 0$. Comme $\|\cdot\|$ est une norme, cela signifie que $v = 0$. Mais ceci est absurde puisque la $(n+1)$-ième coordonnée de v vaut 1. On a donc obtenu une contradiction, ce qui achève la démonstration par récurrence. □

Au total, les faits 1 et 2 donnent le théorème.

□

Exercice. Soit $N \in \mathbb{N}^*$. Déterminer les "meilleures" constantes c_n et C_n telles que $c_n \|u\|_\infty \leq \|u\|_2 \leq C_n \|u\|_\infty$ pour tout $u \in \mathbb{R}^n$.

2. Convergence, continuité

2.1. Suites convergentes.

DÉFINITION 2.1. *Soient E un evn, $(u_k)_{k \in \mathbb{N}}$ une suite d'éléments de E, et $u \in E$. On dit que la suite (u_k)* ***converge vers*** *u si $\|u_k - u\|$ tend vers 0 quand $k \to \infty$. Avec des quantificateurs, cela s'écrit*

$$\forall \varepsilon > 0 \, \exists K \, \forall k \geq K \, : \, \|u_k - u\| \leq \varepsilon \, .$$

(On pourrait tout aussi bien écrire "$< \varepsilon$" au lieu de "$\leq \varepsilon$" : pourquoi ?).

REMARQUE. Les suites convergentes restent les mêmes si on remplace la norme de E par une norme équivalente. En particulier, dans un espace vectoriel *de dimension finie*, les suites convergentes sont les mêmes quelle que soit la norme utilisée, et on peut donc parler de convergence sans faire explicitement référence à aucune norme.

Exercice. Montrer qu'on a *unicité de la limite* ; autrement dit, qu'une suite (u_k) ne peut pas converger vers deux points différents.

Exemple 1. Une suite réelle ou complexe converge pour la norme "valeur absolue" ou "module" si et seulement si elle converge au sens usuel.

Exemple 2. Une suite $(u_k) \subset \mathbb{R}^n$ converge dans $\mathbb{R}^n$ (pour n'importe quelle norme) si et seulement si elle converge "coordonnée par coordonnée". Autrement dit, si on écrit $u_k = (u_k(1), \dots, u_k(n))$ et $u = (u(1), \dots, u(n))$, alors $u_k \to u$ si et seulement si $u_k(j) \to u(j)$ pour tout $j \in \{1, \dots, n\}$.

DÉMONSTRATION. Par équivalence des normes, il suffit de le voir pour la norme $\| \cdot \|_\infty$. C'est un exercice facile, qu'il *faut absolument* savoir faire (et qui d'ailleurs a déjà été utilisé dans la preuve du théorème 1.3). $\square$

Exemple 3. Une suite $(u_k) \subset \mathcal{C}([a,b])$ converge pour la norme $\| \cdot \|_\infty$ si et seulement si elle converge *uniformément* sur $[a,b]$. Pour cette raison, la norme $\| \cdot \|_\infty$ s'appelle la **norme de la convergence uniforme**.

Exemple 4. Dans un evn produit $E = E_1 \times \cdots \times E_n$, une suite converge si et seulement si elle converge "coordonnée par coordonnée".

DÉMONSTRATION. Exercice (le même que dans l'exemple 2). $\square$

2.2. Le théorème de Bolzano-Weierstrass. Une suite (u_k) dans un evn E est dite **bornée** s'il existe une constante $M < \infty$ (*indépendante de* k) telle que $\|u_k\| \leq M$ pour tout $k \in \mathbb{N}$. Il n'est pas difficile de montrer que toute suite convergente est bornée (*exercice*). La réciproque est évidemment fausse : considérer par exemple $u_k = (-1)^k$ dans $E = \mathbb{R}$. Le très important résultat suivant donne cependant une réciproque partielle en dimension finie. Rappelons qu'une **sous-suite** de (u_k) est une suite (u'_k) de la forme $u'_k = u_{n_k}$, où (n_k) est une suite croissante d'entiers tendant vers l'infini.

THÉORÈME 2.2. (Bolzano-Weierstrass)
Toute suite bornée dans un evn de dimension finie *possède une sous-suite convergente.*

DÉMONSTRATION. Par équivalence des normes en dimension finie, il suffit de le voir lorsque l'evn est $\mathbb{R}^n$ muni de la norme $\| \cdot \|_\infty$.

Soit (u_k) une suite bornée dans $(\mathbb{R}^n, \| \cdot \|_\infty)$, et écrivons $u_k = (u_k(1), \dots, u_k(n))$. Fixons également une constante M telle que $\|u_k\|_\infty \leq M$ pour tout $k \in \mathbb{N}$.

Comme $|u_k(1)| \leq \|u_k\|_\infty \leq M$ pour tout k, la suite $(u_k(1))_{k\in\mathbb{N}}$ est bornée dans $\mathbb{R}$. D'après le théorème de Bolzano-Weierstrass usuel, on peut donc trouver une sous-suite (u'_k) de (u_k) et un nombre réel l_1 tels que $u'_k(1) \to l_1$ quand $k \to \infty$. Maintenant, la suite $(u'_k(2))$ est bornée dans $\mathbb{R}$ car $|u'_k(2)| \leq \|u'_k\|_\infty \leq M$, donc on peut trouver $l_2 \in \mathbb{R}$ et une sous-suite (u''_k) de (u'_k) tels que $u''_k(2) \to l_2$. Alors (u''_k) est une sous-suite de (u_k) telle que $u''_k(1) \to l_1$ et $u''_k(2) \to l_2$. En poursuivant ce raisonnement, on obtient après n extractions une sous-suite (v_k) de (u_k) et n nombres réels $l_1, \dots, l_n$ tels que $v_k(j) \to l_j$ pour tout $j \in \{1, \dots, n\}$. Si on pose $u = (l_1, \dots, l_n) \in \mathbb{R}^n$, alors (v_k) converge vers u "coordonnée par coordonnée", et donc au sens de la norme $\| \cdot \|_\infty$. $\square$

Exercice. Pour $k \in \mathbb{N}^*$, soit $u_k : [0,1] \to \mathbb{R}$ la fonction nulle sur les intervalles $[0, 1/(2k+1)]$ et $[1/(2k-1), 1]$, valant 1 au point $t = 1/2k$, et affine sur les intervalles $[1/(2k-1), 1/2k]$ et $[1/2k, 1/(2k+1)]$. Montrer que la suite (u_k) est bornée dans $(\mathcal{C}([0,1]), \| \cdot \|_\infty)$ mais ne possède aucune sous-suite convergente.

2.3. Applications continues.

DÉFINITION 2.3. *Soient E et F deux evn, et soit $A \subset E$. On dit qu'une application $f : A \to F$ est **continue en un point** $a \in A$ si "$f(u)$ tend vers $f(a)$ quand u tend vers a"; autrement dit (avec des quantificateurs) :*

$$\forall \varepsilon > 0 \, \exists \delta = \delta(a, \varepsilon) > 0 \, \forall u \in A \ : \ \|u - a\| < \delta \implies \|f(u) - f(a)\| \leq \varepsilon \, .$$

(C'est la même chose si on écrit "$\leq \delta$" et/ou "$< \varepsilon$"). On dit qu'une application $f : A \to F$ est ***continue sur*** *A si elle est continue en tout point $a \in A$.*

Remarque 1. Cette définition est exactement la même que pour les fonctions réelles d'une variable réelle : on remplace simplement les valeurs absolues par des normes.

Remarque 2. Les applications continues restent les mêmes si on remplace les normes par des normes équivalentes. En particulier, on peut parler d'applications continues entre deux evn de dimension finie sans faire explicitement référence à aucune norme.

Exemple 1. On dit qu'une application $f : A \to F$ est **lipschitzienne** s'il existe une constante $C < \infty$ telle que

$$\forall u, v \in A \;:\; \|f(v) - f(u)\| \leq C\,\|v - u\|\,.$$

(Dans ce cas, on dit que f est **C-lipschitzienne**). Il est immédiat que toute application lipschitzienne est continue sur a (on peut prendre $\delta(a, \varepsilon) = \varepsilon/C$ pour tout $a \in A$).

Exemple 2. Les "applications coordonnées" sont continues sur $\mathbb{R}^n$. Autrement dit, pour tout $j \in \{1, \dots, n\}$, l'application $\pi_j : \mathbb{R}^n \to \mathbb{R}$ définie par $\pi_j(x_1, \dots, x_n) = x_j$ est continue sur $\mathbb{R}^n$

Démonstration. Par équivalence des normes, on peut supposer que $\mathbb{R}^n$ est muni de la norme $\|\cdot\|_\infty$. Si $u = (x_1, \dots, x_n)$ et $v = (y_1, \dots, y_n)$, alors $|\pi_j(v) - \pi_j(u)| = |y_j - x_j| \leq \|v - u\|_\infty$, ce qui prouve que l'application π_j est lipschitzienne. $\square$

Exercice. Montrer que si E est un evn, alors l'application $u \mapsto \|u\|$ est continue sur E.

Le résultat suivant est souvent très utile.

Proposition 2.4. (caractérisation séquentielle de la continuité)
Pour une application $f : A \to F$, les propriétés suivantes sont équivalentes :
(i) *f est continue en un point $a \in A$;*
(ii) *pour toute suite $(u_k) \subset A$ convergeant vers a, la suite $(f(u_k))$ tend vers $f(a)$.*

Démonstration. Exercice (qu'il *faut absolument* savoir faire). $\square$

Corollaire 2.5. *Soit $f : A \to \mathbb{R}^m$, et écrivons $f(u) = \begin{pmatrix} f_1(u) \\ \vdots \\ f_m(u) \end{pmatrix}$. Alors f est continue si et seulement si les applications $f_1, \dots, f_m$ sont continues.*

Démonstration. C'est immédiat en utilisant des suites, puisque la convergence dans $\mathbb{R}^m$ est la convergence "coordonnée par coordonnée". $\square$

Remarque. Ce corollaire est important car il permet de se ramener au cas des fonctions à valeurs réelles (et non plus vectorielles).

Proposition 2.6. *La continuité est préservée par somme, produit (pour les fonctions à valeurs réelles ou complexes), quotient (quand il est défini), composition (quand elle a un sens).*

Démonstration. Autre exercice *à savoir faire absolument.* Le plus délicat est en fait le cas du produit : voir le cours d'analyse de 1ère année si besoin est, ou la preuve de la proposition 5.9. $\square$

Conséquence pratique. Soit $f(x_1,\dots,x_n)$ une "formule explicite" dépendant de n variables $x_1,\dots,x_n\in\mathbb{R}$ et n'utilisant que des fonctions "usuelles". Alors la fonction f est continue *sur son domaine de définition.*

Ce n'est pas un énoncé précis, mais il est clair que cela découle immédiatement de la proposition : toute "formule explicite" $f(x_1,\dots,x_n)$ est construite à partir des applications coordonnés (qui sont continues) en utilisant des fonctions usuelles (donc continues), des sommes, des produits et des compositions. *La* chose importante est qu'il faut déterminer précisément le domaine de définition de la fonction.

Exemple 1. La formule $f(x,y)=\frac{\log(x-y)}{e^{xy}-1}$ définit une fonction continue. Quel est son domaine de définition ?

Exemple 2. L'application $M\mapsto\det(M)$ est continue sur $M_n(\mathbb{R})$.

Démonstration. On identifie $M_n(\mathbb{R})$ à $\mathbb{R}^{n^2}$ en assimilant une matrice M à la liste de ses coefficients (après avoir choisi un ordre d'énumération). La formule définissant le déterminant montre que $\det(M)$ est une fonction polynomiale des coefficients de M, d'où la continuité. □

Remarque. Il faut tout de même faire un peu attention avec l'expression "formule explicite". Il ne doit y avoir qu'*une seule* formule : une définition "par cas" ne rentre pas dans la catégorie "formule explicite". Par exemple, on ne peut pas invoquer la proposition précédente pour justifier que la fonction $f:\mathbb{R}^2\to\mathbb{R}$ définie par $f(x,y)=\frac{x^2y}{x^2+y^2}$ si $(x,y)\neq(0,0)$ et $f(0,0)=0$ est continue sur $\mathbb{R}^2$.

Exercice 1. Soient E et F deux evn, et soit $f:E\to F$. Soit également Ω un *ouvert* de E (voir la section suivante). Montrer que f est continue en tout point de Ω si et seulement si la restriction de f à Ω est continue. Est-ce encore vrai si Ω est une partie quelconque de E ?

Exercice 2. Soit $f:\mathbb{R}^2\to\mathbb{R}$. Montrer que la fonction f est continue en $(0,0)$ si et seulement si $f(r\cos\theta,r\sin\theta)$ tend vers $f(0,0)$ *uniformément par rapport à* $\theta\in\mathbb{R}$ quand $r\to0^+$.

Exercice 3. Montrer que la fonction f de la remarque précédente est effectivement continue sur $\mathbb{R}^2$.

3. Vocabulaire topologique

3.1. Boules.

Définition 3.1. *Soit E un evn, $a\in E$ et $\varepsilon\geq0$.*

- *La **boule ouverte de centre** a **et de rayon** ε, notée $B(a,\varepsilon)$, est l'ensemble des points $u\in E$ vérifiant $\|u-a\|<\varepsilon$:*
$$B(a,\varepsilon)=\{u\in E;\ \|u-a\|<\varepsilon\}\,.$$
- *La boule **fermée** de centre a et de rayon ε, notée $\overline{B}(a,\varepsilon)$, est définie par*
$$\overline{B}(a,\varepsilon)=\{u\in E;\ \|u-a\|\leq\varepsilon\}\,.$$

Exemple 1. Si $E=\mathbb{R}$ muni de la valeur absolue, alors $B(a,\varepsilon)$ est l'intervalle ouvert $]a-\varepsilon,a+\varepsilon[$ et $\overline{B}(a,\varepsilon)=[a-\varepsilon,a+\varepsilon]$. Si $E=\mathbb{C}$ muni du module, alors $B(a,\varepsilon)$ est le disque ouvert de centre a et de rayon ε, et $\overline{B}(a,\varepsilon)$ est le disque fermé correspondant.

Exemple 2. Si $E = (\mathbb{R}^2, \|\cdot\|_\infty)$, alors $\overline{B}(a, \varepsilon)$ est le carré de centre a dont les côtés sont parallèles aux axes de coordonnée et de longueur 2ε.

Exercice 1. Dessiner la boule fermée $\overline{B}(0, 1)$ dans $E = (\mathbb{R}^2, \|\cdot\|_1)$.

Exercice 2. Montrer que dans tout evn E, une boule B (ouverte ou fermée) est toujours un ensemble **convexe** : si $u, v \in B$, alors le segment $[u, v]$ est entièrement contenu dans B. On rappelle que le segment $[u, v]$ est défini analytiquement par

$$[u, v] = \{(1-t)u + tv;\ t \in [0, 1]\}.$$

Exercice 3. Dans la définition des boules, la valeur $\varepsilon = 0$ est autorisée. Déterminer $B(a, 0)$ et $\overline{B}(a, 0)$.

3.2. Ouverts et fermés.

Définition 3.2. *Soit E un evn. On dit qu'un ensemble $\Omega \subset E$ est un* ***ouvert*** *de E si, pour tout point $a \in \Omega$, on peut trouver $r = r(a) > 0$ tel que $B(a, r) \subset \Omega$.*

Remarque 1. On obtient une définition équivalente en écrivant "$\overline{B}(a, r)$" au lieu de $B(a, r)$. Pourquoi ?

Remarque 2. Les ouverts restent les mêmes si on remplace la norme de E par une norme équivalente.

Démonstration. Exercice. $\square$

Exemple. Dans $\mathbb{R}$, tout intervalle ouvert (borné ou non) est un ouvert.

Exercice 1. Montrer que $\Omega = \{(x, y) \in \mathbb{R}^2;\ xy > 1\}$ est un ouvert de $\mathbb{R}^2$.

Exercice 2. Montrer que toute boule ouverte d'un evn E est un ouvert de E.

Proposition 3.3. *Soit E un evn. La famille des ouverts de E vérifie les propriétés suivantes :*

(0) *$\emptyset$ et E sont ouverts ;*
(1) *toute réunion d'ouverts est un ouvert ;*
(2) *toute intersection* finie *d'ouverts est un ouvert.*

Démonstration. Un exercice qu'il faut savoir faire. $\square$

Remarque. Si E est un ensemble quelconque, une famille de parties de E vérifiant (0), (1) et (2) s'appelle une **topologie** sur E. C'est ce qui explique le titre de cette section ("vocabulaire topologique").

Définition 3.4. *Soit E un evn. On dit qu'un ensemble $C \subset E$ est un* ***fermé*** *de E s'il possède la propriété suivante : chaque fois qu'une suite $(u_k) \subset C$ converge dans E, sa limite appartient encore à C.*

Exemple. Dans $\mathbb{R}$, tout intervalle fermé (borné ou non, éventuellement réduit à un point) est un fermé.

Exercice. Montrer que $\emptyset$, E et toute boule fermée de E sont des fermé de E.

Proposition 3.5. *Soit E un evn. Un ensemble $C \subset E$ est fermé si et seulement si son complémentaire $E \setminus C$ est ouvert.*

Démonstration. Encore un exercice qu'il faut savoir faire. $\square$

COROLLAIRE 3.6. *Si E est un evn, alors la famille des fermés de E est stable par intersections quelconques et réunions* finies.

DÉMONSTRATION. C'est immédiat par passage aux complémentaires, en utilisant la proposition 3.3. □

La proposition suivante caractérise la continuité en termes d'ouverts et/ou de fermés. Dans la pratique, c'est presque toujours ce résultat qu'on utilise pour vérifier qu'un ensemble est ouvert ou fermé.

PROPOSITION 3.7. *Soient E et F deux evn, et soit $\Phi : E \to F$. Les propriétés suivantes sont équivalentes :*
(i) *Φ est continue sur E ;*
(ii) *$\Phi^{-1}(V)$ est ouvert dans E, pour tout ouvert $V \subset F$;*
(ii') *$\Phi^{-1}(C)$ est fermé dans E, pour tout fermé $C \subset F$.*

DÉMONSTRATION. D'après la proposition 3.5 et le fait que $f^{-1}(F \setminus A) = E \setminus f^{-1}(A)$ pour tout ensemble $A \subset F$, il et clair que (ii) et (ii') sont équivalentes. L'équivalence de (i) et (ii) est un exercice qu'il faut savoir faire. □

Exemple 1. L'ensemble $\Omega = \{(x,y) \in \mathbb{R}^2;\ x \neq y\,, x+y > 3 \,\text{et}\, x^2+y^2 < 6\}$ est un ouvert de $\mathbb{R}^2$, et l'ensemble $C = \{(x,y,z) \in \mathbb{R}^3;\ |x-y| \leq 5 \,\text{et}\, x^6 + y^2z^4 = 1\}$ est un fermé de $\mathbb{R}^3$.

DÉMONSTRATION. Soient Φ_1, Φ_2 et Φ_3 les trois fonctions de $\mathbb{R}^2$ dans $\mathbb{R}$ définies par $\Phi_1(x,y) = x-y$, $\Phi_2(x,y) = x+y$ et $\Phi_3(x,y) = x^2+y^2$. Ces trois fonctions sont continues (formules explicites), et on a

$$\Omega = \Phi_1^{-1}(\mathbb{R} \setminus \{0\}) \cap \Phi_2^{-1}(]3,+\infty[) \cap \Phi_3^{-1}(]-\infty,6[)\,.$$

Comme $\mathbb{R} \setminus \{0\}$, $]3,+\infty[$ et $]-\infty,6[$ sont des ouverts de $\mathbb{R}$, on voit donc que Ω apparait comme l'intersection de trois ouverts de $\mathbb{R}^2$. D'après la proposition 3.3, on en déduit que Ω est ouvert. On montre de la même manière que C est un fermé de $\mathbb{R}^3$, en l'écrivant comme intersection de deux fermés définis par des fonctions continues bien choisies. □

REMARQUE. Ce qu'il faut retenir de ces deux exemples est le "slogan" suivant : un ensemble défini par un nombre fini de "non-égalités" et d'inégalités *strictes* est ouvert, et un ensemble défini par un nombre fini d'égalités et d'inégalités *larges* est fermé.

Exemple 2. L'ensemble des matrices inversibles est un ouvert de $M_n(\mathbb{R})$.

DÉMONSTRATION. Si on pose $\Phi(M) = \det(M)$, alors la fonction Φ est continue sur $M_n(\mathbb{R})$ et l'ensemble des matrices inversibles est exactement $\Phi^{-1}(\mathbb{R} \setminus \{0\})$. □

Exercice. Donner un exemple d'ensemble $A \subset \mathbb{R}$ qui n'est ni ouvert ni fermé.

3.3. D'autres mots. On renvoie au cours de topologie pour la définition d'autres termes "topologiques" comme **voisinage**, **intérieur**, **adhérence** et **frontière**. Si A est une partie d'un evn E, on notera $\mathring{A}$ son intérieur, $\overline{A}$ son adhérence, et ∂A sa frontière.

Exercice 1. Quelle est la frontière de $\Omega = \{(x,y) \in \mathbb{R}^2;\ xy > 1\}$ dans $\mathbb{R}^2$?

Exercice 2. Montrer que dans un evn E, l'adhérence d'une boule ouverte de rayon $\varepsilon > 0$ est la boule fermée correspondante.

4. Compacité, connexité, complétude

4.1. Parties compactes d'un evn.

DÉFINITION 4.1. *Soit E un evn, et soit $K \subset E$. On dit que K est* ***compact*** *si, de* toute *suite $(u_k) \subset K$, on peut extraire une sous-suite convergente* dont la limite appartient à K.

Remarque 1. Les ensembles compacts restent les mêmes si on remplace la norme par une norme équivalente. On peut donc parler de compacts dans un espace vectoriel de dimension finie sans faire explicitement référence à une norme.

Remarque 2. Tout compact $K \subset E$ est *fermé* dans E et **borné** (il existe une constante M telle que $\forall u \in K \, : \, \|u\| \leq M$).

DÉMONSTRATION. Exercice à savoir faire. □

PROPOSITION 4.2. *Dans un evn* de dimension finie*, les ensembles compacts sont exactement les ensembles fermés et bornés.*

DÉMONSTRATION. Soit E un evn de dimension finie. D'après la remarque 2 ci-dessus, il suffit de montrer que si $K \subset E$ est fermé et borné, alors K est compact. Soit donc (u_k) une suite quelconque d'éléments de K. Alors (u_k) est bornée car K est borné. Comme E est de dimension finie, on peut appliquer le théorème de Bolzano-Weierstrass (théorème 2.2) : la suite (u_k) possède une sous-suite convergente (u'_k). De plus, comme K est fermé dans E, la limite de la suite (u'_k) appartient à K. □

Remarque. En dimension finie, pour montrer qu'un ensemble est borné, on peut choisir la norme qu'on veut par équivalence des normes.

Exemple 1. L'ensemble $K = \{(x, y) \in \mathbb{R}^2; \; x^2 + 3y^2 \leq 1\}$ est un compact de $\mathbb{R}^2$.

DÉMONSTRATION. Il est "clair" que K est fermé (il est défini par une inégalité large). De plus, si $u = (x, y) \in K$ alors $x^2 \leq 1$ et $y^2 \leq 1/3$ (car la somme de 2 termes positifs est plus grande que chacun des 2 termes), autrement dit $|x| \leq 1$ et $|y| \leq 1/\sqrt{3}$. Donc $\|u\|_\infty \leq 1$ pour tout $u \in K$, et par conséquent, K est borné. □

Exemple 2. L'ensemble des matrices orthogonales est un compact de $M_n(\mathbb{R})$.

DÉMONSTRATION. Par définition, une matrice M est orthogonale si et seulement ${}^tMM = I$. Si on pose $\Phi(M) = {}^tMM$, alors Φ est une application continue de $M_n(\mathbb{R})$ dans $M_n(\mathbb{R})$ (les coefficients de $\Phi(M)$ sont des fonctions polynomiales de ceux de M), et $O_n(\mathbb{R}) = \Phi^{-1}(\{I\})$. Par conséquent, $O_n(\mathbb{R})$ est un fermé de $M_n(\mathbb{R})$. D'autre part, on sait que si M est une matrice orthogonale, alors ses vecteurs colonnes sont de norme euclidienne égale à 1. En particulier, tous les coefficients de M sont au plus égaux à 1 en valeur absolue, et donc $\|M\|_\infty \leq 1$, où la norme $\| \cdot \|_\infty$ est définie par $\|(a_{ij})\|_\infty = \max_{i,j} |a_{ij}|$. Par conséquent, $O_n(\mathbb{R})$ est borné dans $M_n(\mathbb{R})$. □

Exercice 1. Soit $E = (\mathcal{C}([0, 1], \| \cdot \|_\infty)$. Donner un exemple d'ensemble fermé et borné dans E qui ne soit pas compact.

Exercice 2. Montrer que si (x_n) est une suite convergente dans un evn E et si on pose $x_\infty = \lim x_n$, alors l'ensemble $K = \{x_\infty\} \cup \{x_n; \; n \in \mathbb{N}\}$ est un compact de E.

Exercice 3. Montrer que si K_1 et K_2 sont deux compacts dans des evn E_1 et E_2, alors $K_1 \times K_2$ est compact dans $E_1 \times E_2$.

L'importance de la compacité vient des deux théorèmes suivants.

THÉORÈME 4.3. *Soit K un compact d'un evn E. Si $f : K \to \mathbb{R}$ est une fonction continue, alors f est bornée et atteint ses bornes.*

DÉMONSTRATION. On se contente de montrer que f est majorée et atteint sa borne supérieure. Soit $M = \sup\{f(u);\ u \in K\} \leq \infty$. (Par convention, on pose $\sup A = \infty$ pour un ensemble A non majoré). Par définition de M, on peut trouver une suite $(u_k) \subset K$ telle que $f(u_k) \to M$. Comme K est compact, on peut supposer, quitte à extraire une sous-suite, que la suite (u_k) converge vers un certain $u \in K$. Alors $f(u_k) \to f(u)$ par continuité, et donc $f(u) = M$. Cela prouve à la fois que $M < \infty$ (i.e. que f est majorée) et que f atteint sa borne supérieure. □

Exercice 1. Soit K un compact de $\mathbb{C}$ contenu dans $\Omega = \{z \in \mathbb{C};\ \mathrm{Re}(z) > 0\}$. Montrer qu'on peut trouver $\alpha > 0$ tel que $\forall z \in K\ :\ \mathrm{Re}(z) \geq \alpha$.

Exercice 2. Soit $\Omega = \{z \in \mathbb{C};\ \mathrm{Re}(z) > 0\}$. Montrer que la série $\sum e^{-nz}$ converge normalement sur tout compact de Ω.

THÉORÈME 4.4. *Soit K un compact d'un evn E, et soit F un autre evn. Si* $: K \to F$ *est continue, alors f est* ***uniformément continue****. Autrement dit, étant donné $\varepsilon > 0$, on peut prendre le même "δ de continuité" pour tous les points $a \in K$.*

DÉMONSTRATION. Avec des quantificateurs, il s'agit de montrer la chose suivante (comparer avec la définition de la continuité) :

$$\forall \varepsilon > 0\ \exists \delta = \delta(\varepsilon) > 0\ \forall a, u \in K\ :\ \|u - a\| < \delta \implies \|f(u) - f(a)\| < \varepsilon\,.$$

Supposons que cela ne soit pas vrai. Alors on peut trouver $\varepsilon_0 > 0$ tel que

$$\forall \delta > 0\ \exists a, u \in K\ :\ \|u - a\| < \delta \ \text{ et } \ \|f(u) - f(a)\| \geq \varepsilon_0\,.$$

En prenant $\delta = 1/k$ (où $k \in \mathbb{N}^*$), on obtient deux suites $(a_k), (u_k) \subset K$ telles que $\|u_k - a_k\| \to 0$ et $\|f(u_k) - f(a_k)\| \geq \varepsilon_0$ pour tout $k \in \mathbb{N}^*$. Comme K est compact, on peut supposer, quitte à extraire une sous-suite, que (a_k) converge vers un certain $a \in K$. Alors $u_k = (u_k - a_k) + a_k$ tend également vers a puisque $u_k - a_k \to 0$. Par continuité, $f(a_k)$ et $f(u_k)$ tendent tous les deux vers $f(a)$, et donc $f(u_k) - f(a_k) \to 0$. Comme $\|f(u_k) - f(a_k)\| \geq \varepsilon_0 > 0$ pour tout n, on obtient donc une contradiction. □

Voici une conséquence ce théorème qui nous sera utile plus tard.

COROLLAIRE 4.5. *Soit $\Psi : I \times [a, b] \to F$ une application continue, où I est une partie d'un evn E et $[a, b]$ est un intervalle compact de $\mathbb{R}$. Pour tout $t \in I$, notons $\Psi(t, \cdot)$ la fonction continue $x \mapsto \Psi(t, x)$. Alors l'application $t \mapsto \Psi(t, \cdot)$ est continue de I dans $\mathcal{C}([a, b])$.*

DÉMONSTRATION. Soit (t_n) une suite de points de I convergeant vers $t_\infty \in I$. Alors l'ensemble $K = \{t_n;\ n \in \mathbb{N}\} \cup \{t_\infty\}$ est compact, donc $K \times [a, b]$ également, *et donc* la fonction Ψ est uniformément continue sur $K \times [a, b]$. Il est alors facile de vérifier que $\Psi(t_n, x)$ tend vers $\Psi(t_\infty, x)$ uniformément sur $[a, b]$, ce qui signifie que $\Psi(t_n, \cdot)$ tend vers $\Psi(t_\infty, \cdot)$ dans l'espace $\mathcal{C}([a, b])$. □

4.2. Ouverts connexes d'un evn.

DÉFINITION 4.6. *Soit Ω un ouvert d'un evn E. On dit que Ω est* **connexe** *s'il est impossible de partitionner Ω en deux ouverts non vides. Autrement dit, s'il est impossible d'écrire $\Omega = \Omega_0 \cup \Omega_1$, où Ω_0 et Ω_1 sont des ouverts non vides et $\Omega_1 \cap \Omega_2 = \emptyset$.*

Il faut connaitre cette définition, mais on renvoie au cours de topologie pour plus de détails. Le seul résultat dont on aura besoin (et en fait, on pourrait s'en passer) est la proposition ci-dessous. Appelons **ligne brisée** dans un evn E tout ensemble L constitué d'un nombre fini de segments mis bout à bout (et pouvant éventuellement se recouper) ; autrement dit, tout ensemble du type $L = [a_0, a_1] \cup [a_1, a_2] \cup \cdots \cup [a_{N-1}, a_N]$, où les a_i sont des points de E (ce qu'on vient d'écrire suppose $N \geq 2$, mais bien entendu un segment $[a_0, a_1]$ est aussi une ligne brisée). De manière équivalente, une ligne brisée est l'image d'une application $\gamma : [0,1] \to E$ continue et affine par morceaux. (*Exercice* : montrer que c'est bien la même chose).

PROPOSITION 4.7. *Soit E un evn. Pour un ouvert $\Omega \subset E$, les propriétés suivantes sont équivalentes :*

(i) *Ω est connexe ;*
(ii) *deux points quelconques de Ω peuvent toujours être reliés par une ligne brisée entièrement contenue dans Ω.*

DÉMONSTRATION. Pour montrer que (i) entraine (ii), supposons Ω connexe. Soit $u \in \Omega$ fixé. Il suffit de montrer que tout point $v \in \Omega$ peut être relié à u par une ligne brisée contenue dans Ω, puisque u est un point arbitrairement choisi de Ω.

Notons Ω_0 l'ensemble des points $v \in \Omega$ qui peuvent être reliés à u, et Ω_1 l'ensemble des points qui ne peuvent pas être reliés à u. Il s'agit de montrer que Ω_1 est vide. Comme Ω est supposé connexe et comme Ω_0 n'est pas vide (il contient u grâce à la ligne brisé "triviale" $L = [u, u]$), il suffit pour cela de montrer que Ω_0 et Ω_1 sont tous les deux *ouverts*.

Soit $v \in \Omega_0$ quelconque, et choisissons une ligne brisée $L \subset \Omega$ joignant u à v. Comme Ω est ouvert, on peut trouver $r > 0$ tel que $B(v, r) \subset \Omega$. Si $w \in B(v, r)$, alors le segment $[v, w]$ est contenu dans $B(v, r)$ (et donc dans Ω) car la boule $B(v, r)$ est *convexe*. Par conséquent, la ligne brisée $L_w = L \cup [v, w]$ est contenue dans Ω, et relie évidemment u à w. Ainsi, tout point $w \in B(v, r)$ appartient à Ω_0, autrement dit $B(v, r) \subset \Omega_0$. On a donc montré que Ω_0 est ouvert.

Soit maintenant $v \in \Omega_1$ quelconque, et choisissons $r > 0$ tel que $B(v, r) \subset \Omega$. Si on pouvait relier u à un point $w \in B(v, r)$ par une ligne brisée $L \subset \Omega$, alors on pourrait relier u à v par la ligne brisée $L \cup [w, v]$ (qui est contenue dans Ω par convexité de $B(v, r)$) ; mais ceci est exclu puisque $v \in \Omega_1$. Par conséquent, *aucun* point $w \in B(v, r)$ ne peut être relié à u, autrement dit $B(v, r) \subset \Omega_1$. Cela montre que Ω_1 est également ouvert.

Montrons maintenant que (ii) entraine (i). On raisonne par l'absurde en supposant que (ii) est vérifiée mais que (i) ne l'est pas. On peut donc écrire $\Omega = \Omega_0 \cup \Omega_1$, où Ω_0 et Ω_1 sont des ouverts non vides et $\Omega_0 \cap \Omega_1 = \emptyset$. Choisissons un point $u_0 \in \Omega_0$ et un point $u_1 \in \Omega_1$. D'après (ii), on peut trouver une application continue (et affine par morceaux) $\gamma : [0,1] \to E$ telle que $\gamma(0) = u_0$, $\gamma(1) = u_1$ et $\gamma(t) \in \Omega$ pour tout $t \in [0,1]$.

Posons $a = \sup \{t \in [0,1];\ \gamma(t) \in \Omega_0\}$. Cela a bien un sens car l'ensemble $A = \{t \in [0,1];\ \gamma(t) \in \Omega_0\}$ est non vide (il contient $t = 0$ car $\gamma(0) = u_0 \in \Omega_0$) et majoré par 1. On va montrer que $\gamma(a)$ n'appartient ni à Ω_0, ni à Ω_1, ce qui fournira une contradiction puisque par hypothèse $\gamma(a)$ doit appartenir à Ω et $\Omega = \Omega_0 \cup \Omega_1$.

Par définition de a, on peut trouver une suite $(a_k) \subset A$ telle que $a_k \to a$. Alors $\gamma(a_k) \to \gamma(a)$ par continuité de γ. Comme $\gamma(a_k) \in \Omega_0 \subset E \setminus \Omega_1$ pour tout k et comme $E \setminus \Omega_1$ est *fermé* dans E, on en déduit que $\gamma(a) \in E \setminus \Omega_1$, autrement dit $\gamma(a) \notin \Omega_1$.

Supposons que $\gamma(a) \in \Omega_0$. Alors $a < 1$ car $\gamma(1) = u_1 \in \Omega_1$ et $\Omega_1 \cap \Omega_0 = \emptyset$. Comme Ω_0 est ouvert et que γ est continue, on peut trouver $\delta > 0$ tel que $\gamma(t) \in \Omega_0$ pour tout $t \in [0,1]$ vérifiant $|t-a| \leq \delta$. Comme $a < 1$, on en déduit qu'on peut trouver un $t > a$ tel que $\gamma(t) \in \Omega_0$; mais cela contredit la définition de a. On a donc montré que $\gamma(a) \notin \Omega_0$, ce qui termine la démonstration.

□

COROLLAIRE 4.8. *Si E est un evn, alors tout ouvert* convexe *de E est connexe. En particulier, E lui même est connexe (ce qui n'était pas évident a priori).*

Exercice. Soit Ω un ouvert d'un evn E, et soit $\mathcal{R}$ une *relation d'équivalence* sur Ω. On suppose que la classe d'équivalence $[u]_{\mathcal{R}}$ de tout point $u \in \Omega$ est un ouvert de E. Montrer que $\Omega \setminus [u]_{\mathcal{R}}$ est également un ouvert, pour tout $u \in \Omega$. Pourquoi cet exercice est-il placé ici ?

Remarque 1. La proposition précédente a un grand intérêt *pratique* : il est souvent beaucoup plus facile de montrer qu'un ouvert "concret" vérifie la propriété (ii) que de prouver qu'il ne peut pas être partitionné en deux ouverts non vides disjoints.

Remarque 2. Une partie A de E est dite **connexe par arcs** si, pour tous $u, v \in A$, on peut trouver un "chemin continu reliant u à v dans A", i.e. une application continue $\gamma : [0,1] \to E$ telle que $\gamma(0) = u$, $\gamma(1) = v$ et $\gamma(t) \in A$ pour tout $t \in [0,1]$. La démonstration précédente a établi qu'un ouvert de E est connexe si et seulement si il est connexe par arcs. (*Exercice :* vérifier).

4.3. Espaces complets.

DÉFINITION 4.9. *Soit E un evn. On dit qu'une suite $(u_k) \subset E$ est une* ***suite de Cauchy*** *si $\|u_q - u_p\| \to 0$ quand $p, q \to \infty$; autrement dit :*

$$\forall \varepsilon > 0\, \exists K\, \forall p, q \geq K \ : \ \|u_q - u_p\| \leq \varepsilon\,.$$

On dit que l'espace E est ***complet****, ou encore que E est un* ***espace de Banach****, si toute suite de Cauchy $(u_k) \subset E$ est convergente.*

Exercice. Montrer qu'une suite convergente est toujours de Cauchy, et que toute suite de Cauchy est bornée.

REMARQUE. Un evn complet reste complet si on remplace la norme par une norme équivalente.

Exemple 1. $\mathbb{R}$ et $\mathbb{C}$ sont complets : c'est le **critère de Cauchy** pour les suites numériques.

Exemple 2. Tout evn *de dimension finie* est complet.

DÉMONSTRATION. Par équivalence des normes, il suffit de montrer que $\mathbb{R}^n$ est complet pour la norme $\|\cdot\|_\infty$. Soit donc $(u_k) \subset \mathbb{R}^n$ une suite de Cauchy pour $\|\cdot\|_\infty$, et écrivons $u_k = (u_k(1), \dots, u_k(n))$. Si $j \in \{1, \dots, n\}$, alors

$$|u_q(j) - u_p(j)| \leq \|u_q - u_p\|_\infty$$

pour tous $p, q \in \mathbb{N}$. On en déduit que la suite $(u_k(j))_{k \in \mathbb{N}}$ est de Cauchy dans $\mathbb{R}$, et converge donc vers un certain nombre réel l_j. Si on pose $u = (l_1, \dots, l_n) \in \mathbb{R}^n$, alors

(u_k) converge vers u coordonnée par coordonnée, et donc au sens de la norme $\|\cdot\|_\infty$. Ainsi, on a bien montré que la suite de Cauchy (u_k) est convergente. $\square$

Exemple 3. L'espace $\mathcal{C}([0,1])$ est complet pour la norme $\|\cdot\|_\infty$: c'est la traduction du *critère de Cauchy uniforme* pour les suites de fonctions continues.

Le résultat suivant est souvent très utile. Si (u_k) est une suite dans un evn E, on dira que **la série $\sum u_k$ est convergente** si la suite des sommes partielles $S_k = \sum_{i=0}^k u_i$ converge dans E ; et on dira que la série $\sum u_k$ est **normalement convergente** si la série à termes positifs $\sum \|u_k\|$ est convergente.

PROPOSITION 4.10. *Dans un evn* complet*, toute série normalement convergente est convergente.*

DÉMONSTRATION. Soit $\sum u_k$ une série normalement convergente dans un evn complet E, et posons $S_k = \sum_{i=0}^k u_i$. Si $p < q$, alors

$$\begin{aligned} \|S_q - S_p\| &= \left\| \sum_{i=p+1}^{q} u_i \right\| \\ &\leq \sum_{i=p+1}^{q} \|u_i\| \\ &\leq \sum_{i=p+1}^{\infty} \|u_i\| \,. \end{aligned}$$

Comme la série $\sum \|u_k\|$ est convergente, le "reste" $\sum_{p+1}^\infty \|u_i\|$ tend vers 0 quand $p \to \infty$. Par conséquent, $\|S_q - S_p\|$ tend vers 0 quand $p, q \to \infty$; autrement dit la suite (S_k) est de Cauchy dans E. Comme E est supposé complet, cela montre que la suite (S_k) est convergente, i.e. que la série $\sum u_k$ converge dans E. $\square$

Remarque. En fait, on peut montrer que la proposition précédente *caractérise* les espaces complets : si E est un evn dans lequel toute série normalement convergente est convergente, alors E est complet. C'est un exercice intéressant.

5. Applications linéaires et multilinéaires

5.1. Applications linéaires continues. Bien que facile à démontrer, le théorème suivant est *très* important. Dans la pratique, c'est *toujours* en utilisant ce théorème qu'on montre qu'une application linéaire donnée est continue, et pas en revenant à la définition de la continuité.

THÉORÈME 5.1. (critère de continuité)
Soient E et F deux evn, et soit $L : E \to F$ une application linéaire*. Les propriétés suivantes sont équivalentes :*

(i) *L est continue ;*
(ii) *il existe une constante $C < \infty$ telle que $\forall u \in E \,:\, \|L(u)\| \leq C\,\|u\|$.*

DÉMONSTRATION. Supposons L continue. Alors L est continue en 0, donc on peut trouver $\delta > 0$ tel que $\|L(v) - L(0)\| \leq 1$ pour tout $v \in E$ vérifiant $\|v - 0\| \leq \delta$; autrement dit :

$$\|v\| \leq \delta \implies \|L(v)\| \leq 1\,.$$

Soit maintenant $u \in E$ quelconque, avec $u \neq 0$. Alors $v = \delta \frac{u}{\|u\|}$ vérifie $\|v\| = \delta$, et on a donc

$$\left\| L\left(\delta \frac{u}{\|u\|}\right) \right\| \leq 1\,.$$

Mais $\left\| L\left(\delta \frac{u}{\|u\|}\right) \right\| = \left\| \frac{\delta}{\|u\|} L(u) \right\| = \frac{\delta}{\|u\|} \|L(u)\|$, donc l'inégalité précédente s'écrit

$$\|L(u)\| \leq \frac{1}{\delta} \|u\|\,.$$

Ceci reste évidemment vrai pour $u = 0$, et on a donc montré que (ii) est vérifiée avec $C = 1/\delta$.

Supposons maintenant que (ii) soit vérifiée pour une certaine constante C. Pour tous $u, v \in E$, on a alors (par linéarité de L)

$$\|L(v) - L(u)\| = \|L(v - u)\| \leq C\,\|v - u\|\,,$$

ce qui montre que L est lipschitzienne et donc continue. □

Remarque. La fin de la démonstration a établi que si C est comme dans (ii), alors L est C-lipschitzienne.

Corollaire 5.2. *Si l'espace* de départ *E est* de dimension finie*, alors toute application linéaire $L : E \to F$ est continue.*

Démonstration. Par équivalence des normes, on peut supposer que $E = (\mathbb{R}^n, \|\cdot\|_\infty)$. Notons $(e_1, \dots, e_n)$ la base canonique de $\mathbb{R}^n$. Si $u = (u(1), \dots, u(n)) = \sum_1^n u(j)e_j \in E$, alors

$$\begin{aligned}
\|L(u)\| &= \left\| L\left(\sum_{j=1}^n u(j)e_j\right) \right\| \\
&= \left\| \sum_{j=1}^n u(j)\, L(e_j) \right\| \\
&\leq \sum_{j=1}^n |u(j)|\, \|L(e_j)\| \\
&\leq \|u\|_\infty \times \sum_{j=1}^n \|L(e_j)\|\,.
\end{aligned}$$

La propriété (ii) du critère de continuité est donc vérifiée avec $C = \sum_1^n \|L(e_j)\|$ (qui est bien indépendante de $u \in \mathbb{R}^n$). □

Exercice 1. Une démonstration identique a déjà été faite. A quel endroit ?

Exercice 2. Montrer que si E est un evn, alors l'application $(u, v) \mapsto u + v$ est continue de $E \times E$ dans E.

Notation. Si E et F sont deux evn, on notera toujours $\mathcal{L}(E, F)$ l'espace vectoriel constitué par les applications linéaires *continues* de E dans F. Si $E = F$, on écrit $\mathcal{L}(E)$ au lieu de $\mathcal{L}(E, E)$. Pour $F = \mathbb{R}$, on écrira souvent E^* au lieu de $\mathcal{L}(E, \mathbb{R})$. Un élément de E^* s'appelle une **forme linéaire** continue (sur E), et l'espace E^* s'appelle le **dual** de E.

DÉFINITION 5.3. *Soient E et F deux evn. Si $L : E \to F$ est une application linéaire continue, on pose*

$$\|L\|_{\mathcal{L}(E,F)} = \sup\left\{ \frac{\|L(u)\|_F}{\|u\|_E};\ u \in E, u \neq 0 \right\}.$$

REFORMULATION. $\|L\|_{\mathcal{L}(E,F)}$ est la plus petite constante C telle que $\|L(u)\| \leq C\,\|u\|$ pour tout $u \in E$. Donc, pour un nombre réel positif C, on a l'équivalence

$$\|L\|_{\mathcal{L}(E,F)} \leq C \iff \forall u \in E\ :\ \|L(u)\| \leq C\,\|u\|\,.$$

Une conséquence est qu'il est en général plus ou moins "automatique" de majorer $\|L\|_{\mathcal{L}(E,F)}$. En revanche, une minoration demande a priori plus d'initiative.

Remarque 1. La notation n'est pas fantaisiste : $\|\cdot\|_{\mathcal{L}(E,F)}$ est effectivement une *norme* sur $\mathcal{L}(E,F)$, qui est dite **subordonnée** aux normes de E et de F. On supposera toujours que $\mathcal{L}(E,F)$ est muni de cette norme, et en conséquence on écrira le plus souvent $\|\cdot\|$ au lieu de $\|\cdot\|_{\mathcal{L}(E,F)}$.

Remarque 2. La norme $\|L\|_{\mathcal{L}(E,F)}$ dépend des normes choisies sur E et F. Mais si on remplace les normes de E et F par des normes équivalentes, on obtient une norme équivalente sur $\mathcal{L}(E,F)$.

Exercice. Démontrer ce qui est dit dans les remarques 1 et 2.

Exemple 1. Si E est un evn, alors $\|Id_E\| = 1$.

Exemple 2. Soit $E = (\mathbb{R}^n, \|\cdot\|_\infty)$. Pour $b = (b_1, \dots, b_n) \in \mathbb{R}^n$, notons $\Theta_b : E \to \mathbb{R}$ la forme linéaire (continue) définie par

$$\Theta_b(x_1, \dots, x_n) = \sum_{j=1}^n b_j x_j\,.$$

Alors $\|\Theta_b\|_{E^*} = \|b\|_1 = \sum_{j=1}^n |b_j|$. De manière un peu pédante, on dit que "$(\mathbb{R}^n, \|\cdot\|_\infty)^*$ s'identifie isométriquement à $(\mathbb{R}^n, \|\cdot\|_1)$".

DÉMONSTRATION. Pour tout $u = (x_1, \dots, x_n) \in \mathbb{R}^n$, on a

$$\begin{aligned} |\Theta_b(u)| &\leq \sum_{j=1}^n |b_j|\,|x_j| \\ &\leq \max_j |x_j| \times \sum_{j=1}^n |b_j| \\ &= \|b\|_1\,\|u\|_\infty\,. \end{aligned}$$

Par définition de la norme d'une forme linéaire, on a donc $\|\Theta_b\| \leq \|b\|_1$.

Pour montrer qu'on a aussi $\|\Theta_b\| \geq \|b\|_1$, il suffit de trouver $u \in \mathbb{R}^n$ tel que $\|u\|_\infty = 1$ et $\Theta_b(u) = \|b\|_1$. Pour tout $j \in \{1, \dots, n\}$, choisissons $\varepsilon_j = \pm 1$ tel que $\varepsilon_j b_j = |b_j|$. Alors $u = (\varepsilon_1, \dots, \varepsilon_n)$ convient : on a $\|u\|_\infty = 1$, et $\Theta_b(u) = \sum_1^n \varepsilon_j b_j = \sum_1^n |b_j| = \|b\|_1$. $\square$

Exercice. On garde les notations de l'exemple 2. Exprimer $\Theta_b(u)$ à l'aide du produit scalaire usuel de $\mathbb{R}^n$, puis montrer que si on munit $\mathbb{R}^n$ de la norme euclidienne, alors $\|\Theta_b\| = \|b\|_2$.

NOTATION. On notera toujours AB la *composée* de deux applications linéaires A et B, i.e. $AB = A \circ B$: si $B \in \mathcal{L}(E, F)$ et $A \in \mathcal{L}(F, G)$ (où E, F, G sont trois evn), alors $AB \in \mathcal{L}(E, G)$ et $AB(u) = A(B(u))$. En accord avec cette notation, si $A \in \mathcal{L}(E)$ et $k \in \mathbb{N}^*$ on note A^k la composée $A \circ \cdots \circ A$ (k fois). On pose aussi $A^0 = I$.

PROPOSITION 5.4. *Si $B \in \mathcal{L}(E, F)$ et $A \in \mathcal{L}(F, G)$, alors $\|AB\| \leq \|A\| \, \|B\|$.*

DÉMONSTRATION. C'est “mécanique” : on a

$$\|AB(u)\| = \|A(B(u))\| \leq \|A\| \, \|B(u)\| \leq \|A\| \, \|B\| \, \|u\|$$

pour tout $u \in E$, d'où le résultat. □

COROLLAIRE 5.5. *Si $A \in \mathcal{L}(E)$, alors $\|A^k\| \leq \|A\|^k$ pour tout $k \in \mathbb{N}$.*

DÉMONSTRATION. Une récurrence “immédiate”. □

Remarque importante. On sait bien que $M_n(\mathbb{R})$ s'identifie canoniquement à $\mathcal{L}(\mathbb{R}^n)$. A chaque norme sur $\mathbb{R}^n$ correspond donc une norme subordonnée sur $M_n(\mathbb{R})$. On dira qu'une telle norme sur $M_n(\mathbb{R})$ est une **norme matricielle**. D'après la proposition précédente, une norme matricielle est toujours **sous-multiplicative** ($\|AB\| \leq \|A\| \, \|B\|$) et on a $\|I\| = 1$.

Exercice. Montrer que pour toute matrice $A \in M_n(\mathbb{R})$, la série $\sum \frac{A^k}{k!}$ converge dans $M_n(\mathbb{R})$.

PROPOSITION 5.6. *Soient E et F deux evn. On suppose que F est* complet. *Alors $\mathcal{L}(E, F)$ est complet pour la norme $\| \cdot \|_{\mathcal{L}(E,F)}$.*

DÉMONSTRATION. Soit (L_k) une suite de Cauchy dans $\mathcal{L}(E, F)$. Il s'agit de montrer que (L_k) converge dans $\mathcal{L}(E, F)$, autrement dit de trouver une $L \in \mathcal{L}(E, F)$ telle que $\|L_k - L\|_{\mathcal{L}(E,F)} \to 0$. Cela va se faire en 3 étapes, ce qui est toujours le cas lors d'une démonstration de complétude : on commence par identifier un “candidat limite”, puis on montre que ce candidat appartient bien à l'espace considéré (ici $\mathcal{L}(E, F)$), et enfin que la suite converge effectivement vers le candidat limite au sens de la norme de l'espace en question.

Pour tout $u \in E$ et pour $p, q \in \mathbb{N}$, on a

$$\|L_q(u) - L_p(u)\| \leq \|L_q - L_p\|_{\mathcal{L}(E,F)} \times \|u\| \,.$$

Comme la suite (L_k) est de Cauchy, on en déduit que la suite $(L_k(u))$ est de Cauchy dans F, et est donc convergente puisque F est supposé complet. Pour tout $u \in E$, on peut donc poser

$$L(u) = \lim_{k \to \infty} L_k(u) \,.$$

L'application $L : E \to F$ ainsi définie est notre “candidat limite”.

Il est clair que L est linéaire car les L_k le sont. De plus, la suite (L_k) est *bornée* dans $\mathcal{L}(E, F)$ car elle est de Cauchy : a donc une constante M telle que $\|L_k\| \leq M$ pour tout $k \in \mathbb{N}$, ce qui s'écrit encore (par définition de la norme de $\mathcal{L}(E, F)$)

$$\forall u \in E \, \forall k \ : \ \|L_k(u)\| \leq M \, \|u\| \,.$$

En faisant tendre k vers l'infini, on en déduit qu'on a $\|L(u)\| \leq M \, \|u\|$ pour tout $u \in E$, ce qui prouve que l'application linéaire L est continue. Ainsi, $L \in \mathcal{L}(E, F)$.

Pour montrer que $\|L_k - L\|_{\mathcal{L}(E,F)} \to 0$, fixons $\varepsilon > 0$. Comme (L_k) est de Cauchy dans $\mathcal{L}(E,F)$, on peut trouver K tel que $\|L_q - L_p\|_{\mathcal{L}(E,F)} \leq \varepsilon$ pour $p, q \geq K$. Par définition de $\|\cdot\|_{\mathcal{L}(E,F)}$, cela s'écrit

$$\forall u \in E\, \forall p, q \geq K \ :\ \|(L_q - L_p)(u)\| \leq \varepsilon \, \|u\| \, .$$

En fixant p et en faisant tendre q vers l'infini, on en déduit (puisque $L_q(u) \to L(u)$)

$$\forall u \in E\, \forall p \geq K \ :\ \|(L - L_p)(u)\| \leq \varepsilon \, \|u\| \, .$$

Par définition de $\|\cdot\|_{\mathcal{L}(E,F)}$ (à nouveau), cela signifie qu'on a $\|L - L_p\|_{\mathcal{L}(E,F)} \leq \varepsilon$ pour tout $p \geq K$, ce qui termine la démonstration. □

Corollaire 5.7. *Si E est un evn, alors son dual E^* est un espace de Banach.*

Remarque. Si E et F sont tous les deux de dimension finie, alors la preuve précédente est inutile puisque $\mathcal{L}(E,F)$ est également de dimension finie, et donc complet pour n'importe quelle norme.

5.2. Applications bilinéaires continues.

Définition 5.8. *Soient E_1, E_2 et F trois evn. Une application $B : E_1 \times E_2 \to F$ est dite* ***bilinéaire*** *si $B(u_1, u_2)$ est linéaire par rapport à chaque variable u_1, u_2 séparément ; autrement dit, si pour tout $u_1 \in E_1$ fixé, l'application $u_2 \mapsto B(u_1, u_2)$ est linéaire, et pour tout $u_2 \in E_2$ fixé, l'application $u_1 \mapsto B(u_1, u_2)$ est linéaire.*

Remarque. Intuitivement, une application bilinéaire est un "produit". Il est toujours bon de garder cette idée en tête.

Exemple 1. Le produit usuel $(x, y) \mapsto xy$ est bilinéaire de $\mathbb{R} \times \mathbb{R}$ dans $\mathbb{R}$. Plus généralement, si E est un espace vectoriel, alors l'application $(\lambda, u) \mapsto \lambda u$ est bilinéaire de $\mathbb{R} \times E$ dans E.

Exemple 2. Si on note $\langle\, ,\, \rangle$ le produit scalaire usuel sur $\mathbb{R}^n$, alors l'application $(x, y) \mapsto \langle x, y \rangle$ est bilinéaire de $\mathbb{R}^n \times \mathbb{R}^n$ dans $\mathbb{R}$.

Exemple 3. Si E et F sont des evn, alors l'application $(L, u) \mapsto L(u)$ est bilinéaire de $\mathcal{L}(E,F) \times E$ dans F.

Exemple 4. Si E, F, G sont des evn, alors l'application $(S, T) \mapsto TS$ est bilinéaire de $\mathcal{L}(E,F) \times \mathcal{L}(F,G)$ dans $\mathcal{L}(E,G)$.

La proposition suivante caractérise la continuité pour les applications bilinéaires de manière analogue à ce que fait le théorème 5.1 pour les applications linéaires. Rappelons que si E_1 et E_2 sont deux evn, alors $E_1 \times E_2$ est muni de la norme produit, et qu'une suite converge dans $E_1 \times E_2$ si et seulement si elle converge coordonnée par coordonnée (cf l'exemple 4 de la section 1.1 et l'exemple 4 de la section 2.1)

Proposition 5.9. (critère de continuité)
Soient E_1, E_2 et F trois evn. Pour une application bilinéaire $B : E_1 \times E_2 \to F$, les propriétés suivantes sont équivalentes :
(i) *B est continue ;*
(ii) *il existe une constante $C < \infty$ telle que*

$$\forall (x, y) \in E_1 \times E_2 \ :\ \|B(x, y)\| \leq C \, \|x\| \, \|y\| \, .$$

Démonstration. L'implication "(i) entraine (ii)" se démontre exactement comme pour les applications linéaires (voir la preuve du théorème 5.1, mais attention : la constante qui sort est cette fois $C = 1/\delta^2$. Pourquoi ?). Inversement, supposons (ii) vérifiée. Soit $(u_k)_{k\in\mathbb{N}} = (x_k, y_k)$ une suite dans $E_1 \times E_2$ convergeant vers $u = (x, y) \in E_1 \times E_2$. Par bilinéarité et en utilisant (ii), on a

$$\begin{aligned}\|B(x_k, y_k) - B(x, y)\| &= \|B(x_k - x, y_k) + B(x, y_k - y)\| \\ &\leq \|B(x_k - x, y_k)\| + \|B(x, y_k - y)\| \\ &\leq C\left(\|x_k - x\|\,\|y_k\| + \|x\|\,\|y_k - y\|\right).\end{aligned}$$

D'autre part, la suite (y_k) est *bornée* puisqu'elle converge vers y Il existe donc une constante M telle que $\|y_k\| \leq M$ pour tout $k \in \mathbb{N}$, et on obtient ainsi

$$\|B(x_k, y_k) - B(x, y)\| \leq CM\,\|x_k - x\| + C\,\|y_k - y\|$$

Comme $x_k \to x$ et $y_k \to y$, cela montre que $B(x_k, y_k)$ tend vers $B(x, y)$. Par conséquent, B est continue. □

Exercice. Comparer la preuve précédente avec celle (vue en première année) du fait que si (u_k) et (v_k) sont deux suites numériques convergentes de limites u et v, alors $u_k v_k \to uv$.

Corollaire 5.10. *Si E, F et G sont trois evn, alors l'application $(S, T) \mapsto TS$ est continue de $\mathcal{L}(E, F) \times \mathcal{L}(F, G)$ dans $\mathcal{L}(E, G)$.*

Démonstration. Cette application est bilinéaire, et on a $\|TS\| \leq \|T\|\,\|S\|$. □

Corollaire 5.11. *Si E_1 et E_2 sont des evn* de dimension finie, *alors toute application bilinéaire $B : E_1 \times E_2 \to F$ est continue.*

Démonstration. Pour $u_1 \in E_1$, notons $B_{u_1} : E_2 \to F$ l'application définie par $B_{u_1}(u_2) = B(u_1, u_2)$. Cette application est linéaire par bilinéarité de B, et elle est donc continue puisque $\dim(E_2) < \infty$. Ainsi, $B_{u_1} \in \mathcal{L}(E_2, F)$, et on a donc sous la main une application $u_1 \mapsto B_{u_1}$ de E_1 dans $\mathcal{L}(E_2, F)$. Cette application est linéaire par bilinéarité de B, donc continue puisque $\dim(E_1) < \infty$. Il existe donc une constante C telle que

$$\|B_{u_1}\|_{\mathcal{L}(E_2,F)} \leq C\,\|u_1\|$$

pour tout $u_1 \in E_1$. On en déduit

$$\begin{aligned}\|B(u_1, u_2)\| &= \|B_{u_1}(u_2)\| \\ &\leq \|B_{u_1}\|\,\|u_2\| \\ &\leq C\,\|u_1\|\,\|u_2\|\end{aligned}$$

pour tout $(u_1, u_2) \in E_1 \times E_2$, ce qui prouve que B est continue. □

Autre démonstration. On procède comme dans la preuve de la continuité des aplications *linéaires* en dimension finie (corollaire 5.2). Par équivalence des normes, on peut supposer que $E_1 = \mathbb{R}^m$ et $E_2 = \mathbb{R}^n$, tous deux munis de la norme $\|\cdot\|_\infty$. Soient $(e_1, \dots, e_m)$ la base canonique de $\mathbb{R}^n$ et $(f_1, \dots, f_n)$ la base canonique de $\mathbb{R}^m$. Si $x = \sum_1^m x_i e_i \in \mathbb{R}^m$ et $y = \sum_1^n y_j f_j \in \mathbb{R}^n$ alors, par bilinéarité,

$$B(x, y) = \sum_{i,j} x_i y_j\, B(e_i, f_j)\,.$$

On en déduit qu'on a $\|B(x, y)\| \leq C\,\|x\|_\infty\,\|y\|_\infty$ pour tout $(x, y) \in \mathbb{R}^m \times \mathbb{R}^n$, où $C = \sum_{i,j} \|B(e_i; e_j)\|$. □

Exercice 1. Soient E et F des evn. Montrer que l'application $(T, u) \mapsto T(u)$ est continue de $\mathcal{L}(E, F) \times E$ dans F.

Exercice 2. Soit E un evn. Montrer que l'application $(\lambda, u) \mapsto \lambda u$ est continue de $\mathbb{R} \times E$ dans E.

5.3. Applications multilinéaires continues. Soient $E_1, \ldots, E_n$ et F des evn. Une application $P : E_1 \times \cdots \times E_n \to F$ est dite **n-linéaire** si $P(u_1, \ldots, u_n)$ est linéaire par rapport à chaque variable séparément. Par exemple, l'application $(u_1, \ldots, u_n) \mapsto \det(u_1, \ldots, u_n)$ est n-linéaire de $\mathbb{R}^n \times \cdots \times \mathbb{R}^n$ dans $\mathbb{R}$.

Exactement comme pour les applications bilinéaires, on montre qu'une application n-linéaire $P : E_1 \times \cdots \times E_n \to F$ est continue si et seulement si il existe une constante C telle que

$$\forall (u_1, \ldots, u_n) \in E_1 \times \cdots \times E_n \ : \ \|P(u_1, \ldots, u_n)\| \leq C \, \|u_1\| \cdots \|u_n\| \,. \tag{5.1}$$

La démonstration est identique bien qu'un peu plus pénible à écrire et constitue un exercice instructif (essayer avec $n = 3$).

On peut également démontrer *par récurrence sur* n que si une application n-linéaire $P : E_1 \times \cdots \times E_n \to F$ vérifie (5.1), alors elle est continue. Supposons le résultat acquis pour $n - 1$ (avec $n \geq 2$), et soit $P : E_1 \times \cdots \times E_n \to F$ une application n-linéaire vérifiant (5.1). Pour $(u_1, \ldots, u_{n-1}) \in E_1 \times \cdots \times E_{n-1}$, notons $P_{(u_1, \ldots, u_{n-1})} : E_n \to F$ l'application linéaire définie par $P_{(u_1, \ldots, u_{n-1})}(v) = P(u_1, \ldots, u_{n-1}, v)$. Par (5.1), on a

$$\|P_{(u_1, \ldots, u_{n-1})}(v)\| \leq C \, \|u_1\| \cdots \|u_{n-1}\| \, \|v\| \,.$$

Par conséquent l'application linéaire $P_{(u_1, \ldots, u_{n-1})}$ est continue, avec $\|P_{(u_1, \ldots, u_{n-1})}\| \leq C \, \|u_1\| \cdots \|u_{n-1}\|$. D'après l'hypothèse de récurrence, on en déduit que l'application $(n-1)$*-linéaire* $(u_1, \ldots, u_{n-1}) \mapsto P_{(u_1, \ldots, u_{n-1})}$ est continue de $E_1 \times \cdots \times E_{n-1}$ dans $\mathcal{L}(E_n, F)$. Maintenant, si $u = (u_1, \ldots, u_n) \in E_1 \times \cdots \times E_n$, alors

$$P(u_1, \ldots, u_n) = P_{(u_1, \ldots, u_{n-1})}(u_n) = B(P_{(u_1, \ldots, u_{n-1})}, u_n) \,,$$

où $B : \mathcal{L}(E_n, F) \times E_n \to F$ est l'application bilinéaire définie par $B(T, v) = T(v)$. On a vu plus haut que cette application B est continue, donc P est continue par composition.

On déduit comme plus haut du critère de continuité que si $E_1, \ldots, E_n$ sont de dimension finie, alors toute application n-linéaire de $E_1 \times \cdots \times E_n$ dans un evn F est continue.

Exercice. Soit E un evn. Montrer que pour tout $n \in \mathbb{N}$, l'application $L \mapsto L^n$ est continue sur $\mathcal{L}(E)$.

6. Scalarisation

Le but de cette section est de démontrer le lemme suivant, qui sera très important pour nous. La preuve est loin d'être facile, et il n'est pas nécessaire de savoir la refaire. Mais bien entendu, il faut *absolument* connaitre le résultat et savoir l'utiliser.

LEMME 6.1. (lemme de scalarisation)
Soit E un evn. Pour tout $a \in E$, on peut trouver une forme linéaire continue $\Theta \in E^*$ telle que $\|\Theta\| \leq 1$ et $\Theta(a) = \|a\|$.

Remarque. Si $a \neq 0$, on a alors en fait $\|\Theta\| = 1$ puisque $\|\Theta\| \times \|a\| \geq \Theta(a)$ et donc $\|\Theta\| \geq 1$ en divisant par $\|a\|$.

Démonstration du lemme lorsque $E = (\mathbb{R}^n, \| \cdot \|_\infty)$. Dans ce cas, c'est très simple et il *faut* savoir refaire la preuve. Soit $a \in E$ quelconque. Par définition de la norme $\| \cdot \|_\infty$, on peut trouver $j_0 \in \{1, \dots, n\}$ tel que $|a_{j_0}| = \|a\|_\infty$. Soit $\varepsilon_0 = \pm 1$ tel que $\varepsilon_0 a_{j_0} = |a_{j_0}|$, et soit $\Theta \in E^*$ la forme linéaire défijnie par $\Theta(x_1, \dots, x_n) = \varepsilon_0 x_{j_0}$; autrement dit et avec les notations de l'exemple 2 donné après la définition 5.3, $\Theta = \Theta_b$ où $b = (0, \dots, \varepsilon_0, \dots 0)$ (avec ε_0 à la place j_0). Alors $\|\Theta\| = \|b\|_1 = 1$ et $\Theta(a) = |a_{j_0}| = \|a\|_\infty$. □

Exercice. Démontrer le lemme de scalarisation lorsque $E = (\mathbb{R}^n, \| \cdot \|_2)$.

Démonstration du lemme dans le cas où $\dim(E) < \infty$. Ici, on ne peut pas s'en sortir en invoquant l'équivalence des normes : c'est un peu plus compliqué. Appelons *solution partielle* (au problème qu'on est en train de considérer) toute forme linéaire (continue) $\Phi : F \to \mathbb{R}$ définie sur un sous-espace vectoriel $F \subset E$ contenant a, telle que $\|\Phi\| \leq 1$ et $\Phi(a) = \|a\|$. Avec cette terminologie, ce qu'il faut faire est montrer qu'il existe une solution partielle Θ *définie sur E tout entier*. Dans la suite, on supposera $a \neq 0$ (si $a = 0$, il suffit de prendre $\Theta = 0$).

Fait 1. Il existe au moins une solution partielle Φ_0 définie sur $F_0 = \mathbb{R}a$.

Démonstration. Soit $\Phi_0 : F_0 \to \mathbb{R}$ la forme linéaire définie comme suit :

$$\forall \lambda \in \mathbb{R} \ : \ \Phi_0(\lambda a) = \|a\| \, \lambda \, .$$

Pour $u = \lambda a \in F_0$, on a $|\Phi_0(u)| = |\lambda| \, \|a\| = \|u\|$, donc $\|\Phi_0\| \leq 1$; et bien sûr $\Phi_0(a) = \|a\|$. Ainsi, Φ_0 est une solution partielle. □

Fait 2. Si $\Phi : F \to \mathbb{R}$ est une solution partielle et si $e \in E \setminus F$, alors Φ peut se prolonger en une solution partielle définie sur $F \oplus \mathbb{R}e$.

Démonstration. Fixons $\Phi : F \to \mathbb{R}$ et $e \in E \setminus F$. Il est clair que toute forme linéaire $\widetilde{\Phi} : F \oplus \mathbb{R}e \to \mathbb{R}$ prolongeant Φ vérifie $\widetilde{\Phi}(a) = \|a\|$. Donc il s'agit simplement d'en trouver une qui vérifie aussi $\|\widetilde{\Phi}\| \leq 1$, autrement dit $|\widetilde{\Phi}(u)| \leq \|u\|$ pour tout $u \in F \oplus \mathbb{R}\, e$. En fait, il suffira de vérifier qu'on a

$$\widetilde{\Phi}(u) \leq \|u\|$$

pour tout $u \in F \oplus \mathbb{R}e$, car si on applique cette inégalité avec $-u$ au lieu de u on obtient $-\widetilde{\Phi}(u) \leq \| - u\| = \|u\|$ (autrement dit $\widetilde{\Phi}(u) \geq -\|u\|$) et donc au total $|\widetilde{\Phi}(u)| \leq \|u\|$.

Si $\widetilde{\Phi} : F \oplus \mathbb{R}e \to \mathbb{R}$ est une forme linéaire prolongeant Φ, alors $\widetilde{\Phi}$ est entièrement déterminée par $\alpha := \widetilde{\Phi}(e)$: pour tout $u = v + \lambda e \in F \oplus \mathbb{R}e$, on a

$$\widetilde{\Phi}(u) = \Phi(v) + \alpha\lambda \, .$$

Il s'agit donc de montrer qu'il existe un nombre réel α tel que

$$\Phi(v) + \alpha\lambda \leq \|v + \alpha e\|$$

pour tout couple $(v, \lambda) \in F \times \mathbb{R}$. Pour $\lambda = 0$, l'inégalité devient $\Phi(v) \leq \|v\|$, ce qui est vérifé car $\|\Phi\| \leq 1$; donc on peut supposer $\lambda \neq 0$. Il reste donc à prouver l'existence d'un $\alpha \in \mathbb{R}$ tel que

$$\forall \lambda > 0 \ \forall v \in F \ : \ \Phi(v) \pm \alpha\lambda \leq \|v \pm \alpha\lambda\| \, .$$

En divisant par λ et en manipulant les inégalités, cela s'écrit encore

$$\forall (\lambda, v) \in]0, \infty[\times F \ : \ \Phi\left(\frac{v}{\lambda}\right) - \left\|\frac{v}{\lambda} - e\right\| \leq \alpha \leq \left\|\frac{v}{\lambda} + e\right\| - \Phi\left(\frac{v}{\lambda}\right) .$$

Pour montrer qu'un tel α existe bien, il suffit de vérifier qu'on a

$$\sup\{\Phi(x) - \|x - e\|;\ x \in F\} \leq \inf\{\|y + e\| - \Phi(y);\ y \in F\}\,, \tag{6.1}$$

car alors n'importe quel α compris entre le sup et l'inf conviendra.

Comme $\|\Phi\| \leq 1$, on a

$$\Phi(x) + \Phi(y) \leq \|x + y\| \leq \|x - e\| + \|e + y\|$$

et donc $\Phi(x) - \|x - e\| \leq \|y + e\| - \Phi(y)$ pour tous $x, y \in F$. En prenant le sup en x et l'inf en y on obtient (6.1), ce qui termine la preuve du fait 2. □

La démonstration est maintenant presque terminée. Comme E est de dimension finie, on peut choisir $e_1, \dots, e_N \in E$ tels que $(a, e_1, \dots, e_N)$ soit une base de E. Posons $F_0 = \mathbb{R}a$ et $F_i = \mathbb{R}a \oplus \mathbb{R}e_1 \oplus \cdots \oplus \mathbb{R}e_i$ pour $i \in \{1, \dots, N\}$. En partant de la solution partielle Φ_0 définie sur F_0 (donnée par le fait 1) et en appliquant N fois le fait 2, on obtient des solutions partielles $\Phi_1, \dots, \Phi_N$ définies sur $F_1, \dots, F_N$. Comme $F_N = E$, la forme linéaire $\Theta := \Phi_N$ convient. □

Démonstration dans le cas général. La preuve est en fait la même, mais lorsque E n'est pas de dimension finie on a besoin d'utiliser un résultat de théorie des ensembles qu'on appelle le **lemme de Zorn**. Sans entrer dans les détails, ce lemme assure l'existence d'une solution partielle $\Theta : F \to \mathbb{R}$ qui est *maximale* au sens où elle ne peut pas être prolongée en une solution partielle définie sur un sous-espace strictement plus grand que F. D'après le fait 2 (dont la preuve n'utilise pas la dimension finie), on a alors nécessairement $F = E$, et donc la forme linéaire Θ convient. □

Corollaire 6.2. *Soit $M \in \mathbb{R}_+$. Pour tout vecteur $\xi \in E$, on a l'équivalence*

$$\|\xi\| \leq M \iff \forall \Theta \in F^* \ :\ |\Theta(\xi)| \leq M\,\|\Theta\|\,.$$

Démonstration. L'implication "$\Longrightarrow$" découle de la définition de la norme d'une forme linéaire, et la réciproque est une conséquence immédiate du lemme. □

Corollaire 6.3. *Si E est un evn, alors son dual E^* **sépare les points de** E : si $u, v \in E$ et $u \neq v$, alors on peut trouver une forme linéaire $\Theta \in E^*$ telle que $\Theta(u) \neq \Theta(v)$. Dit autrement : si on a $\Theta(u) = \Theta(v)$ pour toute forme linéaire $\Theta \in E^*$, alors $u = v$.*

Démonstration. Il suffit d'appliquer le lemme à $a = v - u$: cela fournit $\Theta \in E^*$ telle que $\Theta(v) - \Theta(u) = \Theta(v - u) = \|v - u\| \neq 0$. □

Exercice. Démontrer directement le corollaire (i.e. sans utiliser le lemme de scalarisation) lorsque E est de dimension finie.

Remarque. Habituellement, le lemme de scalarisation est déduit d'un théorème célèbre qu'on appelle le **théorème de Hahn-Banach** (dont la preuve est à peu près identique à celle donnée plus haut). Comme se cultiver n'a jamais fait de mal, il n'est pas interdit d'aller voir dans un livre d'*analyse fonctionnelle* (ou sur internet) ce que dit ce fameux théorème.

CHAPITRE 2

Fonctions d'une variable

1. Dérivation des fonctions à valeurs vectorielles

1.1. Définitions. Dans ce qui suit, I est un intervalle de $\mathbb{R}$ et F est un evn.

DÉFINITION 1.1. *On dit qu'une fonction $\varphi : I \to F$ est **dérivable** en un point $t_0 \in I$ si le quotient*

$$\frac{\varphi(t_0+h)-\varphi(t_0)}{h}$$

admet une limite dans F quand $h \to 0$. Dans ce cas, on pose

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

*et on dit que $\varphi'(t_0)$ est la **dérivée** de φ en t_0, ou encore le **vecteur dérivé** de φ en t_0. On dit que φ est dérivable **sur** I si elle est dérivable en tout point $t_0 \in I$.*

Remarque 1. L'expression "vecteur dérivé" est meilleure : il est physiquement correct de considérer une dérivée comme un vecteur et non comme un "point" de F (penser au vecteur vitesse).

Remarque 2. On écrira parfois $\frac{d\varphi}{dt}$ au lieu de φ'.

REFORMULATION. La fonction φ est dérivable en t_0 si et seulement si elle admet un **développement limité à l'ordre 1** en t_0 ; autrement dit, s'il existe un vecteur $l \in F$ tel qu'on puisse écrire

$$\varphi(t_0+h) = \varphi(t_0) + h\,l + h\,\varepsilon(h)\,, \tag{1.1}$$

où $\varepsilon(h) \in F$ et $\varepsilon(h) \to 0$ quand $h \to 0$. Dans ce cas, on a $l = \varphi'(t_0)$.

DÉMONSTRATION. Si φ est dérivable en t_0, alors on peut écrire $\frac{\varphi(t_0+h)-\varphi(t_0)}{h} = \varphi'(t_0)+\varepsilon(h)$, où $\varepsilon(h)$ tend vers 0 quand $h \to 0$; d'où (1.1) avec $l = \varphi'(t_0)$ en multipliant par h. Inversement, si (1.1) est vérifiée alors $\frac{\varphi(t_0+h)-\varphi(t_0)}{h} = l+\varepsilon(h)$, donc φ est dérivable en t_0 avec $\varphi'(t_0) = l$. $\square$

Exemple. Soit $\varphi : I \to \mathbb{R}^m$, et écrivons $\varphi(t) = \begin{pmatrix} \varphi_1(t) \\ \vdots \\ \varphi_m(t) \end{pmatrix}$. Alors φ est dérivable en t_0 si et seulement si les fonctions numériques $\varphi_1, \dots, \varphi_m$ le sont ; et dans ce cas, on a

$$\varphi(t_0) = \begin{pmatrix} \varphi_1'(t) \\ \vdots \\ \varphi_m'(t) \end{pmatrix}.$$

DÉMONSTRATION. C'est clair puisque la convergence dans $\mathbb{R}^m$ est la convergence coordonnée par coordonnée.

$\square$

Exercice. Montrer que si $\varphi : I \to F$ est dérivable en un point t_0, alors elle continue en t_0.

DÉFINITION 1.2. *On dit qu'une fonction $\varphi : I \to F$ est* **de classe $\mathcal{C}^1$ sur** *I si φ est dérivable sur I et si la fonction $\varphi' : I \to F$ est continue sur I.*

NOTATION. On notera $\mathcal{C}^1(I, F)$ l'ensemble de toutes les fonctions $\varphi : I \to F$ de classe $\mathcal{C}^1$. Bien entendu, on peut définir par récurrence la notion de fonction k **fois dérivable** et de fonction **de classe** $\mathcal{C}^k$ pour tout $k \in \mathbb{N}^*$. On notera $\mathcal{C}^k(I, F)$ l'ensembles des fonctions de classe $\mathcal{C}^k$ de I dans F. Par analogie, on écrira $\mathcal{C}^0(I, F)$ pour l'ensemble des fonctions *continues.*

Exercice. Soit $[a, b]$ un intervalle compact de $\mathbb{R}$. Montrer que si F est un espace de Banach, alors l'espace $\mathcal{C}^0([a, b], F)$ est complet pour la norme $\|\cdot\|_\infty$ définie par $\|u\|_\infty = \sup\{\|u(t)\|;\ t \in [a, b]\}$.

1.2. Deux calculs importants. Dans ce qui suit, les lettres E, F, G désignent des evn, et I est un intervalle de $\mathbb{R}$.

NOTATION. Si $L : E \to F$ est une application linéaire continue. On écrira souvent Lu au lieu de $L(u)$ pour désigner l'image par L d'un point $u \in E$. Il faut faire l'effort de s'habituer à cette notation, car elle est réellement très commode.

LEMME 1.3. (effet d'une application linéaire)
Soit $\varphi : I \to E$ et soit $L \in \mathcal{L}(E, F)$. Notons $L\varphi : I \to F$ la fonction composée $L \circ \varphi$: $(L\varphi)(t) = L\varphi(t) = L(\varphi(t))$. Si φ est dérivable, alors $L\varphi$ l'est également, et on a

$$(L\varphi)'(t) = L\varphi'(t)\,.$$

DÉMONSTRATION. Par linéarité de L, on a

$$\frac{L\varphi(t+h) - L\varphi(t)}{h} = L\left(\frac{\varphi(t+h) - \varphi(t)}{h}\right),$$

et ceci tend vers $L(\varphi'(t)) = L\varphi'(t)$ quand $h \to 0$, par continuité de L. $\square$

LEMME 1.4. (dérivée d'un produit)
Soient $u : I \to E$ et $v : I \to F$. Soit également $B : E \times F \to G$ une application bilinéaire continue. Notons $B(u, v) : I \to G$ la fonction $t \mapsto B(u(t), v(t))$. Si u et v sont dérivables, alors $B(u, v)$ l'est également, et on a

$$B(u, v)'(t) = B(u'(t), v(t)) + B(u(t), v'(t))\,.$$

Remarque. Écrivons "$u \cdot v$" au lieu de "$B(u, v)$" (une application bilinéaire est un "produit"...). Alors la formule précédente prend un aspect beaucoup plus familier, qui explique le titre du lemme :

$$(u \cdot v)' = u' \cdot v + u \cdot v'\,.$$

DÉMONSTRATION DU LEMME. On utilise la notation $u \cdot v$ plutôt que $B(u, v)$. Soit $t \in I$ fixé. Par bilinéarité, on a

$$(u \cdot v)(t+h) - (u \cdot v)(t) = (u(t+h) - u(t)) \cdot v(t+h) + u(t) \cdot (v(t+h) - v(t))$$

et donc (en utilisant à nouveau la bilinéarité)

$$\frac{(u \cdot v)(t+h) - (u \cdot v)(t)}{h} = \left(\frac{u(t+h) - u(t)}{h}\right) \cdot v(t+h) + u(t) \cdot \left(\frac{v(t+h) - v(t)}{h}\right).$$

Comme u et v sont dérivables, que v est continue (car dérivable) et que l'opération "$\cdot$" est continue, le membre de droite tend vers $u'(t) \cdot v(t) + u(t) \cdot v'(t)$ quand $h \to 0$, ce qui est la conclusion souhaitée. $\square$

Exercice 1. Soient $u, v : I \to \mathbb{R}^n$ et $M, N : I \to M_n(\mathbb{R})$ des fonctions dérivables. Quelles sont les dérivées des fonctions $t \mapsto \langle u(t), v(t) \rangle$ et $t \mapsto M(t)N(t)$?

Exercice 2. Soit $\gamma : I \to \mathbb{R}^n$ une fonction dérivable. On suppose que $\|\gamma(t)\|_2$ reste constante sur I. Montrer que $\gamma'(t)$ est orthogonal à $\gamma(t)$, pour tout $t \in I$.

1.3. L'inégalité des accroissements finis. Dans ce qui suit, $[a, b]$ est un intervalle compact de $\mathbb{R}$. Rappelons la "formule des accroissements finis" pour les fonctions réelles d'une variable réelle :

FORMULE DES ACCROISSEMENTS FINIS. *Si $\varphi : [a, b] \to \mathbb{R}$ est continue sur $[a, b]$ et dérivable sur $]a, b[$, alors il existe un point $c \in]a, b[$ tel que $\varphi(b) - \varphi(a) = (b - a)\, \varphi'(c)$.*

DÉMONSTRATION. Supposons d'abord qu'on a $\varphi(b) = \varphi(a)$. Dans ce cas, il s'agit de montrer qu'il existe $c \in]a, b[$ tel que $\varphi'(c) = 0$ (c'est le **théorème de Rolle**). Si φ est constante, il n'y a rien à démontrer. Sinon, φ prend par exemple au moins une fois une valeur strictement inférieure à $\varphi(a) = \varphi(b)$. Comme φ est continue sur le compact $[a, b]$, on peut trouver $c \in [a, b]$ tels que $\varphi(c) = \min\{\varphi(t);\ t \in [a, b]\}$. Alors $c \neq a, b$ par hypothèse sur φ, autrement dit c est *intérieur* à l'intervalle $[a, b]$; et on sait qu'alors on doit avoir $\varphi'(c) = 0$.

Dans le cas général, on applique le théorème de Rolle à la fonction "auxiliaire" ψ définie par

$$\psi(t) = \varphi(t) - \frac{\varphi(b) - \varphi(a)}{b - a}\,(t - a)\,,$$

qui vérifie $\psi(b) = \varphi(a) = \psi(a)$.

$\square$

Exercice. Soient f et g deux fonctions continues sur $[a, b]$ et dérivables sur $]a, b[$, avec $g'(t) \neq 0$ pour tout $t \in]a, b[$. Établir la "formule des accroissements finis généralisée" : il existe $c \in]a, b[$ tel que

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}\,.$$

La formule des accroissements finis n'est plus valable pour une fonction $\varphi : [a, b] \to F$ à valeurs dans un evn F de dimension strictement plus grande que 1. Par exemple, la fonction $\varphi : [0, 2\pi] \to \mathbb{C}$ définie par $\varphi(t) = e^{it}$ est de classe $\mathcal{C}^1$ avec $\varphi(2\pi) = \varphi(0)$ et $\varphi'(t) = ie^it \neq 0$ pour tout $t \in [0, 2\pi]$, donc on ne peut certainement trouver aucun c tel que $\varphi(2\pi) - \varphi(0) = (2\pi - 0)\, \varphi'(c)$.

Cependant, pour les fonctions à valeurs vectorielles il existe un substitut au théorème des accroissements finis sous la forme d'une **inégalité**, qui dans la pratique rend exactement les mêmes services :

THÉORÈME 1.5. (inégalité des accroissements finis)
Soit F un evn. Si $\varphi : [a, b] \to F$ est une fonction continue sur $[a, b]$ et dérivable sur $]a, b[$, alors il existe un point $c \in]a, b[$ tel que $\|\varphi(b) - \varphi(a)\| \leq (b - a)\, \|\varphi'(c)\|$.

DÉMONSTRATION. D'après le lemme de scalarisation, il existe une forme linéaire $\Theta \in F^*$ telle que $\|\Theta\| \leq 1$ et

$$\Theta(\varphi(b) - \varphi(a)) = \|\varphi(b) - \varphi(a)\|\,.$$

Définissons alors $\widetilde{\varphi} : [a,b] \to \mathbb{R}$ par

$$\widetilde{\varphi}(t) = \Theta\varphi(t) = \Theta(\varphi(t))\,.$$

La fonction $\widetilde{\varphi}$ est continue sur $[a,b]$ par composition, et d'après le lemme 1.3 elle est dérivable sur $]a,b[$ avec

$$\widetilde{\varphi}'(t) = \Theta(\varphi'(t))\,.$$

D'après la formule des accroissements finis (applicable puisque $\widetilde{\varphi}$ est à valeurs réelles), on peut trouver $c \in]a,b[$ tel que $\widetilde{\varphi}(b) - \widetilde{\varphi}(a) = (b-a)\,\widetilde{\varphi}'(c) = \Theta(\varphi'(c))$. On en déduit

$$\begin{aligned} \|\varphi(b) - \varphi(a)\| &= \Theta(\varphi(b)) - \Theta(\varphi(a)) \\ &= \widetilde{\varphi}(b) - \widetilde{\varphi}(a) \\ &= (b-a)\,\Theta(\varphi'(c)) \\ &\leq (b-a)\|\varphi'(c)\|\,, \end{aligned}$$

où on a utilisé la linéarité de Θ à la première ligne et le fait que $\|\Theta\| \leq 1$ à la dernière. □

Remarque. Cette démonstration illustre un "principe" très utile : pour dire quelque chose d'intéressant sur une fonction φ à valeurs vectorielles, il est souvent judicieux de **scalariser**, c'est-à-dire de composer φ avec une forme linéaire bien choisie.

COROLLAIRE 1.6. *Soit I un intervalle de $\mathbb{R}$, et soit $\varphi : I \to F$ une fonction dérivable à valeurs dans un evn F. Supposons qu'on ait $\|\varphi'(t)\| \leq M$ sur I, pour une certaine constante M. Alors φ est M-lipschitzienne sur I :*

$$\forall x, y \in I \ :\ \|\varphi(y) - \varphi(x)\| \leq M\,\|y - x\|\,.$$

DÉMONSTRATION. C'est immédiat en appliquant l'inégalité des accroissements finis sur l'intervalle $[a,b] = [x,y]$ (ou $[y,x]$ si $x > y$). □

COROLLAIRE 1.7. *Sous les hypothèses précédentes, si $\varphi'(t) \equiv 0$ sur I, alors φ est constante sur l'intervalle I.*

DÉMONSTRATION. Appliquer le corollaire précédent avec $M = 0$. □

COROLLAIRE 1.8. *Si $\varphi : [a,b] \to F$ est de classe $\mathcal{C}^1$ sur l'intervalle* fermé borné $[a,b]$, *alors φ est lipschitzienne.*

DÉMONSTRATION. La fonction $t \mapsto \|\varphi'(t)\|$ est continue sur le *compact* $[a,b]$, donc elle y est bornée : on a ainsi une constante M telle que $\|\varphi'(t)\| \leq M$ sur $[a,b]$, et on peut appliquer le corollaire 1.6. □

2. Intégration des fonctions à valeurs vectorielles

2.1. Définition de l'intégrale. Dans ce qui suit, $[a,b]$ est un intervalle compact de $\mathbb{R}$ et F est un evn *complet* (par exemple un evn de dimension finie).

LEMME 2.1. *Si $u : [a,b] \to F$ est une fonction continue, il existe un unique $\xi \in F$ tel que*

$$\forall \Theta \in F^* \ :\ \Theta(\xi) = \int_a^b \Theta(u(t))\,dt\,.$$

*On dit que ξ est l'**intégrale de** u **sur** $[a,b]$, et on écrit $\xi = \int_a^b u(t)\,dt$.*

Démonstration. L'unicité vient du fait que F^* sépare les points de F (corollaire 6.3) : si ξ_1 et ξ_2 vérifient la propriété du lemme, alors $\Theta(\xi_1) = \Theta(\xi_2)$ pour toute $\Theta \in F^*$ et donc $\xi_1 = \xi_2$.

L'existence est très facile à démontrer dans le cas où F est de dimension finie : si $(e_1, \ldots, e_m)$ est une base de F, on peut écrire

$$u(t) = \sum_{i=1}^{m} x_i(t)\, e_i$$

où les x_i sont des fonctions continues à valeurs réelles déterminées de manière unique, et il suffit de poser

$$\xi = \sum_{i=1}^{m} \left(\int_a^b x_i(t)\, dt \right) e_i\,.$$

Exercice. Vérifier que ξ convient effectivement.

L'existence dans le cas général est plus difficile à établir. On va utiliser un argument de **sommes de Riemann**. Pour $k \in \mathbb{N}^*$, soit ξ_k l'élément de F défini par

$$\xi_k = \frac{b-a}{k} \sum_{i=0}^{k-1} u\left(a + i\,\frac{b-a}{k} \right)\,.$$

Admettons provisoirement que la suite (ξ_k) converge dans F, et notons ξ sa limite. Si $\Theta \in F^*$, alors (par linéarité)

$$\Theta(\xi_k) = \frac{b-a}{k} \sum_{i=0}^{k-1} \Theta \circ u \left(a + i\,\frac{b-a}{k} \right)$$

pour tout $k \in \mathbb{N}^*$. On reconnait une somme de Riemann associée à la fonction continue *à valeurs réelles* $\Theta \circ u$ et à la subdivision $\sigma_k = \left(a, a + \frac{b-a}{k}, \ldots, a + (k-1)\,\frac{b-a}{k}, b\right)$, de “pas” égal à $(b-a)/k$. Par conséquent, $\Theta(\xi_k)$ tend vers $\int_a^b \Theta \circ u(t)\, dt$ quand $k \to \infty$, et comme Θ est continue on en déduit qu'on a $\Theta(\xi) = \int_a^b \Theta(u(t))\, dt$. Ceci étant vrai pour toute forme linéaire $\Theta \in F^*$, on voit que $\xi = \lim \xi_k$ convient.

Pour achever la démonstration, il suffit donc de montrer que la suite (ξ_k) converge dans F ; et comme F est supposé *complet*, on peut se contenter de vérifier que (ξ_k) est une suite de Cauchy. Pour alléger les notations, on posera

$$t_{k,i} = a + i\,\frac{b-a}{k}$$

pour $k \in \mathbb{N}^*$ et $i \in \{0, \ldots, k\}$. Ainsi, on a

$$\xi_k = \frac{b-a}{k} \sum_{i=0}^{k-1} u(t_{k,i})\,.$$

Enfin, on notera pour tout $\delta > 0$:

$$\omega_u(\delta) = \sup \left\{ \|u(t) - u(t')\|;\ |t - t'| \leq \delta \right\}\,.$$

Comme u est *uniformément continue* sur le compact $[a, b]$, on sait que $\omega_u(\delta)$ tend vers 0 quand $\delta \to 0$.

Le point clé est de montrer que $\Theta(\xi_k)$ tend vers $\int_a^b \Theta(u(t))\, dt$ pour toute forme linéaire continue $\Theta \in F^*$, avec une convergence *uniforme par rapport à* Θ lorsque $\|\Theta\|$ reste bornée. C'est le contenu du fait suivant.

Fait. Si $\Theta \in F^*$, alors on a la majoration suivante pour tout $k \in \mathbb{N}^*$:

$$\left|\Theta(\xi_k) - \int_a^b \Theta(u(t))\,dt\right| \leq \|\Theta\| \times (b-a) \times \omega_u\left(\frac{b-a}{k}\right).$$

Preuve du fait. D'après la relation de Chasles, on a

$$\int_a^b \Theta(u(t))\,dt = \sum_{i=0}^{k-1} \int_{t_{k,i}}^{t_{q,i+1}} \Theta(u(t))\,dt$$

et d'autre part

$$\Theta(\xi_k) = \frac{b-a}{k} \sum_{i=0}^{k-1} \Theta(u(t_{k,i})) = \sum_{i=0}^{k-1} \int_{t_{k,i}}^{t_{k,i+1}} \Theta(u(t_{k,i}))\,dt\,.$$

En faisant la différence, on obtient

$$\Theta(\xi_k) - \int_a^b \Theta(u(t))\,dt = \sum_{i=0}^{k-1} \int_{t_{k,i}}^{t_{k,i+1}} [\Theta(u(t_{k,i})) - \Theta(u(t))]\,dt\,;$$

et on en déduit

$$\left|\Theta(\xi_k) - \int_a^b \Theta(u(t))\,dt\right| \leq \|\Theta\| \times \sum_{i=0}^{k-1} \int_{t_{k,i}}^{t_{k,i+1}} \|u(t_{k,i}) - u(t)\|\,dt\,.$$

Chaque terme de la somme étant inférieur ou égal à $\omega_u\left(\frac{b-a}{k}\right)$ (car $|t - t_{k,i}| \leq \frac{b-a}{k}$ si $t \in [t_{k,i}, t_{k,i+1}]$), cela donne finalement

$$\left|\Theta(\xi_k) - \int_a^b \Theta(u(t))\,dt\right| \leq \|\Theta\| \times \omega_u\left(\frac{b-a}{k}\right) \times \sum_{i=0}^{k-1} \int_{t_{k,i}}^{t_{k,i+1}} dt\,,$$

d'où le résultat.

□

Soit maintenant $\varepsilon > 0$. Comme $\omega_u(\delta)$ tend vers 0 quand $\delta \to 0$, on peut trouver $K \in \mathbb{N}^*$ tel que $\omega_u\left(\frac{b-a}{k}\right) \leq \frac{\varepsilon}{b-a}$ pour tout $k \geq K$. D'après le fait, on a alors

$$\left|\Theta(\xi_k) - \int_a^b \Theta(u(t))\,dt\right| \leq \|\Theta\|\,\varepsilon$$

pour $\Theta \in F^*$et $k \geq K$; et d'après l'inégalité triangulaire on en déduit

$$|\Theta(\xi_q - \xi_p)| = |\Theta(\xi_q) - \Theta(\xi_p)| \leq 2\varepsilon\,\|\Theta\|\,,$$

pour tous $p, q \geq K$ et pour toute forme linéaire $\Theta \in F^*$. D'après le lemme de scalarisation, cela signifie qu'on a

$$\forall p, q \geq K\ :\ \|\xi_q - \xi_p\| \leq 2\varepsilon\,.$$

On a donc bien montré que la suite (ξ_k) est de Cauchy, ce qui achève la preuve.

□

REMARQUE. Pour une fonction u à valeurs dans $\mathbb{R}^m$, on peut oublier la démonstration précédente et ne retenir qu'une seule chose : si on écrit

$$u(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_m(t) \end{pmatrix},$$

alors

$$\int_a^b u(t)\,dt = \begin{pmatrix} \int_a^b x_1(t)dt \\ \vdots \\ \int_a^b x_m(t)dt \end{pmatrix}.$$

Exercice. Montrer que si $u \in \mathcal{C}^0([a,b],F)$ et si $L : F \to G$ est linéaire continue (où G est un autre espace de Banach), alors

$$L\left(\int_a^b u(t)\,dt\right) = \int_a^b Lu(t)\,dt\,.$$

La proposition suivante résume les propriété essentielles de l'intégrale, qui sont les mêmes que pour les fonctions à valeurs réelles.

PROPOSITION 2.2. (propriétés de l'intégrale)

(1) *L'intégrale est* ***linéaire*** *par rapport à la fonction intégrée : si $u, v \in \mathcal{C}^0([a,b],F)$ et si $\lambda, \mu \in \mathbb{R}$, alors*

$$\int_a^b (\lambda u(t) + \mu v(t))\,dt = \lambda \int_a^b u(t)\,dt + \mu \int_a^b v(t)\,dt\,.$$

(2) *Si $a < c < b$, alors $\int_a^b u(t)\,dt = \int_a^c u(t)\,dt + \int_c^b u(t)\,dt$ (**relation de Chasles**).*

(3) *Si $u \in \mathcal{C}^0([a,b],F)$, alors*

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

DÉMONSTRATION. (1) et (2) se déduisent immédiatement de la définition de l'intégrale et des propriétés correspondantes de l'intégrale des fonctions à valeurs réelles. Pour démontrer (3), on utilise le lemme de scalarisation : on a

$$\begin{aligned} \forall \Theta \in F^* \ :\ \left|\Theta\left(\int_a^b u(t)\,dt\right)\right| &= \left|\int_a^b \Theta(u(t))\,dt\right| \\ &\leq \int_a^b |\Theta(u(t))|\,dt \\ &\leq \|\Theta\| \times \int_a^b \|u(t)\|\,dt\,, \end{aligned}$$

d'où le résultat.

□

NOTATION. Soit I un intervalle de $\mathbb{R}$. Si $u : I \to F$ et continue et si $x, y \in I$ avec $x < y$, on pose $\int_y^x u(t)\,dt = -\int_x^y u(t)\,dt$. Avec cette convention, la relation de Chasles reste vraie *quel que soit l'ordre des points a, b, c.*

Exercice 1. On munit l'espace $\mathcal{C}^0([a,b],F)$ de la norme $\|\cdot\|_\infty$ définie par $\|u\|_\infty = \sup\{\|u(t)\|;\ t \in [a,b]\}$. Montrer que l'application $u \mapsto \int_a^b u(t)\,dt$ est continue de $\mathcal{C}^0([a,b],F)$ dans F.

Exercice 2. Montrer que si une suite $(u_k) \subset \mathcal{C}^0([a,b],F)$ converge **uniformément** vers une fonction u, alors $\int_a^b u_k(t)\,dt$ tend vers $\int_a^b u(t)\,dt$. (La convergence uniforme pour des fonctions à valeurs vectorielles se définit exactement comme pour les fonctions à valeurs réelles, en remplaçant les valeurs absolues par des normes. C'est la notion de convergence associée à la norme $\|\cdot\|_\infty$ sur $\mathcal{C}^0([a,b],F)$).

2.2. Le "théorème fondamental de l'analyse". Comme son nom l'indique, le théorème suivant est particulièrement important. Il se démontre exactement comme pour les fonctions à valeurs réelles, en utilisant uniquement les propriétés de base de l'intégrale (proposition 2.2).

Théorème 2.3. (théorème fondamental de l'analyse)
Soit I un intervalle de $\mathbb{R}$, et soit $a \in I$. Si $u : I \to F$ est continue, alors la fonction $\varphi : I \to F$ définie par $\varphi(t) = \int_a^t u(s)\,ds$ est de classe $\mathcal{C}^1$ et $\varphi' = u$.

Démonstration. Comme u est continue, il suffit de montrer que φ est dérivable en tout point, avec $\varphi' = u$.

Fixons $t_0 \in I$. Si $t_0+h \in I$, alors (d'après la relation de Chasles) $\varphi(t_0+h)-\varphi(t_0) = \int_a^{t_0+h} u(t)\,dt - \int_a^{t_0} u(t)\,dt = \int_{t_0}^{t_0+h} u(t)\,dt$, et donc

$$\frac{\varphi(t_0+h)-\varphi(t_0)}{h} = \frac{1}{h}\int_{t_0}^{t_0+h} u(t)\,dt\,.$$

D'autre part, on peut également écrire

$$u(t_0) = \frac{1}{h}\int_{t_0}^{t_0+h} u(t_0)\,dt\,,$$

et on en déduit

$$\frac{\varphi(t_0+h)-\varphi(t_0)}{h} - u(t_0) = \frac{1}{h}\int_{t_0}^{t_0+h} (u(t)-u(t_0))\,dt\,.$$

Par conséquent, on a

$$\left\|\frac{\varphi(t_0+h)-\varphi(t_0)}{h} - u(t_0)\right\| \leq \left|\frac{1}{h}\int_{t_0}^{t_0+h} \|u(t)-u(t_0)\|\,dt\right|.$$

(Il faut mettre une valeur absolue autour de l'intégrale dans le membre de droite, car on a peut-être $h<0$, auquel cas l'intégrale est négative).

En utilisant la continuité de u au point t_0, il est maintenant facile de conclure que $\frac{\varphi(t_0+h)-\varphi(t_0)}{h}$ tend vers $u(t_0)$ quand $h \to 0$. (*Écrire les détails* !). Ainsi, φ est dérivable en t_0 et $\varphi'(t_0) = u(t_0)$.

$\square$

Remarque. Le théorème fondamental de l'analyse montre en particulier que *toute fonction continue $u : I \to F$ possède des primitives.*

Le corollaire suivant est au moins aussi important que le théorème, et mérite également de s'appeler "théorème fondamental de l'analyse".

Corollaire 2.4. *Si $\varphi : I \to F$ est de classe $\mathcal{C}^1$, alors $\varphi(x) = \varphi(a) + \int_a^x \varphi'(t)\,dt$ pour tout $x \in I$.*

DÉMONSTRATION. Si on pose $\psi(x) = \varphi(a) + \int_a^x \varphi'(t)\,dt$, alors ψ est de classe $\mathcal{C}^1$ et $\psi' = \varphi'$ d'après le théorème appliqué à $u = \varphi'$. Par conséquent, la fonction $v : \varphi - \psi$ a une dérivée identiquement nulle. Donc v est constante, et la constante vaut 0 car $v(a) = \varphi(a) - \varphi(a) = 0$. □

Une autre conséquence est la formule d'intégration par parties :

COROLLAIRE 2.5. *Soient $\varphi_1 : [a,b] \to F_1$ et $\varphi_2([a,b] \to F_2$ de classe $\mathcal{C}^1$, et soit $B : F_1 \times F_2 \to G$ une application bilinéaire continue, notée $(u_1,u_2) \mapsto u_1 \cdot u_2$. Alors*

$$\int_a^b \varphi_1(t) \cdot \varphi_2'(t)\,dt = [\varphi_1 \cdot \varphi_2]_a^b - \int_a^b \varphi_1'(t) \cdot \varphi_2(t)\,dt\,.$$

DÉMONSTRATION. C'est évident puisque $(\varphi_1 \cdot \varphi_2)'(t) = \varphi_1'(t) \cdot \varphi_2(t) + \varphi_1(t) \cdot \varphi_2'(t)$. □

APPLICATION 1. Le théorème fondamental de l'analyse permet de retrouver très rapidement et de façon "mécanique" l'inégalité des accroissements finis (plus exactement, le corollaire 1.6) pour une fonction de classe $\mathcal{C}^1$ à valeurs dans un espace de Banach. En effet, si $\varphi : [a,b] \to F$ est de classe $\mathcal{C}^1$ et $\|\varphi'(t)\| \leq M$ sur $]a,b[$, il suffit d'écrire

$$\begin{aligned} \|\varphi(b) - \varphi(a)\| &= \left\| \int_a^b \varphi'(t)\,dt \right\| \\ &\leq \int_a^b \|\varphi'(t)\|\,dt \\ &\leq \int_a^b M\,dt \\ &= M(b-a)\,. \end{aligned}$$

Exercice. Soient $\varphi : [a,b] \to F$ de classe $\mathcal{C}^1$. On suppose qu'on a $\|\varphi'(t)\| \leq \rho(t)$ pour tout $t \in [a,b]$, où $\rho : [a,b] \to \mathbb{R}^+$ est une fonction continue. Montrer qu'on a $\|\varphi(b) - \varphi(a)\| \leq \int_a^b \rho(t)\,dt$.

APPLICATION 2. Comme deuxième illustration du théorème fondamental de l'analyse, on va démontrer un résultat très utile sur les suites de fonctions de classe $\mathcal{C}^1$. Pour des fonctions à valeurs réelles, ce résultat a été vu en 2è année (sous une forme sans doute plus générale).

PROPOSITION 2.6. *Soit I un intervalle de $\mathbb{R}$, et soit (φ_k) une suite de fonctions de classe $\mathcal{C}^1$ sur I, à valeurs dans F. Soit également $\varphi : I \to F$. On suppose que*
(i) *$\varphi_k(t) \to \varphi(t)$ pour tout $t \in I$;*
(ii) *la suite (φ_k') converge simplement vers une fonction $u : I \to F$, avec* convergence uniforme sur tout intervalle compact $J \subset I$.
Alors φ est de classe $\mathcal{C}^1$ et $\varphi' = u = \lim_{k\to\infty} \varphi_k'$.

Remarque. On l'a déjà dit : la convergence uniforme se définit exactement comme pour les fonctions à valeurs réelles, en remplaçant les valeurs absolues par des normes.

DÉMONSTRATION. Fixons un point $a \in I$. Si $x \in I$ alors

$$\varphi_k(x) = \varphi_k(a) + \int_a^x \varphi_k'(t)\,dt \tag{2.1}$$

pour tout $k \in \mathbb{N}$, d'après le théorème fondamental de l'analyse. Comme $\varphi_k' \to u$ *uniformément* sur $[a,x]$ (ou $[x,a]$ si $x > a$) par (ii), l'intégrale du second membre tend vers $\int_a^x u(t)\,dt$. Comme de plus $\varphi_k(a) \to \varphi(a)$ par (i), on obtient donc en passant à la limite dans (2.1) :

$$\forall x \in I \;:\; \varphi(x) = \varphi(a) + \int_a^x u(t)\,dt\,.$$

D'après le théorème fondamental de l'analyse, cela prouve que φ est de classe $\mathcal{C}^1$ avec $\varphi' = u$.

□

Corollaire 2.7. *Soit (v_k) une suite de fonctions de classe $\mathcal{C}^1$ sur I, à valeurs dans F. On suppose que la série $\sum v_k$ converge simplement sur I, et que la série $\sum v_k'$ converge* normalement sur tout compact de *I. Alors la fonction $v = \sum_0^\infty v_k$ est de classe $\mathcal{C}^1$, et $v' = \sum_0^\infty v_k'$.*

Démonstration. C'est immédiat en appliquant la proposition aux sommes partielles $\varphi_k = \sum_0^k v_j$, car la convergence normale de la série $\sum v_k'$ entraine la convergence uniforme de la suite (φ_k'). (On utilise ici le fait que l'espace F est *complet* : écrire les détails). □

Application 3. Comme troisième application du théorème fondamental de l'analyse (et également comme illustration d'autres idées vues dans ce chapitre), on va démontrer un résultat élémentaire mais très utile concernant les **intégrales à paramètres**.

Proposition 2.8. *Soit I et $[a,b]$ deux intervalles de $\mathbb{R}$, et soit $\Phi : I \times [a,b] \to \mathbb{R}$. On suppose que Φ admet en tout point $(t,x) \in I \times [a,b]$ une dérivée partielle $\frac{\partial \Phi}{\partial t}(t,x)$, et que la fonction $\frac{\partial \Phi}{\partial t}$ est continue sur $I \times [a,b]$. On définit $f : I \to \mathbb{R}$ par*

$$f(t) = \int_a^b \Phi(t,x)\,dx\,.$$

Alors f est de classe $\mathcal{C}^1$ sur I et on peut "dériver sous l'intégrale" :

$$f'(t) = \int_a^b \frac{\partial \Phi}{\partial t}(t,x)\,dx\,.$$

Démonstration. Pour $t \in I$, notons $\Phi(t,\cdot)$ la fonction $[a,b] \ni x \mapsto \Phi(t,x)$. Tout repose sur le fait suivant.

Fait. L'application $\varphi : I \to \mathcal{C}([a,b])$ définie par $\varphi(t) = \Phi(t,\cdot)$ est de classe $\mathcal{C}^1$, avec $\varphi'(t) = \frac{\partial \Phi}{\partial t}(t,\cdot)$ pour tout $t \in I$.

Preuve du fait. Fixons un point $t_0 \in I$. D'après le théorème fondamental de l'analyse, on a $\Phi(t,x) = \Phi(t_0,x) + \int_{t_0}^t \frac{\partial \Phi}{\partial s}(s,x)\,ds$ pour tout $x \in [a,b]$. En notant $\delta_x : \mathcal{C}([a,b]) \to \mathbb{R}$ la forme linéaire définie par $\delta_x(u) = u(x)$, cela s'écrit encore

$$\delta_x(\varphi(t)) = \delta_x(\varphi(t_0)) + \int_{t_0}^t \delta_x\left(\frac{\partial \Phi}{\partial s}(s,\cdot)\right) ds\,.$$

D'autre part, on sait d'après le corollaire 4.5 du chapitre 1 que l'application $s \mapsto \frac{\partial \Phi}{\partial s}(s,\cdot)$ est continue de I dans $\mathcal{C}([a,b])$. Donc l'intégrale vectorielle $\int_{t_0}^t \frac{\partial \Phi}{\partial s}(s,\cdot)\,ds$ est bien définie, et comme comme δ_x est une forme linéaire *continue* sur $\mathcal{C}([a,b])$ (exercice) on a

$$\int_{t_0}^t \delta_x\left(\frac{\partial \Phi}{\partial s}(s,\cdot)\right) ds = \delta_x\left(\int_{t_0}^t \frac{\partial \Phi}{\partial s}(s,\cdot)\,ds\right).$$

On en déduit $\delta_x(\varphi(t)) = \delta_x\Big(\varphi(t_0) + \int_{t_0}^t \frac{\partial\Phi}{\partial s}(s,\cdot)\,ds\Big)$, autrement dit

$$\varphi(t)(x) = \left(\varphi(t_0) + \int_{t_0}^t \frac{\partial\Phi}{\partial s}(s,\cdot)\,ds\right)(x)$$

pour tout $x \in [a,b]$. Ainsi, on a

$$\varphi(t) = \varphi(t_0) + \int_{t_0}^t \frac{\partial\Phi}{\partial s}(s,\cdot)\,ds$$

pour tout $t \in [a,b]$; et d'après le théorème fondamental de l'analyse on en déduit que φ est de classe $\mathcal{C}^1$ avec $\varphi'(t) = \frac{\partial\Phi}{\partial t}(t,\cdot)$.

$\square$

Si on introduit maintenant la forme linéaire $\Theta : \mathcal{C}([a,b]) \to \mathbb{R}$ définie par $\Theta(u) = \int_a^b u(x)\,dx$, alors on a par définition $f = \Theta \circ \varphi$. De plus, Θ est continue car $|\Theta(u)| \leq (b-a) \times \|u\|_\infty$ pour toute $u \in \mathcal{C}([a,b])$. D'après le fait et le lemme 1.3, on en déduit que f est de classe $\mathcal{C}^1$ avec $f'(t) = \Theta(\varphi'(t))$; ce qui s'écrit

$$f'(t) = \int_a^b \frac{\partial\Phi}{\partial x}(t,x)\,dx\,.$$

$\square$

3. La formule de Taylor

La formule de Taylor avec “reste intégrale” est un résultat fondamental de la théorie des fonctions numériques d'une variable réelle. La version “vectorielle” s'énonce exactement de la même façon :

Théorème 3.1. (formule de Taylor à l'ordre $r+1$)
Soient I un intervalle de $\mathbb{R}$, F un espace de Banach, $r \geq 0$ et $\varphi : I \to F$ de classe $\mathcal{C}^{r+1}$. Pour tous $a, b \in I$, on a

$$\varphi(b) = \varphi(a) + (b-a)\varphi'(a) + \cdots + \frac{(b-a)^r}{r!}\varphi^{(r)}(a) + \int_a^b \frac{(b-t)^r}{r!}\varphi^{(r+1)}(t)\,dt\,.$$

Remarque 3.2. Le “reste intégrale” peut s'écrire différemment : en posant $h = b-a$ et en utilisant le changement de variable $t = a + sh$, on obtient (*exercice*)

$$\int_a^b \frac{(b-t)^r}{r!}\varphi^{(r+1)}(t)\,dt = \int_0^1 \frac{(1-s)^r}{r!}\varphi^{(r+1)}(a+sh)\,h^{r+1}\,ds.$$

Preuve de la formule de Taylor. D'après le théorème fondamental de l'analyse, on a

$$\varphi(b) = \varphi(a) + \int_a^b \varphi'(t)\,dt\,,$$

ce qui est la formule souhaitée pour $r = 0$. Si on suppose $r \geq 1$ (et donc φ au moins $\mathcal{C}^2$) une intégration par parties donne

$$\begin{aligned}\int_a^b \varphi'(t)\,dt &= \big[-(b-t)\varphi'(t)\big]_a^b + \int_a^b (b-t)\varphi''(t)\,dt \\ &= (b-a)\varphi'(a) + \int_a^b (b-t)\varphi''(t)\,dt\,,\end{aligned}$$

et donc $\varphi(b) = \varphi(a) + (b-a)\varphi'(a) + \int_a^b (b-t)\varphi''(t)\,dt$. Si on suppose $r \geq 2$, on peut ré-intégrer par parties pour obtenir

$$\begin{aligned}\int_a^b (b-t)\varphi''(t)\,dt &= \left[-\frac{(b-t)^2}{2}\varphi''(t)\right]_a^b + \int_a^b \frac{(b-t)^2}{2}\varphi'''(t)\,dt \\ &= \frac{(b-a)^2}{2}\varphi''(a) + \int_a^b \frac{(b-t)^2}{2}\varphi'''(t)\,dt\,,\end{aligned}$$

et ainsi de suite. En répétant r fois ce raisonnement, on obtient la formule souhaitée. □

Remarque. L'intégration par parties de la preuve précédente peut sembler un peu mystérieuse : pourquoi intégrer par parties avec $-(b-t)$ et pas avec t ? Voici une manière un peu différente de procéder. On commence par écrire

$$\begin{aligned}\varphi(b) &= \varphi(a) + \int_a^b \varphi'(t)\,dt \\ &= \varphi(a) + \int_a^b \left(\varphi'(a) + \int_a^t \varphi''(s)\,ds\right)dt \\ &= \varphi(a) + (b-a)\varphi'(a) + \int_a^b \varphi''(s)\left(\int_s^b dt\right)ds \\ &= \varphi(a) + (b-a)\varphi'(a) + \int_a^b (b-s)\varphi''(s)\,ds\,.\end{aligned}$$

Ensuite, on itère ce raisonnement ; ou si on préfère, on démontre la formule de Taylor par récurrence sur r en écrivant

$$\int_a^b \frac{(b-t)^r}{r!}\varphi^{(r+1)}(t)\,dt = \int_a^b \frac{(b-t)^r}{r!}\left(\varphi^{(r+1)}(a) + \int_a^t \varphi^{(r+2)}(s)\,ds\right)dt$$

et en finissant le calcul comme plus haut.

Autre démonstration. On peut également *déduire* la formule de Taylor pour les fonctions à valeurs vectorielles de la formule correspondante pour les fonctions à valeurs réelles, en utilisant le lemme de scalarisation. Il suffit d'observer que si $\Theta \in F^*$, alors la fonctions à valeurs réelles $\Theta\varphi = \Theta \circ \varphi$ est de classe $\mathcal{C}^{r+1}$, avec $(\Theta\varphi)^{(k)}(t) = \Theta\varphi^{(k)}(t)$ pour tout $k \in \{0, \dots, r+1\}$ (preuve immédiate par récurrence). En notant ξ le second membre dans le formule de Taylor, on a donc

$$\begin{aligned}\Theta(\xi) &= \Theta\varphi(a) + (b-a)(\Theta\varphi)'(a) + \cdots + \frac{(b-a)^r}{r!}(\Theta\varphi)^{(r)}(a) \\ &\qquad + \Theta\left(\int_a^b \frac{(b-t)^r}{r!}\varphi^{(r+1)}(t)\,dt\right) \\ &= \Theta\varphi(a) + (b-a)(\Theta\varphi)'(a) + \cdots + \frac{(b-a)^r}{r!}(\Theta\varphi)^{(r)}(a) \\ &\qquad + \int_a^b \frac{(b-t)^r}{r!}(\Theta\varphi)^{(r+1)}(t)\,dt \\ &= \Theta(\varphi(b)),\end{aligned}$$

d'après la formule de Taylor appliquée à la fonction à valeurs réelles $\Theta\varphi$. Comme $\Theta \in F^*$ est quelconque, on a donc bien $\xi = \varphi(b)$ d'après le lemme de scalarisation. □

Corollaire 3.3. (inégalité de Taylor-Lagrange)
Si $\varphi : I \to F$ est de classe $\mathcal{C}^{r+1}$ et si $\|\varphi^{(r+1)}(t)\| \leq M$ sur I pour une certaine constante M alors, pour $a, x \in I$, on a

$$\left\|\varphi(x) - \left(\varphi(a) + (x-a)\varphi'(a) + \cdots + \frac{(x-a)^r}{r!}\,\varphi^{(r)}(a)\right)\right\| \leq M\,\frac{|x-a|^{r+1}}{(r+1)!}\,.$$

Démonstration. Posons $\xi = \varphi(x) - \left(\varphi(a) + (x-a)\varphi'(a) + \cdots + \frac{(x-a)^r}{r!}\,\varphi^{(r)}(a)\right)$. D'après la formule de Taylor, on a

$$\begin{aligned}\|\xi\| &= \left\|\int_0^1 \frac{(1-t)^r}{r!}\varphi'^{(r+1)}(a+t(x-a))\,(x-a)^{r+1}\,dt\right\| \\ &\leq \int_0^1 \|\cdots\|\,dt \\ &\leq M\,\frac{|x-a|^{r+1}}{r!}\int_0^1 (1-t)^r\,dt\,,\end{aligned}$$

d'où le résultat puisque $\int_0^1 (1-t)^r\,dt = \frac{1}{r+1}$.

□

Exercice. Soit $\varphi : [a,b] \to F$ de classe $\mathcal{C}^2$. On suppose qu'on a $\varphi'(a) = 0$ et $\|\varphi''(t)\| \leq \rho(t)$ pour tout $t \in [a,b]$, où $\rho : [a,b] \to \mathbb{R}^+$ est une fonction continue. Majorer $\|\varphi(b) - \varphi(a)\|$ par une intégrale faisant intervenir α.

4. Fonctions convexes

Définition 4.1. *Soit I un intervalle de $\mathbb{R}$. Une fonction $\varphi : I \to \mathbb{R}$ est dite **convexe** si "le graphe de f est en dessous de toutes ses cordes" ; ce qui s'écrit analytiquement comme suit :*

$$\forall u, v \in I\ \forall \lambda \in [0,1]\ :\ f((1-\lambda)u + \lambda v) \leq (1-\lambda)f(u) + \lambda f(v)\,. \tag{4.1}$$

*La fonction f est dite **concave** si $-f$ est convexe, autrement dit si l'inégalité (4.1) est inversée*

Exercice 1. Faire un dessin pour illustrer la définition.

Exercice 2. Montrer que la fonction $t \mapsto |t|$ est convexe sur $\mathbb{R}$.

Exercice 3. Soit $\varphi : I \to J$ une bijection croissante entre deux intervalles de $\mathbb{R}$. Montrer que si φ est convexe, alors φ^{-1} est concave.

Remarque. Une fonction $\varphi : I \to \mathbb{R}$ est à la fois convexe et concave si et seulement si elle est *affine*, i.e. de la forme $\varphi(t) = at + b$.

Démonstration. Il est "clair" qu'une fonction affine est à la fois convexe et concave (exercice). Inversement, supposons φ convexe et concave, autrement dit qu'on a

$$\varphi((1-\lambda)u + \lambda v) = (1-\lambda)\varphi(u) + \lambda\varphi(v)$$

pour tous $u, v \in I$ et pour tout $\lambda \in [0,1]$. Cela signifie que pour tous u, v dans I avec $u < v$, la fonction φ coïncide sur $[u,v]$ avec l'unique fonction affine $\alpha_{u,v}$ telle que $\alpha_{u,v}(u) = \varphi(u)$ et $\alpha_{u,v}(v) = \varphi(v)$. Ainsi, $\varphi(t)$ est de la forme $a_{u,v}t + b_{u,v}$ sur tout intervalle $[u,v] \subset I$. Comme deux fonctions affines qui coincident sur un intervalle non-trivial sont égales partout (il suffit qu'elles soient égales en deux points distincts), on voit très facilement que les coefficients $a_{u,v}$ et $b_{u,v}$ ne dépendent en fait pas de u et v ; donc φ est affine sur I. $\square$

On ne dira pas grand chose ici sur les fonctions convexes. En fait, on va se contenter de démontrer le résultat suivant.

Théorème 4.2. *Soit I un intervalle de $\mathbb{R}$ et soit $\varphi : I \to \mathbb{R}$ une fonction dérivable. Les propriétés suivantes sont équivalentes :*

(i) *φ est convexe ;*

(ii) *le graphe de φ est "au dessus de toutes ses tangentes", i.e.*

$$\forall x, y \in \mathring{I}\ :\ \varphi(y) \geq \varphi(x) + \varphi'(x)(y-x)\,;$$

(iii) *φ' est croissante.*

Exercice. Faire un dessin pour illustrer (ii).

Corollaire 4.3. *Soit $\varphi : I \to \mathbb{R}$ une fonction continue sur I et dérivable sur $\mathring{I}$. Alors φ est convexe sur I si et seulement si φ' est croissante sur $\mathring{I}$.*

Démonstration. Par continuité, φ est convexe sur I si et seulement si elle l'est sur $\mathring{I}$; donc le résultat découle immédiatement du théorème.

$\square$

Corollaire 4.4. *Une fonction φ continue sur I et deux fois dérivable sur $\mathring{I}$ est convexe si et seulement si $\varphi'' \geq 0$ sur $\mathring{I}$, et concave si et seulement si $\varphi'' \leq 0$.*

Corollaire 4.5. *La fonction $t \mapsto t^\alpha$ est convexe sur $[0, \infty[$ si $\alpha \geq 1$, et concave si $\alpha \in [0, 1]$. La fonction exponentielle est convexe sur $\mathbb{R}$, et la fonction logarithme est concave sur $]0, \infty[$.*

Preuve du théorème. Par continuité, φ est convexe sur I si et seulement si elle l'est sur $\mathring{I}$; donc on peut supposer que φ est dérivable *sur I*.

L'implication "(i) $\implies$ (iii)" va découler très facilement du fait suivant (qu'on appelle l'**inégalité des 3 pentes**). Il est *indispensable* d'illustrer l'énoncé par un dessin.

Fait. Si φ est convexe et si $u < v < w$, alors

$$\frac{\varphi(v)-\varphi(u)}{v-u} \leq \frac{\varphi(w)-\varphi(u)}{w-u} \leq \frac{\varphi(w)-\varphi(v)}{w-v}\,. \tag{4.2}$$

Preuve du fait. On écrit $v = (1-\lambda)u + \lambda w$, où $\lambda \in [0, 1]$ est donné par

$$\lambda = \frac{v-u}{w-u}\ ,\ 1-\lambda = \frac{w-v}{w-u}\cdot$$

L'inégalité de convexité $\varphi(v) \leq (1-\lambda)\varphi(u) + \lambda\varphi(w)$ donne d'une part

$$(1-\lambda)\left(\varphi(w)-\varphi(u)\right) \leq \varphi(w)-\varphi(v)\,,$$

qui est l'inégalité de droite dans (4.2) ; et d'autre part

$$\varphi(v)-\varphi(u) \leq \lambda\left(\varphi(w)-\varphi(u)\right),$$

qui est l'inégalité de gauche dans (4.2). $\square$

Preuve de "(i) $\Longrightarrow$ (iii)". Si φ est convexe alors, en faisant tendre v vers u puis vers w dans l'inégalité des 3 pentes (4.2), on obtient

$$\varphi'(u) \le \frac{\varphi(w)-\varphi(u)}{w-u} \le \varphi'(w)$$

pour tous $u, w \in I$ vérifiant $u < w$; donc φ' est croissante sur I. □

Preuve de "(iii) $\Longrightarrow$ (ii)". Supposons φ' croissante, et soient $x, y \in I$. D'après le théorème des accroissements finis, on peut trouver c entre x et y tel que $\varphi(y)-\varphi(x) = \varphi'(c)(x-y)$. Si $x \le y$, alors $c \ge x$, donc $\varphi'(c) \ge \varphi'(x)$ et donc $\varphi(y)-\varphi(x) \ge \varphi'(x)(y-x)$ puisque $y-x \ge 0$. Si $x \ge y$, alors $c \le x$, donc $\varphi'(c) \le \varphi'(x)$ et donc $\varphi(y)-\varphi(x) \ge \varphi'(x)(y-x)$ à nouveau puisque $y-x \le 0$. □

preuve de "(ii) $\Longrightarrow$ (i)". Supposons (ii) vérifiée. Pour montrer que φ est convexe, il suffit d'établir l'inégalité $\varphi((1-\lambda)u+\lambda v) \le (1-\lambda)\varphi(u)+\lambda\varphi(v)$ lorsque $u < v$ et $0 < \lambda < 1$.

Posons $x = (1-\lambda)u + \lambda v$, de sorte que $u < x < v$ et

$$\lambda = \frac{v-x}{v-u} \;,\; 1-\lambda = \frac{x-u}{v-u} \cdot$$

En appliquant (ii) avec $y = u$ et $y = v$, on obtient (puisque $u - x < 0$ et $v - x > 0$)

$$\frac{\varphi(u)-\varphi(x)}{u-x} \le \varphi'(x) \le \frac{\varphi(v)-\varphi(x)}{v-x}\,.$$

On en déduit

$$\varphi(x) \times \left(\frac{1}{v-x} - \frac{1}{u-x}\right) \le \frac{\varphi(v)}{v-x} - \frac{\varphi(u)}{u-x}\,,$$

autrement dit $\varphi(x) \le \frac{x-u}{v-u}\,\varphi(u) + \frac{v-x}{v-u}\,\varphi(v) = (1-\lambda)\varphi(u) + \lambda\,\varphi(v)$. □

□

Exemple. Soient p et q deux nombres réels strictement plus grand que 1 tels que

$$\frac{1}{p} + \frac{1}{q} = 1\,.$$

Par concavité de la fonction logarithme, on a $\log(\frac{1}{p}a^p + \frac{1}{q}a^q) \ge \frac{1}{p}\log(a^p) + \frac{1}{q}\log(b^q) = \log a + \log b = \log(ab)$ pour tous $a, b > 0$, autrement dit

$$ab \le \frac{1}{p}a^p + \frac{1}{q}b^q\,.$$

On en déduit que si $(a_i)_{i=1}^n$ et $(b_i)_{i=1}^n$ sont deux suites finies de nombres strictement positifs tels que $\sum_1^n a_i^p = 1 = \sum_1^n b_i^q$, alors

$$\sum_{i=1}^n a_i\, b_i \le \frac{1}{p} + \frac{1}{q} = 1\,.$$

Si maintenant les a_i et les b_i sont (strictement positifs et) quelconques et si on pose $\widetilde{a}_i = \frac{a_i}{A}$ et $\widetilde{b}_i = \frac{b_i}{B}$, où $A = (\sum_1^n a_i^p)^{1/p}$ et $B = (\sum_1^n b_i^q)^q$, alors $\sum_1^n \widetilde{a}_i^p = 1 = \sum_1^n \widetilde{b}_i^q$. On a donc $\sum_{i=1}^n \widetilde{a}_i\widetilde{b}_i \le 1$, autrement dit

$$\sum_{i=1}^n a_i\, b_i \le \left(\sum_{i=1}^n a_i^p\right)^{1/p} \left(\sum_{i=1}^n b_i^q\right)^q\,.$$

Cette inégalité s'appelle l'**inégalité de Hölder**.

Exercice 1. Montrer que si $\varphi : I \to \mathbb{R}$ est convexe, alors possède une dérivée à gauche et à droite en tout point de $\mathring{I}$, avec $\varphi'_g(s) \leq \varphi'_d(s) \leq \varphi'_g(t) \leq \varphi'_d(t)$ pour tous s, t vérifiant $s < t$.

Exercice 2. Montrer que si $\varphi : I \to \mathbb{R}$ est convexe, alors, pour tout $x \in \mathring{I}$, on peut trouver une fonction affine a telle que $a \leq \varphi$ et $a(x) = \varphi(x)$. En déduire que si $x_1, \ldots, x_n \in I$ et si $\lambda_1, \ldots, \lambda_n$ vérifient $\lambda_i \geq 0$ et $\sum_i \lambda_1^n = 1$, alors

$$\varphi(\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i \, \varphi(x_i) \, .$$

Exercice 3. Démontrer l'**inégalité arithmético-géométrique** : si $a_1, \ldots, a_n > 0$, alors

$$(a_1 \cdots a_n)^{1/n} \leq \frac{a_1 + \cdots + a_n}{n} \cdot$$

CHAPITRE 3

Applications différentiables

1. Définitions

1.1. Notations "o" et "O".

DÉFINITION 1.1. *Soient E et F deux evn, et soient $R : V \to F$ et $\alpha : V \to [0, \infty[$ deux applications définies sur une partie V de E*

(1) *Suposons que V soit un voisinage de 0. On dit que $R(h)$* ***est un petit o de*** *$\alpha(h)$* ***quand*** *h* ***tend vers*** *0, et on écrit $R(h) = o(\alpha(h))$, si on a $\alpha(h) > 0$ pour $h \neq 0$ suffisamment proche de 0, et $\lim\limits_{h\to 0} \frac{\|R(h)\|}{\alpha(h)} = 0$.*

(2) *On dit que $R(h)$* ***est un grand O de*** *$\alpha(h)$ (sur V), et on écrit $R(h) = O(\alpha(h))$, s'il existe une constante $C < \infty$ telle que $\|R(h)\| \leq C\,\alpha(h)$ pour tout $h \in V$.*

Remarque. Les notions de "o et de "O" ne changent pas si o remplace les normes par des normes équivalentes.

Exemple 0. Par définition, "$R(h) = o(1)$" signifie que $R(h)$ tend vers 0.

Exemple 1. On a $\|h\|^2 = o(\|h\|)$ quand $h \to 0$.

Exemple 2. Si $L : E \to F$ est linéaire continue, alors $L(h) = O(\|h\|)$ sur E.

DÉMONSTRATION. C'est immédiat d'après le critère de continuité pour les applications linéaires. □

Exemple 3. Si $B : E \times E \to F$ est bilinéaire continue, alors $B(h, k) = O(\|h\|\,\|k\|)$, et donc $B(h, k) = o(\|(h, k)\|)$ quand $(h, k) \to 0$.

DÉMONSTRATION. La première partie découle du critère de continuité pour les applications bilinéaires. Pour la seconde, il suffit de remarquer qu'on a $\|h\|\,\|k\| \leq \|(h, k)\|^2$ par définition de la norme produit ($\|(h, k)\| = \max(\|h\|, \|k\|)$), ce qui entraine que $B(h, k) = O(\|(h, k)\|^2)$ au voisinage de 0.

□

Exemple 4. Si $E = M_n(\mathbb{R}) = F$ et si $\|\cdot\|$ est une norme quelconque sur $M_n(\mathbb{R})$, alors $H^2 = o(\|H\|)$ quand $H \to 0$.

DÉMONSTRATION. On peut le voir comme une conséquence de l'exemple 3 en considérant l'application bilinéaire $B : M_n(\mathbb{R}) \times M_n(\mathbb{R}) \to M_n(\mathbb{R})$ définie par $B(H, K) = HK$. Cette application B est continue car on est en dimension finie (ou simplement parceque les coefficients de HK sont des fonctions polynomiales de ceux de H et K). On a donc $H^2 = B(H, H) = o(\|(H, H)\|) = o(\|H\|)$ quand $H \to 0$.

On peut aussi raisonner comme suit : par équivalence des normes, on peut supposer que $\|\cdot\|$ est une norme matricielle ; et dans ce cas tout est clair car $\|H^2\| \leq \|H\|^2$ et donc $H^2 = O(\|H\|^2)$. □

Remarque. Les notations "o" et "O" sont extrêmement commodes, et il *faut* faire l'effort de s'y habituer. Par exemple, on peut écrire (exactement comme quand on fait des développements limités) des identités du style

$$\begin{aligned} o(\alpha(h)) \pm o(\alpha(h)) &= o(\alpha(h)), \\ O(\alpha(h)) \pm O(\alpha(h)) &= O(\alpha(h)), \\ O(\alpha(h)) \pm o(\alpha(h)) &= O(\alpha(h)). \end{aligned}$$

Exercice 1. Justifier les identités précédentes.

Exercice 2. Montrer que si $R(h) = o(\alpha(h))$ quand $h \to 0$ et $\alpha(h) = O(\beta(h))$, alors $R(h) = o(\beta(h))$ quand $h \to 0$. En particulier, on peut écrire $o(C\,\alpha(h)) = o(\alpha(h))$ pour toute constante C.

1.2. Différentiabilité. Dans ce qui suit, E et F sont des evn. On rappelle que si $L : E \to F$ est linéaire continue, la notation "Lh" est synonyme de "$L(h)$".

DÉFINITION 1.2. *Soit $f : \Omega \to F$, où Ω est un ouvert de E et . On dit que f est **différentiable** en un point $p \in \Omega$ s'il existe une application linéaire continue $L : E \to F$ telle que*

$$f(p+h) = f(p) + Lh + o(\|h\|) \tag{1.1}$$

quand h tend vers 0.

Cette définition a bien un sens car $f(p+h)$ est effectivement défini dans un voisinage de 0 (on suppose que Ω est *ouvert*).

Remarque 1. Si f est différentiable en p, il existe une *unique* $L \in \mathcal{L}(E, F)$ vérifiant (1.1). On dit que L est la **différentielle de f au point** p, et on écrit $L = Df(p)$. Ainsi, quand $h \to 0$ on a

$$f(p+h) = f(p) + Df(p)h + o(\|h\|)\,.$$

DÉMONSTRATION. Soient L_1 et L_2 deux applications linéaires continues vérifiant (1.1). Alors

$$\begin{aligned} (L_1 - L_2)h &= (f(p+h) - f(p) + o(\|h\|)) - (f(p+h) - f(p) + o(\|h\|)) \\ &= o(\|h\|) \end{aligned}$$

quand $h \to 0$. Pour $\xi \in E \setminus \{0\}$ fixé, on a donc $(L_1 - L_2)(t\xi) = o(\|t\xi\|) = o(|t|)$ quand $t \to 0$, autrement dit

$$\lim_{t \to 0} \frac{\|(L_1 - L_2)(t\xi)\|}{|t|} = 0$$

Mais comme $L_1 - L_2$ est linéaire, on a $\frac{\|(L_1-L_2)(t\xi)\|}{|t|} = \frac{\|t\,(L_1-L_2)\xi\|}{|t|} = \|(L_1 - L_2)\xi\|$, expression qui ne dépend pas de t. Par conséquent $\|(L_1 - L_2)\xi\| = 0$, pour tout $\xi \in E \setminus \{0\}$. Ceci est évidemment encore vrai pour $\xi = 0$, et on a donc $L_1 = L_2$. □

Remarque 2. Si f est à valeurs *réelles*, on utilise aussi la notation $df(p)$ pour désigner la différentielle de f au point p.

Remarque 3. Si f est différentiable en p, alors f est *continue* au point p.

DÉMONSTRATION. On a $f(p+h) - f(p) + Df(p)h + o(\|h\|)$ quand $h \to 0$, et $Df(p)h = O(\|h\|)$ car $Df(p)$ est linéaire continue. Donc

$$f(p+h) - f(p) = o(\|h\|) + O(\|h\|) = O(\|h\|)$$

au voisinage de 0, et en particulier $f(p+h) - f(p)$ tend vers 0 quand $h \to 0$. □

Le lemme suivant montre que dans le cas des fonctions d'une variable, on n'a rien défini de nouveau : la différentiabilité est équivalente à la dérivabilité.

LEMME 1.3. *Soit $\varphi : I \to F$ où I est un intervalle ouvert de $\mathbb{R}$, et soit $t_0 \in I$. Alors φ est différentiable en t_0 si et seulement si elle est dérivable en t_0, et dans ce cas sa différentielle en t_0 est donnée par*

$$D\varphi(t_0)h = h \times \varphi'(t_0)\,.$$

En particulier, on a $\varphi'(t_0) = D\varphi(t_0)1$.

DÉMONSTRATION. Si φ est dérivable en t_0, alors on peut écrire $\varphi(t_0+h) = \varphi(t_0) + h\varphi'(t_0) + h\varepsilon(h)$, où $\varepsilon(h) \to 0$ quand $h \to 0$. Comme $h\varepsilon(h) = o(|h|)$ quand $h \to 0$ et que l'application $h \mapsto h\,\varphi'(t_0)$ est linéaire continue, cela montre que φ est différentiable en t_0 avec $D\varphi(t_0)h = h\varphi'(t_0)$.

Inversement, si φ est différentiable en t_0, alors $\varphi(t_0+h) = \varphi(t_0) + D\varphi(t_0)h + o(|h|)$ quand $h \to 0$. Comme $D\varphi(t_0)$ est linéaire, on a $D\varphi(t_0)h = D\varphi(t_0)(h\times 1) = h \times D\varphi(t_0)1$ pour tout $h \in \mathbb{R}$. En posant $\xi = D\varphi(t_0)1$, l'identité précédente s'écrit donc $\varphi(t_0+h) = \varphi(t_0) + h \times \xi + o(|h|)$. Ainsi, φ possède un développement limité à l'ordre 1 en t_0, et par conséquent φ est dérivable en t_0 avec $\varphi'(t_0) = \xi = D\varphi(t_0)1$. □

Remarque. De cette démonstration, il est bon de retenir le fait suivant : *toute application linéaire $L : \mathbb{R} \to F$ est de la forme $L(h) = h \times \xi$, et on a alors $\xi = L(1)$*. La correspondance $L \longleftrightarrow \xi = L(1)$ est un isomorphisme de F sur $\mathcal{L}(\mathbb{R}, F)$, et on a de plus $\|L\|_{\mathcal{L}(\mathbb{R},F)} = \|L(1)\|_F$ (*exercice*). Ainsi, $\mathcal{L}(\mathbb{R}, F)$ s'identifie isométriquement à F de façon "canonique".

DÉFINITION 1.4. *Soit Ω un ouvert de E, et soit $f : \Omega \to F$.*

(i) *On dit que f est **différentiable sur** Ω si elle est différentiable en tout point $p \in \Omega$. On a alors sous la main une* application $Df : \Omega \to \mathcal{L}(E, F)$*, qu'on appelle la **différentielle** de f.*

(ii) *On dit que f est **de classe $\mathcal{C}^1$ sur** Ω si f est différentiable sur Ω et si sa différentielle $Df : \Omega \to \mathcal{L}(E, F)$ est continue.*

Remarque. La différentielle de f est donc une application de Ω dans un "espace d'applications", à savoir l'espace$\mathcal{L}(E, F)$. Il est important d'être toujours bien conscient de ce fait : une différentielle est un objet compliqué.

Exemple 1. Soit I un intervalle ouvert de $\mathbb{R}$. Une application $\varphi : I \to F$ est de classe $\mathcal{C}^1$ au sens de la définition précédente si et seulement si elle est $\mathcal{C}^1$ au sens de la définition donnée au chapitre 1, i.e. φ est dérivable sur I et $\varphi' : I \to F$ est continue.

DÉMONSTRATION. On sait déjà que $\varphi : I \to F$ est différentiable sur I si et seulement si elle est dérivable sur I, et qu'on a alors $\varphi'(t) = D\varphi(t)1$ pour tout $t \in I$; autrement dit, la correspondance canonique entre $\mathcal{L}(\mathbb{R}, F)$ et F identifie $D\varphi(t)$ et $\varphi'(t)$, pour tout $t \in I$. Ainsi, lorsque φ est différentiable sur I, sa différentielle $D\varphi : I \to \mathcal{L}(\mathbb{R}, F)$ s'identifie à son application dérivée $\varphi' : I \to F$ et on a $\|D\varphi(t) - D\varphi(s)\|_{\mathcal{L}(\mathbb{R},F)} = \|\varphi'(t) - \varphi'(s)\|_F$ pour tous $s, t \in I$. En particulier, $D\varphi$ est continue si et seulement si φ' est continue. □

Exemple 2. Supposons qu'il existe une application linéaire continue $L : E \to F$ telle que $\forall u \in \Omega\ :\ f(u) = L(u)$. Alors f est de classe $\mathcal{C}^1$ et

$$\forall u \in \Omega\ :\ Df(u) = L\,.$$

La différentielle Df est donc *constante* sur Ω.

Démonstration. C'est évident puisque pour tout $p \in \Omega$ et h tel que $p+h \in \Omega$, on a $f(p+h) = L(p+h) = L(p) + L(h) = f(p) + Lh + 0$. □

Cas particulier. Soit Ω un ouvert de $\mathbb{R}^n$. Pour $j \in \{1, \ldots, n\}$, notons $x_j : \Omega \to \mathbb{R}$ la j-ième **fonction coordonnée** : si $u = (u_1, \ldots, u_n) \in \Omega$, alors $x_j(u) = u_j$. Alors les x_j sont de classe $\mathcal{C}^1$ sur Ω, et pour $u \in \Omega$ et $h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \in \mathbb{R}^n$, on a

$$dx_j(u)h = h_j\,.$$

Démonstration. Par définition, on a $x_j(u) = \pi_j(u)$ pour tout $u \in \Omega$, où $\pi_j : \mathbb{R}^n \to \mathbb{R}$ est la j-ième projection, qui est linéaire continue. Donc x_j est $\mathcal{C}^1$ avec $dx_j(u) = \pi_j$ pour tout $u \in \Omega$. □

Exemple 3. Soient E, F, G trois evn. Si $B : E \times F \to G$ est une application bilinéaire continue, alors B est de classe $\mathcal{C}^1$ et sa différentielle est donnée par la formule

$$DB(u,v)(h,k) = B(u,k) + B(h,v)\,. \tag{1.2}$$

En particulier, DB est une application *linéaire* continue de $E \times F$ dans $\mathcal{L}(E \times F, G)$.

Démonstration. On a

$$\begin{aligned} B((u,v)+(h,k)) &= B(u+h, v+k) \\ &= B(u,v) + B(u,k) + B(h,v) + B(h,k) \\ &= B(u,v) + B(u,k) + B(h,v) + O(\|(h,k)\|^2) \\ &= B(u,v) + B(u,k) + B(h,v) + o(\|(h,k)\|) \end{aligned}$$

quand $(h,k) \to 0$. Comme pour tout $(u,v) \in E \times F$ l'application $(h,k) \mapsto B(u,k) + B(h,v)$ est linéaire continue (car B est bilinéaire continue), cela montre que B est différentiable en tout point $(u,v) \in E \times F$ et que DB est donnée par (1.2).

Comme B est bilinéaire, (1.2) montre que l'application DB est linéaire de $E \times F$ dans $\mathcal{L}(E \times F, G)$. De plus, par continuité de B, il existe une constante C telle que $\|B(x,y)\| \leq C\,\|x\|\,\|y\|$ pour tout $(x,y) \in E \times F$. On en déduit

$$\begin{aligned} \|DB(u,v)(h,k)\| &\leq C\,(\|u\|\,\|k\| + \|h\|\,\|v\|) \\ &\leq 2C\,\|(u,v)\|\,\|(h,k)\|\,. \end{aligned}$$

Par conséquent, $\|DB(u,v)\| \leq 2C\,\|(u,v)\|$ pour tout $(u,v) \in E \times F$. Cela montre que l'application linéaire DB est continue, autrement dit que B est de classe $\mathcal{C}^1$. □

Cas particulier. L'application $P : M_n(\mathbb{R}) \times M_n(\mathbb{R}) \to M_n(\mathbb{R})$ définie par $P(A,B) = AB$ est de classe $\mathcal{C}^1$ sur $M_n(\mathbb{R})$, avec

$$DP(A,B)(H,K) = HB + AK\,.$$

Démonstration. C'est immédiat puisque P est bilinéaire continue. □

Exercice. Montrer que l'application $A \mapsto A^2$ est de classe $\mathcal{C}^1$ sur $M_n(\mathbb{R})$ et trouver sa différentielle.

1.3. Dérivées directionnelles. Dans ce qui suit, E et F sont toujours des evn.

DÉFINITION 1.5. *Soit $f : \Omega \to F$ où Ω est un ouvert de E, et soit $p \in \Omega$. Soit également $e \in E \setminus \{0\}$. On dit que f admet au point p une* ***dérivée dans la direction de*** *e si la fonction d'*une variable *$t \mapsto f(p + te)$ est dérivable en 0. On pose alors*

$$\begin{aligned} \partial_e f(p) &= \frac{d}{dt}\Big[f(p+te)\Big]_{t=0} \\ &= \lim_{t\to 0} \frac{f(p+te) - f(p)}{t}\,. \end{aligned}$$

Exercice. Montrer que si $\partial_e f(p)$ existe, alors $\partial_{\lambda e} f(p)$ existe pour tout $\lambda \in \mathbb{R}^*$, et $\partial_{\lambda e} f(p) = \lambda \partial_e f(p)$.

La proposition suivante est très simple, mais très importante.

PROPOSITION 1.6. *Si $f : \Omega \to F$ est différentiable en p, alors f admet au point p des dérivées dans toutes les directions, et on a*

$$\partial_e f(p) = Df(p)e$$

pour tout $e \in E \setminus \{0\}$.

DÉMONSTRATION. On a $f(p + h) = f(p) + Df(p)h + o(\|h\|)$ quand $h \to 0$. Par conséquent, si $e \in E \setminus \{0\}$ est fixé alors

$$\begin{aligned} f(p+te) &= f(p) + Df(p)(te) + o(\|te\|) \quad \text{quand } t \to 0 \\ &= f(p) + tDf(p)e + o(|t|)\,, \end{aligned}$$

car $Df(p)$ est linéaire et $\|te\| = |t|\,\|e\| = O(|t|)$. Ainsi, $f(p+te)$ possède un développement limité à l'ordre 1 en 0, et par conséquent $\frac{d}{dt}\big[f(p+te)\big]_{t=0}$ existe et vaut $Df(p)e$. $\square$

L'existence de dérivées dans toutes les directions ne suffit cependant pas à assurer la différentiabilité :

Exemple. Soit $f : \mathbb{R}^2 \to \mathbb{R}$ la fonction définie par $f(0,0) = 0$, $f(x,y) = 0$ si $y \neq x^2$ et $f(x,x^2) = 1$ pour tout $x \neq 0$. Alors

(i) f admet en $(0,0)$ des dérivées dans toutes les directions, et $\partial_e f(0,0) = 0$ pour tout $e \in \mathbb{R}^2 \setminus \{0\}$;

(ii) f n'est pas continue en $(0,0)$ (et donc pas différentiable).

DÉMONSTRATION. Il est clair que f n'est pas continue en $(0,0)$ puisque $f(x,x^2) = 1$ ne tend pas vers $0 = f(0,0)$ quand $x \to 0$.

Pour démontrer (i), fixons $e = (\alpha,\beta) \in \mathbb{R}^2 \setminus \{0\}$. On a $f((0,0)+te) = f(t\alpha, t\beta)$ pour tout $t \in \mathbb{R}$. L'équation $t\beta = (t\alpha)^2$ possède au plus 2 solutions (et en fait exactement 2 si α et β sont tous les deux non nuls), dont l'une est $t = 0$. Donc on a certainement $t\beta \neq (t\alpha)^2$ si $t \neq 0$ est assez proche de 0, et donc $f(t\alpha, t\beta) = 0$. Par conséquent, $f((0,0) + te) - f(0,0) = 0$ pour $t \neq 0$ assez proche de 0, donc $\partial_e f(0,0)$ existe et vaut 0. $\square$

2. Fonctions à valeurs dans $\mathbb{R}^m$; fonctions définies sur $\mathbb{R}^n$

2.1. Fonctions à valeurs dans $\mathbb{R}^m$. Le résultat suivant est très simple, mais important car il signifie que l'étude des fonctions à valeurs dans $\mathbb{R}^m$ se ramène à celle des fonctions *à valeurs réelles.*

Proposition 2.1. *Soit $f : \Omega \to \mathbb{R}^m$ (où Ω est un ouvert d'un evn E). On écrit*

$$f(u) = \begin{pmatrix} f_1(u) \\ \vdots \\ f_m(u) \end{pmatrix}.$$

Alors f est différentiable en un point $p \in \Omega$ si et seulement si les fonctions $f_1, \ldots, f_m$ le sont ; et dans ce cas $Df(p)$ est donnée par

$$Df(p)h = \begin{pmatrix} Df_1(p)h \\ \vdots \\ Df_m(p)h \end{pmatrix}. \tag{2.1}$$

Démonstration. Dans ce qui suit, on munit $\mathbb{R}^m$ de la norme $\| \cdot \|_\infty$. On aura besoin du fait suivant, qui sera également utilisé en d'autres occasions.

Fait. Soit $L : E \to \mathbb{R}^m$ une application linéaire, et écrivons

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

Alors L est continue si et seulement si ses "composantes" $L_1, \ldots, L_m$ le sont, et dans ce cas on a $\|L\| = \max(\|L_1\|, \ldots, \|L_m\|)$.

Preuve du fait. La première partie est claire puisque la convergence dans $\mathbb{R}^m$ est la convergence "coordonnée par coordonnée" (voir le chapitre 1, corollaire 2.5). Supposons les L_i (et donc L) continues. Si $h \in E$, alors $|L_i(h)| \leq \|L(h)\| \leq \|L\| \times \|h\|$ pour tout $i \in \{1, \ldots, m\}$ et $\|L(h)\|_\infty = \max(|L_1(h)|, \ldots, |L_m(h)|) \leq \max(\|L_1\|, \ldots, \|L_m\|) \times \|h\|$. On en déduit d'une part $\|L_i\| \leq \|L\|$ pour tout $i \in \{1, \ldots, m\}$, et d'autre part $\|L\| \leq \max(\|L_1\|, \ldots, \|L_m\|)$. Autrement dit : $\|L\| = \max(\|L_1\|, \ldots, \|L_m\|)$.

□

Supposons f différentiable en p. On peut alors écrire $f(p+h) = f(p) + Df(p)h + R(h)$, où $R(h) = o(\|h\|)$ quand $h \to 0$. En notant $\pi_1, \ldots, \pi_m$ les "formes linéaires coordonnées" sur $\mathbb{R}^m$, on en déduit $\pi_i(f(p+h)) = \pi_i(f(p)) + \pi_i(Df(p)h) + R_i(h)$ pour tout $i \in \{1, \ldots, m\}$; autrement dit $f_i(p+h) = f_i(p) + \pi_i(Df(p)h) + R_i(h)$, où $R_i(h) = \pi_i(R(h))$. Par continuité des formes linéaires π_i, on a $R_i(h) = O(\|R(h))\|)$, et donc $R_i(h) = o(\|h\|))$ quand $h \to 0$. Cela montre que chaque f_i est différentiable en p avec $Df_i(p)h = \pi_i(Df(p)h)$ pour tout $i \in \{1, \ldots, m\}$; d'où (2.1).

Inversement, supposons que chaque f_i soit différentiable en p. On peut alors écrire $f_i(p+h) = f_i(p) + Df_i(p)h + R_i(h)$, où $R_i(h) = o(\|h\|)$. Par conséquent,

$$f(p+h) = f(p) + \begin{pmatrix} Df_1(p)h \\ \vdots \\ Df_m(p)h \end{pmatrix} + R(h),$$

où $R(h) = \begin{pmatrix} R_1(h) \\ \vdots \\ R_m(h) \end{pmatrix}$. On a $\|R(h)\|_\infty = \max(|R_1(h)|, \ldots, |R_m(h)|) = o(\|h\|)$ car $R_i(h) = o(\|h\|)$ pour tout i. De plus, l'application linéaire L définie par $L(h) = \begin{pmatrix} Df_1(p)h \\ \vdots \\ Df_m(p)h \end{pmatrix}$ est continue d'après le fait. Ainsi, f est différentiable en p avec $Df(p) = L$.

□

COROLLAIRE 2.2. *Avec les notations précédentes, l'application f est de classe $\mathcal{C}^1$ si et seulement si ses "composantes" $f_1, \ldots, f_m$ le sont.*

DÉMONSTRATION. Munissons $\mathbb{R}^m$ de la norme $\|\cdot\|_\infty$. D'après le fait énoncé dans la preuve de la proposition, l'espace $\mathcal{L}(E, \mathbb{R}^m)$ s'identifie isométriquement à l'espace produit $\mathcal{L}(E, \mathbb{R}) \times \cdots \times \mathcal{L}(E, \mathbb{R})$. Une application $\Phi : \Omega \to \mathcal{L}(E, \mathbb{R}^m) = \mathcal{L}(E, \mathbb{R}) \times \cdots \times \mathcal{L}(E, \mathbb{R})$ est donc continue si et seulement si ses composantes $\Phi_1, \ldots, \Phi_m$ le sont. En particulier (en supposant f différentiable), $Df : \Omega \to \mathcal{L}(E, \mathbb{R}^m)$ est continue si et seulement si les Df_i le sont ; autrement dit, f est de classe $\mathcal{C}^1$ si et seulement si les f_i le sont. □

Exercice. Énoncer et démontrer un analogue de la proposition 2.1 pour une application f à valeurs dans un evn produit $E_1 \times \cdots \times E_m$.

2.2. Dérivées partielles. Dans ce qui suit, F est un evn.

DÉFINITION 2.3. *Soit Ω un ouvert de $\mathbb{R}^n$, et soit $f : \Omega \to F$. Soit également $j \in \{1, \ldots, n\}$. On dit que f admet en un point $p = (p_1, \ldots, p_n) \in \Omega$ une **dérivée partielle par rapport à la j-ième variable** si la fonction d'*une variable $x \mapsto f(p_1, \ldots, p_{j-1}, x, p_{j+1}, \ldots p_n)$ *est dérivable au point p_j. On pose alors*

$$\partial_j f(p) = \frac{d}{dx}\Big[f(p_1, \ldots, p_{j-1}, x, p_{j+1}, \ldots p_n)\Big]_{x=p_j}.$$

Autrement dit, on a

$$\partial_j f(p) = \lim_{t \to 0} \frac{f(p_1, \ldots, p_{j-1}, p_j + t, p_{j+1}, \ldots p_n) - f(p_1, \ldots, p_{j-1}, p_j, p_{j+1}, \ldots p_n)}{t}.$$

Remarque 1. En oubliant les notations peut-être désagréables, cette définition est très simple : on considère toutes les variables comme constantes sauf une, et on dérive par rapport à cette variable.

Remarque 2. Il ne faut pas perdre de vue que $\partial_j f(p)$ est un *élément de F*. Si $\partial_j f(p)$ existe en tout point $p \in \Omega$, on a donc sous la main une application $\partial_j f : \Omega \to F$, qu'on appelle bien entendu la **j-ième dérivée partielle de f**.

AUTRES NOTATIONS.

(i) Si les variables se notent $(x_1, \ldots, x_n)$, on écrit souvent $\frac{\partial f}{\partial x_j}$ au lieu de $\partial_j f$.

(i') Mais si les variables se notent $(u_1, \ldots, u_n)$, on écrit plutôt $\frac{\partial f}{\partial u_j}\cdot$

(ii) Si Ω est un ouvert de $\mathbb{R}^2$, on écrit plutôt (x, y) au lieu de (x_1, x_2), et donc $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ au lieu de $\frac{\partial f}{\partial x_1}$, $\frac{\partial f}{\partial x_2}$.

(iii) De même, si Ω est un ouvert de $\mathbb{R}^3$, les variables se notent en général (x, y, z) et on écrit $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial f}{\partial z}$.

En résumé : la seule notation "canonique" est $\partial_j f$. Les autres dépendent du nom qu'on donne aux variables et peuvent parfois prêter à confusion ; mais elles sont quand même très utilisées, car peut-être plus suggestives.

Exemple. Soit $f : \mathbb{R}^2 \to \mathbb{R}$ définie par $f(x,y) = e^{x^3y} + x^2 y \cos(x - y^3)$. Alors f admet des dérivées partielles en tout point, avec

$$\begin{aligned}\frac{\partial f}{\partial x}(x,y) &= 3x^2 y e^{x^3 y} + 2xy\cos(x-y^3) - x^2 y \sin(x-y^3)\,,\\ \frac{\partial f}{\partial y}(x,y) &= x^3 e^{x^3 y} + x^2\cos(x-y^3) + 3x^2y^3\sin(x-y^3)\,.\end{aligned}$$

Remarque. Pour alléger les notations, on écrit souvent $\frac{\partial f}{\partial x}$ et $\frac{\partial f}{\partial y}$ au lieu de $\frac{\partial f}{\partial x}(x,y)$ et $\frac{\partial f}{\partial y}(x,y)$. C'est commode, mais incorrect et dangereux : en toute rigueur, $\frac{\partial f}{\partial x}$ est une *fonction* dont la valeur en un point $p = (x,y)$ est $\frac{\partial f}{\partial x}(x,y)$. On va au devant de graves ennuis si on ne fait pas très attention à bien comprendre ce qu'on écrit.

Le lemme suivant montre que les dérivées partielles sont des cas particuliers de dérivées directionnelles.

LEMME 2.4. *Soit $f : \Omega \to F$ (où Ω est un ouvert de $\mathbb{R}^n$) et soit $p \in \Omega$. Pour tout $j \in \{1,\dots,n\}$, on a $\partial_j f(p) = \partial_{e_j} f(p)$, où e_j est le j-ième vecteur de la base canonique de $\mathbb{R}^n$.*

DÉMONSTRATION. C'est clair car $f(p + te_j) = f(p_1,\dots,p_{j-1},p_j+t,p_{j+1},\dots,p_n)$. $\square$

THÉORÈME 2.5. *Soit Ω un ouvert de $\mathbb{R}^n$, et soit $f : \Omega \to F$. Si f est différentiable en un point $p \in \Omega$, alors f admet au point p des dérivées partielles par rapport à chaque variable, et pour $h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \in \mathbb{R}^n$ on a*

$$Df(p)h = \sum_{j=1}^n h_j \partial_j f(p)\,.$$

En particulier, $\partial_j f(p) = Df(p)e_j$ pour tout $j \in \{1,\dots,n\}$.

DÉMONSTRATION. Supposons f différentiable en p. D'après la proposition 1.6, on sait que f admet au point p des dérivées dans toutes les directions, avec $\partial_e f(p) = Df(p)e$ pour tout $e \in \mathbb{R}^n \setminus \{0\}$. Donc f admet des dérivées partielles en p, avec

$$\partial_j f(p) = Df(p)e_j$$

pour tout $j \in \{1,\dots,n\}$. Pour $h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \in \mathbb{R}^n$, on a donc

$$\begin{aligned} Df(p)h &= Df(p)\left(\sum_{j=1}^n h_j e_j\right)\\ &= \sum_{j=1}^n h_j\, Df(p)e_j \qquad \text{par linéarité}\\ &= \sum_{j=1}^n h_j \partial_j f(p)\,.\end{aligned}$$

$\square$

On a vu plus haut que l'existence de dérivées partielles en un point ne suffit pas à assurer la différentiabilité (voir l'exemple donné après la proposition 1.6). Avec des hypothèses plus fortes, on obtient cependant une réciproque au théorème précédent :

THÉORÈME 2.6. *Soit Ω un ouvert de $\mathbb{R}^n$, et soit $f : \Omega \to F$. Soit également $p \in \Omega$. On suppose que f admet des dérivées partielles* en tout point de Ω *et que les fonctions $\partial_1 f, \dots, \partial_n f$ sont* continues *au point p. Alors f est différentiable en p.*

DÉMONSTRATION. D'après le théorème précédent, le seul candidat possible pour $Df(p)$ est l'application linéaire (continue) $L : \mathbb{R}^n \to F$ définie par

$$L(h) = \sum_{j=1}^{n} h_j \partial_j f(p)\,.$$

Il s'agit donc de montrer qu'on a $f(p+h) - f(p) - L(h) = o(\|h\|)$ quand $h \to 0$. Dans la suite, on munit $\mathbb{R}^n$ de la norme $\|\cdot\|_\infty$ et on choisit $\delta > 0$ tel que

$$\|h\|_\infty \le \delta \implies p + h \in \Omega\,.$$

Soit $h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \in \mathbb{R}^n$ vérifiant $\|h\|_\infty \le \delta$. En écrivant

$$\begin{aligned} f(p+h) - f(p) &= \big[f(p_1+h_1, p_2+h_2, \dots, p_n+h_n) - f(p_1, p_2+h_2, \dots, p_n+h_n)\big] \\ &+ \big[f(p_1, p_2+h_2, \dots, p_n+h_n) - f(p_1, p_2, p_3+h_3, \dots, p_n+h_n)\big] \\ &+ \dots \\ &+ \big[f(p_1, \dots, p_{n-1}, p_n+h_n) - f(p_1, \dots, p_n)\big]\,, \end{aligned}$$

on obtient

$$f(p+h) - f(p) - L(h) = \sum_{j=1}^{n} \Delta_j(h)\,,$$

où les $\Delta_j(h)$ sont donnés par les formules suivantes :

$$\begin{cases} \Delta_1(h) = f(p_1+h_1, p_2+h_2, \dots, p_n+h_n) - f(p_1, p_2+h_2, \dots, p_n+h_n) - h_1\partial_1 f(p) \\ \Delta_2(h) = f(p_1, p_2+h_2, \dots, p_n+h_n) - f(p_1, p_2, p_3+h_3, \dots, p_n+h_n) - h_2\partial_2 f(p) \\ \quad\vdots \\ \Delta_n(h) = f(p_1, \dots, p_{n-1}, p_n+h_n) - f(p_1, \dots, p_n) - h_n\partial_n f(p) \end{cases}$$

En regardant Δ_1 de plus près, on constate qu'on a

$$\Delta_1(h) = \varphi(1) - \varphi(0)\,,$$

où $\varphi : [0,1] \to F$ est définie par

$$\varphi(t) = f(p_1 + th_1, p_2+h_2, \dots, p_n+h_n) - th_1\partial_1 f(p)\,.$$

Comme $u(t) := (p_1 + th_1, p_2+h_2, \dots, p_n+h_n) \in \overline{B}(p,\delta) \subset \Omega$ pour tout $t \in [0,1]$ et comme $\partial_1 f(u)$ existe en tout point de Ω, la fonction φ est dérivable sur $[0,1]$ avec

$$\varphi'(t) = h_1\partial_1 f(p_1 + th_1, p_2+h_2, \dots, p_n+h_n) - h_1\partial_1 f(p)\,.$$

D'après l'inégalité des accroissements finis, on sait qu'on peut trouver $c \in]0,1[$ tel que $\|\varphi(1) - \varphi(0)\| \le \|\varphi'(c)\|$, autrement dit

$$\|\Delta_1(h)\| \le |h_1|\, \|\partial_1 f(p_1 + ch_1, p_2+h_2, \dots, p_n+h_n) - \partial_1 f(p)\|\,.$$

D'autre part, le point $q^1(h) = (p_1 + ch_1, p_2 + h_2, \dots, p_n + h_n)$ vérifie $\|q^1(h) - p\|_\infty = \max(|ch_1|, |h_2|, \dots, |h_n|) \leq \|h\|_\infty$. Ainsi, on a trouvé un point $q^1(h) \in \Omega$ tel que

$$\|q^1(h) - p\|_\infty \leq \|h\|_\infty \quad \text{et} \quad \|\Delta_1(h)\| \leq |h_1| \, \|\partial_1 f(q^1(h)) - \partial_1 f(p)\| \, .$$

De même, on peut trouver des points $q^2(h), \dots, q^n(h)$ tels que $\|q^j(h) - p\|_\infty \leq \|h\|_\infty$ pour tout j, et

$$\|\Delta_j(h)\| \leq |h_j| \, \|\partial_1 f(q^j(h)) - \partial_1 f(p)\| \, .$$

On en déduit

$$\begin{aligned} \|f(p+h) - f(p) - L(h)\| &\leq \sum_{j=1}^n |h_j| \, \|\partial_1 f(q^j(h)) - \partial_1 f(p)\| \\ &\leq \|h\|_\infty \times \sum_{j=1}^n \|\partial_1 f(q^j(h)) - \partial_1 f(p)\| \\ &= \|h\|_\infty \times \varepsilon(h) \, . \end{aligned}$$

Comme $\|q^j(h) - p\|_\infty \leq \|h\|_\infty$, chaque $q^j(h)$ tend vers p quand $h \to 0$, et donc $\varepsilon(h)$ tend vers 0 puisque les $\partial_j f$ sont *continues* au point p. On a donc bien montré que $f(p+h) - f(p) - L(h) = o(\|h\|)$ quand $h \to 0$, ce qui achève la démonstration. $\square$

Corollaire 2.7. *Soit Ω un ouvert de $\mathbb{R}^n$. Pour une application $f : \Omega \to F$, les propriétés suivantes sont équivalentes :*

(i) *f est de classe $\mathcal{C}^1$ sur Ω ;*

(ii) *f admet des dérivées partielles en tout point de Ω, et les fonctions $\partial_1 f, \dots, \partial_n f$ sont continues sur Ω.*

Démonstration. Si (i) est vérifiée, alors f admet des dérivées partielles en tout point, et on a $\partial_j f(u) = Df(u)e_j$ pour $j \in \{1, \dots, n\}$ et pour tout $u \in \Omega$. Comme l'application (linéaire) $L \mapsto Le_j$ est continue sur $\mathcal{L}(\mathbb{R}^n, F)$ et comme Df est continue, on en déduit par composition que chaque $\partial_j f$ est continue sur Ω.

Inversement, supposons (ii) vérifiée. Alors f est différentiable sur Ω d'après le théorème, et il reste à voir que $Df : \Omega \to \mathcal{L}(\mathbb{R}^n, F)$ est continue. Munissons $\mathbb{R}^n$ de la norme $\| \cdot \|_\infty$. La preuve du fait suivant est laissée en exercice.

Fait. Si $L \in \mathcal{L}(\mathbb{R}^n, F)$, alors $\|L\| = \sum_1^n \|L(e_j)\|$.

D'après le fait, si $u, v \in \Omega$ alors

$$\begin{aligned} \|Df(v) - Df(u)\| &= \sum_{j=1}^n \|Df(v)e_j - Df(u)e_j\| \\ &= \sum_{j=1}^n \|\partial_j f(v) - \partial_j f(u)\| \, . \end{aligned}$$

Comme les $\partial_j f$ sont supposées continues, on en déduit que Df est continue, autrement dit que f est de classe $\mathcal{C}^1$.

$\square$

Corollaire 2.8. *Soit Ω un ouvert de $\mathbb{R}^n$, et soit $f : \Omega \to \mathbb{R}^m$. Écrivons $f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$. Alors f est de classe $\mathcal{C}^1$ si et seulement si $\partial_j f_i$ existe et est continue sur Ω, pour tout $i \in \{1, \dots, m\}$ et pour tout $j \in \{1, \dots, n\}$.*

Remarque. L'intérêt “pratique” des deux corollaires précédents est évident : pour montrer qu'une fonction définie sur un ouvert de $\mathbb{R}^n$ est de classe $\mathcal{C}^1$, il suffit de calculer des dérivées partielles et de vérifier qu'elles sont continues. On peut même oublier la définition de la différentiabilité !

Exemple. La formule $f(x,y) = \begin{pmatrix} e^x \log(x-y) \\ (x+y)\sin(xy) \\ x^5y^3 + \frac{\cos(x^3+y^2)}{x^3-1} \end{pmatrix}$ définit une application de classe $\mathcal{C}^1$ sur $\Omega = \{(x;y) \in \mathbb{R}^2;\ x \neq 1 \text{ et } x > y\}$, à valeurs dans $\mathbb{R}^3$.

2.3. Matrice jacobienne ; gradient.

DÉFINITION 2.9. *Soit Ω un ouvert de $\mathbb{R}^n$, et soit $f : \Omega \to \mathbb{R}$ possèdant des dérivées partielles en un point $p \in \Omega$. Le* ***gradient*** *de f au point p, noté $\nabla f(p)$, est le vecteur de $\mathbb{R}^n$ défini par*

$$\nabla f(p) = \begin{pmatrix} \partial_1 f(p) \\ \vdots \\ \partial_n f(p) \end{pmatrix}.$$

Il est important de se souvenir que le gradient n'est traditionnellement défini que pour une fonction *à valeurs réelles.*

PROPOSITION 2.10. *Soit Ω un ouvert de $\mathbb{R}^n$, et soit $f : \Omega \to \mathbb{R}$. Si f est différentiable en un point $p \in \Omega$, alors*

$$\forall h \in \mathbb{R}^n \ :\ Df(p)h = \langle \nabla f(p), h \rangle\,,$$

où $\langle\ ,\ \rangle$ est le produit scalaire usuel sur $\mathbb{R}^n$.

DÉMONSTRATION. C'est évident : si $h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \in \mathbb{R}^n$, alors

$$Df(p)h = \sum_{j=1}^n h_j \partial_j f(p) = \langle \nabla f(p), h \rangle\,.$$

□

DÉFINITION 2.11. *Soit Ω un ouvert de $\mathbb{R}^n$, et soit $f : \Omega \to \mathbb{R}^m$. Écrivons comme d'habitude $f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$, et supposons que les fonctions f_i possèdent des dérivées partielles en un point $p \in \Omega$. La* ***matrice jacobienne*** *de f au point p est la matrice*

$$\mathrm{Jac}_f(p) = \left(\frac{\partial f_i}{\partial x_j}(p)\right)_{i,j=1}^n$$

REMARQUE. La matrice jacobienne d'une fonction de $\mathbb{R}^n$ dans $\mathbb{R}^m$ a donc m lignes et n colonnes. Pour s'en souvenir, penser aux applications linéaires : la matrice d'une application linéaire $L = \mathbb{R}^n \to \mathbb{R}^m$ a bien m lignes et n colonnes. Pour ne pas se tromper dans l'écriture de Jac_f, il faut juste ne pas oublier d'*écrire $f(x_1,\dots,x_n)$ comme un vecteur colonne.*

Exercice. Écrire la matrice jacobienne de L'application $f : \mathbb{R}^2 \to \mathbb{R}^3$ définie par $f(x,y) = \big(x^3y^4, (x+y^2)\sin(xy), \arctan(x-y^3)\big)$ en tout point $(x,y) \in \mathbb{R}^2$.

Proposition 2.12. *Soit Ω un ouvert de $\mathbb{R}^n$. Si $f : \Omega \to \mathbb{R}^m$ est différentiable en un point $p \in \Omega$, alors $\mathrm{Jac}_f(p)$ est la matrice de l'application linéaire $Df(p)$ relativement aux bases canoniques de $\mathbb{R}^n$ et de $\mathbb{R}^m$.*

Démonstration. On a vu que la différentielle $Df(p)$ est donnée par

$$Df(p)h = \begin{pmatrix} Df_1(p)h \\ \vdots \\ Df_m(p)h \end{pmatrix}.$$

En écrivant $h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$, on voit donc que pour $i \in \{1, \dots, m\}$ la i-ième coordonnée du vecteur $Df(p)h \in \mathbb{R}^m$ est égale à

$$\begin{aligned} \big(Df(p)h\big)_i &= \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(p)\, h_j \\ &= \big(\mathrm{Jac}_f(p)h\big)_i . \end{aligned}$$

□

Exercice. Soit $f : \Omega \to \mathbb{R}$ une fonction à valeurs réelles différentiable en un point p. Quelle relation y a-t-il entre le gradient de f au point p et sa matrice jacobienne au point p ?

2.4. Formes différentielles et "notation physicienne". Dans ce qui suit, Ω est un ouvert de $\mathbb{R}^n$.

Définition 2.13. *Une **forme différentielle** (de degré 1) sur Ω est une application $\alpha : \Omega \to \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ de Ω dans l'espace vectoriel $\mathcal{L}(\mathbb{R}^n, \mathbb{R})$.*

Exemple. Si $f : \Omega \to \mathbb{R}$ est différentiable sur Ω, alors sa différentielle *df* est une forme différentielle sur Ω.

Comme $\mathcal{L}(\mathbb{R}^n, \mathbb{R})$ est un espace vectoriel, on dispose de deux opérations naturelles sur les formes différentielles.

- *Somme* : si α et β sont des formes différentielles sur Ω, on définit une forme différentielle $\alpha + \beta$ en posant
$$(\alpha + \beta)(u) = \alpha(u) + \beta(u)\,.$$
- *Multiplication par une fonction* : si $\alpha : \Omega \to \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ est une forme différentielle et si $\lambda : \Omega \to \mathbb{R}$ est une fonction sur Ω (à valeurs réelles), on définit une forme différentielle $\lambda\,\alpha$ en posant
$$(\lambda\,\alpha)(u) = \lambda(u)\alpha(u)\,.$$

Proposition 2.14. *Soient $x_1, \dots, x_n$ les fonctions coordonnées sur Ω. Toute forme différentielle $\alpha : \Omega \to \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ s'écrit de manière unique sous la forme*

$$\alpha = \sum_{j=1}^{n} \lambda_j \, dx_j$$

où les λ_j sont des fonctions sur Ω. On a $\lambda_j(u) = \alpha(u)e_j$ pour tout $u \in \Omega$ et $j \in \{1, \dots, n\}$.

DÉMONSTRATION. On sait qu'on a $dx_j(u) = \pi_j$ pour tout $u \in \Omega$ et $j \in \{1, \dots, n\}$, où $\pi_j : \mathbb{R}^n \to \mathbb{R}$ est la j-ième forme linéaire coordonnée. D'autre part, on sait également que $(\pi_1, \dots, \pi_n)$ est une base de $\mathcal{L}(\mathbb{R}^n, \mathbb{R}) = (\mathbb{R}^n)^*$. Le résultat s'en déduit immédiatement. Enfin, en évaluant $\alpha(u) = \sum_1^n \lambda_k(u) dx_j(u) = \sum_1^n \lambda_k(u)\pi_k$ sur le vecteur e_j et en se souvenant que $\pi_k(e_j)$ vaut 0 pour $k \neq j$ et 1 pour $k = j$, on trouve $\alpha(u)e_j = \lambda_j(u)$. $\square$

Cas particulier. Si $f : \Omega \to \mathbb{R}$ est différentiable, on a $df(u)e_j = \partial_j f(u)$ pour tout $u \in \Omega$ et $j \in \{1, \dots, n\}$. On en déduit l'écriture de la forme différentielle df dans la "base" $(dx_1, \dots, dx_n)$:

$$df = \sum_{j=1}^n \frac{\partial f}{\partial x_j} \, dx_j \, . \tag{2.2}$$

C'est la mystérieuse formule utilisée systématiquement en physique. On verra plus loin que cette formule est particulièrement bien adaptée aux changements de variables.

Remarque 1. L'écriture (2.2) dépend du *nom* qu'on a donné aux variables, i.e. aux fonctions coordonnées. Si les variables s'appelaient $u_1, \dots, u_n$, on écrirait évidemment $df = \sum_1^n \frac{\partial f}{\partial u_j} \, du_j$.

Remarque 2. Si $n = 1$, i.e. Ω est un ouvert de $\mathbb{R}$, on a une seule variable x, donc $\partial_1 f = f'$ et $df = f' \, dx$. Même si on ne peut pas "diviser par dx", cela explique la notation

$$f' = \frac{df}{dx} \, .$$

3. Fonctions composées

Le théorème suivant généralise la formule de dérivation des fonctions composées

$$(g \circ f)'(t) = g'(f(t)) \times f'(t) \, .$$

Rappellons qu'on note $L_1 L_2$ plutôt que $L_1 \circ L_2$ la composée (le "produit") de deux applications linéaires L_1 et L_2.

THÉORÈME 3.1. (théorème des fonctions composées)
Soient, E, F, G trois evn, $f : \Omega \to F$ où Ω un ouvert de E, et $g : \Omega' \to G$ où Ω' un ouvert de F. On suppose qu'on a $f(\Omega) \subset \Omega'$ (de sorte que la composée $g \circ f : \Omega \to G$ est bien définie). Si f est différentiable en un point $p \in \Omega$ et si g est différentiable en $f(p)$, alors $g \circ f$ est différentiable en p avec

$$D(g \circ f)(p) = Dg(f(p)) Df(p) \, .$$

Remarque. Il faut bien comprendre cette formule. On a $Df(p) \in \mathcal{L}(E, F)$ et $Dg(f(p)) \in \mathcal{L}(F, G)$, donc la composée $Dg(f(p))Df(p) = Dg(f(p)) \circ Df(p)$ est bien définie. La formule signifie que si $h \in E$, alors $D(g \circ f)(p)h = Dg(f(p))(Df(p)h)$.

DÉMONSTRATION. Il s'agit simplement d'une "composition de développements limités". On a

$$f(p + h) = f(p) + Df(p)h + R_1(h)$$

où $R_1(h) = o(\|h\|)$ quand $h \to 0$, et

$$g(f(p) + k) = g(f(p)) + Dg(f(p))k + R_2(k)$$

où $R_2(k) = o(\|k\|)$ quand $k \to 0$. En prenant $k = Df(p)h + R_1(h)$, on en déduit

$$\begin{aligned} g \circ f(p+h) &= g(f(p) + Df(p)h + R_1(h)) \\ &= g(f(p)) + Dg(f(p))\big(Df(p)h + R_1(h)\big) + R_2\big(Df(p)h + R_1(h)\big) \\ &= g \circ f(p) + Dg(f(p)Df(p)h + R_3(h)\,, \end{aligned}$$

où $R_3(h) = Dg(f(p))R_1(h) + R_2\big(Df(p)h + R_1(h)\big)$.

Comme $Dg(f(p))$ est linéaire continue, on a $Dg(f(p)R_1(h) = 0(\|R_1(h)\|) = o(\|h\|)$ quand $h \to 0$; et comme $k = Df(p)h + R_1(h)$ tend vers 0 quand $h \to 0$, on a $R_2(Df(p)h + R_1(h)) = o(\|Df(p)h + R_1(h)\|) = o(\|h\|)$ quand $h \to 0$ (car $Df(p)h + R_1(h) = O(\|h\|) + o(\|h\|) = O(\|h\|)$ au voisinage de 0). Ainsi, $R_3(h) = o(\|h\|)$ quand $h \to 0$, ce qui termine la démonstration. □

Corollaire 3.2. *La composée de deux applications de classe $\mathcal{C}^1$ est de classe $\mathcal{C}^1$ sur son domaine de définition.*

Démonstration. D'après le théorème, si f et g sont différentiables alors $\Phi = g \circ f$ est différentiable sur Ω avec $D\Phi(u) = Dg(f(u))Dg(u)$ pour tout $u \in \Omega$. On peut donc écrire

$$D\Phi(u) = B\Big(Df(u), Dg(f(u))\Big)\,,$$

où $B : \mathcal{L}(E, F) \times \mathcal{L}(F, G)$ est l'application bilinéaire définie par $B(S, T) = TS$. Comme B est *continue* (voir le chapitre 1), on en déduit par composition que si Df et Dg sont continues, alors $D(g \circ f)$ est continue. □

Conséquence pratique. Comme pour la continuité, il découle de ce résultat qu'une fonction donnée par une "formule explicite", construite à partir de fonctions de classe $\mathcal{C}^1$, est de classe $\mathcal{C}^1$ sur tout ouvert contenu dans son domaine de définition.

Exercice. Soit $\alpha > 0$. Montrer que la formule $f(x, y) = (x^2 + y^2)^\alpha$ définit une fonction de classe $\mathcal{C}^1$ sur $\Omega = \mathbb{R}^2 \setminus \{(0, 0)\}$. Pour $\alpha = 1/2$, cette fonction est-elle de classe $\mathcal{C}^1$ sur $\mathbb{R}^2$?

Corollaire 3.3. *Sous les hypothèses du théorème, on prend $E = \mathbb{R}^n$, $F = \mathbb{R}^m$ et $G = \mathbb{R}^l$. Alors la matrice jacobienne de $g \circ f$ au point p est le produit des deux matrices jacobiennes $\mathrm{Jac}_g(f(p))$ et $\mathrm{Jac}_f(p)$:*

$$\mathrm{Jac}_{g\circ f}(p) = \mathrm{Jac}_g(f(p))\,\mathrm{Jac}_f(p)\,.$$

Démonstration. C'est évident puisque les matrices jacobiennes sont les matrices des différentielles et que la composition des applications linéaires correspond au produit des matrices. □

Corollaire 3.4. *Soit $f : \Omega \to \mathbb{R}$, où Ω est un ouvert de $\mathbb{R}^m$. On suppose que f s'écrit sous la forme*

$$f(u_1, \dots, u_m) = F(x_1(u_1, \dots, u_m), \dots, x_n(u_1, \dots, u_m))\,,$$

où les fonctions F et $x_1, \dots, x_n$ sont différentiables. Pour $j \in \{1, \dots, n\}$, notons "$\frac{\partial F}{\partial x_j}$" la fonction $u \mapsto \partial_j F(x_1(u), \dots, x_n(u))$. Alors :

(1) *la différentielle de f est donnée par la formule*

$$df = \sum_{j=1}^{n} \text{“}\frac{\partial F}{\partial x_j}\text{”}\, dx_j\,;$$

(2) *pour* $i \in \{1, \ldots, m\}$, *on a*

$$\frac{\partial f}{\partial u_i} = \sum_{j=1}^{n} \text{“}\frac{\partial F}{\partial x_j}\text{”} \, \frac{\partial x_j}{\partial u_i} \cdot$$

DÉMONSTRATION. Posons $x(u) = (x_1(u), \ldots, x_n(u))$. Alors $x : \Omega \to \mathbb{R}^n$ est différentiable et $f = F \circ x$. Pour $u \in \Omega$ et $h \in \mathbb{R}^m$, on a donc

$$\begin{aligned} df(u)h &= dF(x(u))(dx(u)h) \\ &= \sum_{j=1}^{n} (dx(u)h)_j \, \partial_j F(x(u)) \\ &= \sum_{j=1}^{n} \partial_j F(x(u)) \times dx_j(u)h \\ &= \left(\sum_{j=1}^{n} \text{“}\frac{\partial F}{\partial x_j}\text{”} \, dx_j \right)(u)\, h\,. \end{aligned}$$

Cela prouve (1).

La formule (2) découle immédiatement de (1) : en écrivant $dx_j = \sum_{i=1}^{m} \frac{\partial x_j}{\partial u_i} du_i$ et en reportant dans (1), on obtient

$$\begin{aligned} df &= \sum_{j=1}^{n} \text{“}\frac{\partial F}{\partial x_j}\text{”} \times \left(\sum_{i=1}^{m} \frac{\partial x_j}{\partial u_i} du_i \right) \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \text{“}\frac{\partial F}{\partial x_j}\text{”} \, \frac{\partial x_j}{\partial u_i} \right) du_i \end{aligned}$$

Comme par ailleurs $df = \sum_1^m \frac{\partial f}{\partial u_i} du_i$, on en déduit la formule souhaitée par *unicité* de la décomposition de la forme différentielle df dans la “base” $(du_1, \ldots, du_m)$:

$$\frac{\partial f}{\partial u_i} = \sum_{j=1}^{n} \text{“}\frac{\partial F}{\partial x_j}\text{”} \, \frac{\partial x_j}{\partial u_i} \cdot$$

□

PREUVE DIRECTE DE (2). Pour $u \in \Omega$ et $i \in \{1, \ldots, m\}$ donnés, $\frac{\partial f}{\partial u_i}(u)$ est le i-ème coefficient de la matrice jacobienne $\mathrm{Jac}_f(u)$ (qui est une matrice ligne $1 \times m$ puisque f est à valeurs réelles). Comme $f = F \circ x$, où $x(u) = (x_1(u), \ldots, x_n(u))$, on sait que $\mathrm{Jac}_f(u) = \mathrm{Jac}_F(x(u))\mathrm{Jac}_x(u)$, autrement dit

$$\mathrm{Jac}_f(u) = \begin{pmatrix} \frac{\partial F}{\partial x_1}(x(u)) & \ldots & \frac{\partial F}{\partial x_n}(x(u)) \end{pmatrix} \times \begin{pmatrix} \frac{\partial x_1}{\partial u_1}(u) & \ldots & \frac{\partial x_1}{\partial u_m}(u) \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial u_1}(u) & \ldots & \frac{\partial x_n}{\partial u_m}(u) \end{pmatrix}.$$

En écrivant le i-ème coefficient de ce produit de matrices, on obtient (2). □

Remarque. Évidemment, on va s'empresser d'oublier les guillemets autour de “$\frac{\partial F}{\partial x_j}$” pour écrire simplement $\frac{\partial F}{\partial x_j}\cdot$ Mais il ne faut pas perdre de vue que cette notation est une abbréviation potentiellement source d'erreurs. On peut cependant en rajouter dans l'abus de notation, et écrire f au lieu de F dans (1) : on retombe alors sur la formule magique $df = \sum_{j=1}^n \frac{\partial f}{\partial x_j}\, dx_j$, mais le "$f$" du membre de droite n'a pas le même sens que celui du membre de gauche. Le f de droite est la fonction f considérée comme fonction des nouvelles variables $x_1, \ldots, x_n$, qui sont elles-même des fonctions et à ce titre admettent des différentielles $dx_1, \ldots, dx_n$. En résumé : la formule magique $df = \sum_{j=1}^n \frac{\partial f}{\partial x_j}\, dx_j$ *est valable en toute circonstance*, à condition de l'interpréter correctement.

Exercice. Soit $F : \Omega \to \mathbb{R}$ différentiable, où $\Omega = \mathbb{R}^2 \setminus \{(x, 0);\ x \geq 0\}$, et soit $f :]0, \infty[\times]0, 2\pi[\to \mathbb{R}$ la fonction définie par $f(r, \theta) = F(r\cos\theta, r\sin\theta)$. Exprimer les dérivées partielles de f en fonction de celles de F, puis celles de F en fonction de celles de f.

Corollaire 3.5. *Soit I un intervalle ouvert de $\mathbb{R}$, et soit $f : I \to \mathbb{R}$ de la forme $f(t) = F(x_1(t), \ldots, x_n(t))$, où F est différentiable et les x_j sont dérivables. Alors la dérivée de f est donnée par la formule*

$$f'(t) = \sum_{j=1}^{n} \frac{\partial F}{\partial x_j}(x_1(t), \ldots, x_n(t))\, x_j'(t)\,.$$

Démonstration. C'est un cas particulier du corollaire précédent (avec $\Omega = I$ et $m = 1$). $\square$

Exercice. soit $F : \mathbb{R}^3 \to \mathbb{R}$ différentiable. Donner une formule pour la dérivée de la fonction $t \mapsto F(t^2 + t^5, e^t, t^3 \cos t)$.

Corollaire 3.6. *Soient $\gamma = I \to E$ dérivable avec $\gamma(I) \subset \Omega$ et soit $f : \Omega \to F$ différentiable, où I est un intervalle ouvert de $\mathbb{R}$. Alors la fonction $\varphi : I \to F$ définie par $\varphi(t) = f(\gamma(t))$ est dérivable, avec $\varphi'(t) = Df(\gamma(t))\gamma'(t)$.*

Démonstration. On a $\varphi = f \circ \gamma$, donc φ est différentiable (i.e. dérivable) avec $D\varphi(t)h = Df(\gamma(t))(D\gamma(t)h)$ pour $t \in I$ et $h \in \mathbb{R}$. Autrement dit :

$$h \times \varphi'(t) = Df(\gamma(t))(h \times \gamma'(t)) = h \times Df(\gamma(t))\gamma'(t)\,,$$

d'où le résultat en prenant $h = 1$.

$\square$

Remarque. Le résultat reste vrai même pour un intervalle I quelconque. On peut le voir soit en recopiant la preuve du théorème des fonctions composées dans ce cas particulier, soit en prolongant φ en une fonction dérivable sur un intervalle ouvert contenant I (exercice).

Corollaire 3.7. *Soit Ω un ouvert d'un evn E, et soient $f, g : \Omega \to \mathbb{R}$. Si f et g sont différentiables, alors $f + g$ et fg le sont avec $d(f + g) = df + dg$ et $d(fg) = gdf + fdg$. Si de plus g ne s'annule pas, alors f/g est différentiable avec $d(f/g) = \frac{gdf - fdg}{g^2}$.*

Démonstration. Pour la somme, il n'est pas difficile de démontrer le résultat en utilisant uniquement la définition de la différentiabilité. On peut aussi procéder d'une manière beaucoup plus compliquée mais qui illustre bien le théorème des fonctions composées. Par définition, on a $f + g = L \circ \Phi$ où $\Phi : \Omega \to \mathbb{R} \times \mathbb{R}$ sont définies par $\Phi(u) = (f(u), g(u))$ et $L(x, y) = x + y$. L'application Φ est différentiable avec $D\Phi(u)h = (df(u)h, dg(u)h)$, et l'aplication L est linéaire continue, donc différentiable avec $DL(x,y) = L$ pour tout $(x,y) \in \mathbb{R}^2$. Par composition, $f+g$ est donc différentiable avec

$$\begin{aligned} d(f+g)(u)h &= DL(\Phi(u))(D\Phi(u)h) \\ &= L(df(u)h, dg(u)h) \\ &= df(u)h + dg(u)h\,, \end{aligned}$$

d'où le résultat.

Pour le produit, on écrit $fg = B \circ \Phi$, où $\Phi(u) = (f(u), g(u))$ et $B : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ est définie par $B(x,y) = xy$. Comme B est bilinéaire continue, on en déduit que fg est différentiable avec

$$\begin{aligned} d(fg)(u)h &= DB(\Phi(u))(D\Phi(u)h) \\ &= DB(f(u), g(u))(df(u)h, dg(u)h) \\ &= B(f(u), dg(u)h) + B(df(u)h, g(u)) \\ &= f(u)dg(u)h + g(u)df(u)h\,, \end{aligned}$$

ce qui est la formule souhaitée.

Pour le quotient, on se ramène au produit. En écrivant $h = f/g$, on a $hg = f$, donc $d(hg) = df$ autrement dit $gdh + hdg = df$. Par conséquent :

$$d(f/g) = dh = (df - hdg)/g = \frac{df - (f/g)dg}{g} = \frac{gdf - fdg}{g^2}\,.$$

Il faut quand même justifier que f/g est bien différentiable, ce qui découle du théorème des fonctions composées et de la différentiabilité de la fonction $(x,y) \mapsto x/y$ sur $\mathbb{R}\times\mathbb{R}^*$ (formule explicite ...). □

4. AF et TFA

4.1. L'inégalité des accroissements finis. Rappelons que si u et v sont deux points d'un espace vectoriel (réel) E, on note $[u,v]$ le **segment d'extrémités** u **et** v. C'est le sous-ensemble de E défini analytiquement par

$$[u,v] = \big\{(1-t)u + tv;\ t \in [0,1]\big\}\,.$$

On a bien sûr $[u,v] = [v,u]$. D'autre part, cette définition a un sens pour $E = \mathbb{R}$ et on obtient bien l'intervalle fermé borné d'extrémités u et v.

Théorème 4.1. (inégalités des accroissements finis)
Soient E et F deux evn, et soit $f : \Omega \to F$ différentiable, où Ω est un ouvert de E. Si $u, v \in \Omega$ et si le segment $[u,v]$ est contenu dans Ω, alors il existe un $\xi \in [u,v]$ tel que

$$\|f(v) - f(u)\| \le \|Df(\xi)(v-u)\|\,.$$

En particulier, on peut trouver $\xi \in [u,v]$ tel que

$$\|f(v) - f(u)\| \le \|Df(\xi)\| \times \|v-u\|\,.$$

Démonstration. Soit $\varphi : [0,1] \to F$ la fonction définie par

$$\varphi(t) = f(\gamma(t))\,,$$

où $\gamma(t) = (1-t)u + tv$. Comme $[u,v] \subset \Omega$, la fonction φ est dérivable sur $[0,1]$ par composition, avec (corollaire 3.6)

$$\varphi'(t) = Df(\gamma(t))\gamma'(t) = Df(\gamma(t))(v-u)\,.$$

D'après l'inégalité des accroissements finis pour les fonctions d'une variable, on sait qu'il existe un point $c \in [0,1]$ tel que $\|\varphi(1)-\varphi(0)\| \leq (1-0)\times\|\varphi'(c)\|$. Comme $\gamma(1) = v$ et $\gamma(0) = u$, cela s'écrit

$$\|f(v) - f(u)\| \leq \|Df(\gamma(c))(v-u)\|\,,$$

d'où le résultat avec $\xi = Df(\gamma(c))$. La deuxième inégalité du théorème est claire puisque $\|Df(\xi)(v-u)\| \leq \|Df(\xi)\|\,\|v-u\|$. □

Corollaire 4.2. *Si l'ouvert* Ω *est* convexe *et s'il existe une constante* M *tel que* $\|Df(\xi)\| \leq M$ *pour tout* $\xi \in \Omega$*, alors* f *est* M*-lipschitzienne sur* Ω *:*

$$\forall u,v \in \Omega\ :\ \|f(v)-f(u)\| \leq M\,\|v-u\|\,.$$

Démonstration. C'est évident puisque $[u,v] \subset \Omega$ pour tous $u,v \in \Omega$. □

Corollaire 4.3. *Soit* Ω *un ouvert de* $\mathbb{R}^n$*. Si* $f : \Omega \to F$ *est* de classe $\mathcal{C}^1$ *sur* Ω*, alors* f *est lipschitzienne sur toute boule fermée* $\overline{B} \subset \Omega$*.*

Démonstration. La boule $\overline{B}$ est *compacte* car on est en dimension finie, donc la fonction *continue* $\xi \mapsto \|Df(\xi)\|$ est bornée sur $\overline{B}$: il existe une constante M telle que $\|Df(\xi)\| \leq M$ pour tout $\xi \in \overline{B}$. De plus, la boule $\overline{B}$ est également *convexe*, donc $\|Df(\xi)\| \leq M$ pour tout $\xi \in [u,v]$ si $u,v \in \overline{B}$. D'après l'inégalité des accroissements finis, on en déduit immédiatement le résultat. □

Corollaire 4.4. *Si l'ouvert* $\Omega \subset E$ *est* connexe *et si* $Df \equiv 0$ *dans* Ω*, alors* f *est constante sur* Ω*.*

Démonstration. D'après l'inégalité des accroissements finis (corollaire 4.2 avec $M = 0$), on a $\|f(v)-f(u)\| \leq 0$ (i.e. $f(u) = f(v)$) dès que $u,v \in \Omega$ sont tels que $[u,v] \subset \Omega$. Si maintenant u et v sont deux points quelconques de Ω alors, par connexité, on peut relier u et v par une ligne brisée $L = [a_0, a_1 \cup \cdots \cup [a_{N-1}, a_N]$ entièrement contenue dans Ω (voir la proposition 4.6 du chapitre 1). D'après ce qui précède, on a $f(a_{i+1}) = f(a_i)$ pour tout $i \in \{0,\ldots,N-1\}$, et donc $f(v) = f(a_N) = \cdots = f(a_0) = f(u)$. Comme u et v sont quelconques dans Ω, cela montre que f est constante.

On peut aussi démontrer directement le résultat, sans utiliser la proposition 4.6 du chapitre 1. Fixons un point $p \in \Omega$ et posons $\Omega_0 = \{u \in \Omega;\ f(u) = f(p)\}$ et $\Omega_1 = \{u \in \Omega;\ f(u) \neq f(p)\}$. Il s'agit de montrer que $\Omega_1 = \emptyset$. Comme Ω_0 et Ω_1 forment une partition de Ω, que Ω_1 est visiblement ouvert et que $\Omega_0 \neq \emptyset$ (car $p \in \Omega_0$), il suffit (par connexité de Ω) de montrer que Ω_0 est *ouvert*. Soit donc $u_0 \in \Omega_0$ quelconque, et soit $r > 0$ tel que $B = B(u_0, r) \subset \Omega$. Comme la boule $B(u,r)$ est *convexe*, f est constante sur B d'après l'inégalité des accroissements finis (corollaire 4.2 avec $M = 0$). On a ainsi $f(u) \equiv f(u_0) = f(p)$ sur B, autrement dit $B = B(u_0, r) \subset \Omega_0$. □

4.2. le "théorème fondamental de l'analyse". Dans cette section, E est un evn et F est un evn *complet* (par exemple de dimension finie).

Si Ω est un ouvert de E, on appelera **chemin dans** Ω toute application continue $\gamma : [a,b] \to E$ définie sur un intervalle compact $[a,b] \subset \mathbb{R}$, telle que $\gamma([a,b]) \subset \Omega$.

Théorème 4.5. (théorème fondamental de l'analyse)
Soit Ω *un ouvert de* E*, et soit* $f : \Omega \to F$ *de classe* $\mathcal{C}^1$*. Si* $\gamma : [a,b] \to E$ *est un chemin de classe* $\mathcal{C}^1$ *dans* Ω*, alors*

$$f(\gamma(b)) - f(\gamma(a)) = \int_a^b Df(\gamma(t))\gamma'(t)\,dt\,.$$

Démonstration. Soit $\varphi : [a,b] \to F$ la fonction d'*une variable* définie par $\varphi(t) = f(\gamma(t))$. La fonction φ est effectivement bien définie puisque $\gamma(t) \in \Omega$ pour tout $t \in [a,b]$, et par composition elle est de classe $\mathcal{C}^1$ avec $\varphi'(t) = Df(\gamma(t))\gamma'(t)$ (corollaire 3.6). Il suffit donc d'appliquer à φ le théorème fondamental de l'analyse pour les fonctions d'une variable démontré au chapitre 1. □

Corollaire 4.6. *Soit $f : \Omega \to F$ de classe $\mathcal{C}^1$. Si $u, v \in \Omega$ et si $[u,v] \subset \Omega$, alors*

$$f(v) - f(u) = \int_0^1 Df((1-t)u + tv)(v-u)\,dt\,.$$

Autrement dit, si $p \in \Omega$ et si $h \in E$ est tel que $[p, p+h] \subset \Omega$, alors

$$f(p+h) = f(p) + \int_0^1 Df(p+th)h\,dt\,.$$

Démonstration. C'est le théorème appliqué au chemin $\gamma : [0,1] \to E$ défini par $\gamma(t) = (1-t)u + tv$. □

Exercice. Utiliser le théorème fondamental de l'analyse pour redémontrer l'inégalité des accroissements finis, dans le cas d'une application de classe $\mathcal{C}^1$.

Comme illustration du théorème fondamental de l'analyse, on va maintenant démontrer un résultat concernant les suites d'applications de classe $\mathcal{C}^1$. C'est l'analogue de la proposition 6.13 du chapitre 1.

Proposition 4.7. *Soit Ω un ouvert de E, soit (f_k) une suite de fonctions de classe $\mathcal{C}^1$ sur Ω, à valeurs dans F, et soit $f : \Omega \to F$. On suppose que*

(i) *la suite (f_k) converge simplement vers f sur Ω ;*

(ii) *la suite (Df_k) converge* uniformément sur toute boule fermée $\overline{B} \subset \Omega$ *vers une application $\Phi : \Omega \to \mathcal{L}(E,F)$.*

Alors f est de classe $\mathcal{C}^1$ et $Df = \Phi$.

Démonstration. Comme les Df_k sont continues et convergent uniformément vers Φ sur toute boule boule ouverte $B = B(p,r)$ telle que $\overline{B} \subset \Omega$, l'application Φ est *continue* (car continue au voisinage de chaque point $p \in \Omega$). Il suffit donc de montrer que f est différentiable sur Ω avec $DF = \Phi$.

Soit $p \in \Omega$, et fixons $r > 0$ tel $\overline{B}(p,r) \subset \Omega$. Alors $[p, p+h] \subset \Omega$ pour tout $h \in E$ vérifiant $\|h\| \leq r$, par convexité de la boule $\overline{B} = \overline{B}(p,r)$. Pour un tel h, on a donc

$$\forall k \in \mathbb{N} \ : \ f_k(p+h) = f_k(p) + \int_0^1 Df_k(p+th)h\,dt\,.$$

Comme $Df_k(p+th)h$ tend vers $\Phi(p+th)h$ uniformément sur $[0,1]$ par (ii) (car $[p,p+h] \subset \overline{B}$), on peut passer à la limite sous l'intégrale et on obtient

$$f(p+h) = f(p) + \int_0^1 \Phi(p+th)h\,dt$$

pour tout h vérifiant $\|h\| \leq r$.

Comme Φ est continue, on peut écrire $\Phi(p+k) = \Phi(p) + \varepsilon(k)$, où $\varepsilon(k)$ tend vers 0 quand $k \to 0$. Pour $\|h\| \leq r$, on a donc

$$\begin{aligned} f(p+h) &= f(p) + \int_0^1 \big[\Phi(p) + \varepsilon(th)\big]h\,dt \\ &= f(p) + \Phi(p)h + R(h)\,, \end{aligned}$$

où $R(h) = \int_0^1 \varepsilon(th)h\,dt$. De plus,

$$\begin{aligned}\|R(h)\| &\leq \int_0^1 \|\varepsilon(th)h\|\,dt \\ &\leq \|h\| \times \int_0^1 \|\varepsilon(th)\|dt\end{aligned}$$

Comme $\varepsilon(k)$ tend vers 0 quand $k \to 0$, l'intégrale $\int_0^1 \|\varepsilon(th)\|\,dt$ tend vers 0 quand $h \to 0$ (*exercice*). On a donc $R(h) = o(\|h\|)$ quand $h \to 0$, et comme $\Phi(p) \in \mathcal{L}(E, F)$, cela prouve que f est différentiable en p avec $Df(p) = \Phi(p)$.

□

Corollaire 4.8. *Soit Ω un ouvert de $\mathbb{R}^n$, et soit (f_k) une suite de fonctions de classe $\mathcal{C}^1$ sur Ω, à valeurs réelles. Soit également $f : \Omega \to \mathbb{R}$. On suppose que*

(i) *la suite (f_k) converge simplement vers f sur Ω ;*

(ii) *pour tout $j \in \{1, \dots, n\}$ la suite $\left(\frac{\partial f_k}{\partial x_j}\right)_{k\in\mathbb{N}}$ converge* uniformément sur tout compact de Ω *vers une fonction $g_j : \Omega \to \mathbb{R}$.*

Alors f est de classe $\mathcal{C}^1$ et $\frac{\partial f_k}{\partial x_j} = g_j$ pour tout $j \in \{1, \dots, n\}$.

Démonstration. On applique la proposition avec $E = \mathbb{R}^n$ et $F = \mathbb{R}$, en identifiant $\mathcal{L}(\mathbb{R}^n, \mathbb{R})$ avec l'espace $M_{1,n}(\mathbb{R})$ des matrices $1 \times n$ (et donc en identifiant la différentielle d'une fonction en un point avec sa matrice jacobienne en ce point). L'hypothèse (ii) dit que $\mathrm{Jac}_{f_k}(u)$ tend uniformément sur tout compact vers $(g_1(u), \dots, g_n(u))$. En notant $\Phi(u) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ l'application linéaire de matrice $(g_1(u), \dots, g_n(u))$, cela signifie que $Df_k(u)$ tend vers $\Phi(u)$ uniformément sur tout compact (donc sur toute boule fermée puisqu'on est en dimension finie). On peut donc effectivement appliquer la proposition. □

Corollaire 4.9. *Soit Ω un ouvert de $\mathbb{R}^n$, soit (v_k) une suite de fonctions de classe $\mathcal{C}^1$ sur Ω, à valeurs réelles. On suppose que*

(i) *la série $\sum v_k$ converge simplement sur Ω ;*

(ii) *pour tout $j \in \{1, \dots, n\}$ la série $\sum \frac{\partial v_k}{\partial x_j}$ converge* normalement sur tout compact de Ω.

Alors la fonction $f = \sum_{k=0}^{\infty} v_k$ est de classe $\mathcal{C}^1$ et $\frac{\partial f}{\partial x_j} = \sum_{k=0}^{\infty} \frac{\partial v_k}{\partial x_j}$ pour tout $j \in \{1, \dots, n\}$.

Démonstration. On applique la proposition aux sommes partielles $f_k = \sum_{i=0}^k v_i$, en utilisant le fait que la convergence normale d'une série de fonctions entraine sa convergence uniforme. □

Exercice. Énoncer et démontrer une version de ce dernier corollaire pour un evn de départ E quelconque et des fonctions u_k à valeurs dans F.

5. Fonctions de classe $\mathcal{C}^2$

5.1. Définitions. Dans ce qui suit, Ω est un ouvert de $\mathbb{R}^n$.

Définition 5.1. *On dit qu'une fonction $f : \Omega \to \mathbb{R}$ est **deux fois différentiable** en un point $p \in \Omega$ si f est différentiable au voisinage de p et si ses dérivées partielles* *sont différentiables en p. On dit que f est* ***de classe*** $\mathcal{C}^2$ *sur* Ω *si f est de classe* $\mathcal{C}^1$ *et si ses dérivées partielles* $\partial_1 f, \dots, \partial_n f$ *sont elle mêmes de classe* $\mathcal{C}^1$*. On note* $\mathcal{C}^2(\Omega)$ *l'ensemble des fonctions de classe* $\mathcal{C}^2$ *sur* Ω*.*

REMARQUE. Ces définitions peuvent se reformuler de manière "intrinsèque", i.e. sans faire intervenir les dérivées partielles : (i) f est deux fois différentiable en p si et seulement si elle est différentiable sur un voisinage V de p et $Df = V \to \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ est différentiable de p ; (ii) f est de classe $\mathcal{C}^2$ si et seulement si elle est de classe $\mathcal{C}^1$ et $Df : \Omega \to \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ est de classe $\mathcal{C}^1$.

DÉMONSTRATION. C'est immédiat en identifiant $\mathcal{L}(\mathbb{R}^n, \mathbb{R})$ à $M_{1,n}(\mathbb{R}) = \mathbb{R}^n$, et donc Df à $\mathrm{Jac}_f = (\partial_1 f, \dots, \partial_n f)$. $\square$

NOTATIONS. Pour $i, j \in \{1, \dots, n\}$, on note $\partial_i\partial_j f$ ou $\frac{\partial^2 f}{\partial x_i \partial x_j}$ la dérivée partielle "itérée" $\partial_i(\partial_j f) = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)\cdot$ (Ces fonctions sont définies en tout point où f est deux fois différentiable). Si $i = j$, on écrit $\frac{\partial^2 f}{\partial x_i^2}$ au lieu de $\frac{\partial^2 f}{\partial x_i \partial x_i}\cdot$

Remarque. Bien entendu, si $n = 2$ on emploie plutôt les notations $\frac{\partial^2 f}{\partial x \partial y}$, $\frac{\partial^2 f}{\partial y \partial x}$, $\frac{\partial^2 f}{\partial x^2}$ et $\frac{\partial^2 f}{\partial y^2}\cdot$ Et de même si $n = 3$.

DÉFINITION 5.2. *Soit* $f : \Omega \to \mathbb{R}$ *deux fois différentiable en un point p. La* ***différentielle seconde de f au point*** *p est la forme bilinéaire* $d^2 f(p) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *définie par*

$$d^2 f(p)(h, h') = \sum_{i,j=1}^{n} \partial_i \partial_j f(p)\, h_i h'_j\,,$$

où on a écrit les vecteurs $h, h' \in \mathbb{R}^n$ *sous la forme* $h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$ *et* $h' = \begin{pmatrix} h'_1 \\ \vdots \\ h'_n \end{pmatrix}$.

Remarque 1. Notons $(e_1, \dots, e_n)$ la base canonique de $\mathbb{R}^n$. Pour $i, j \in \{1, \dots, n\}$, on a

$$d^2 f(p)(e_i, e_j) = \partial_i \partial_j f(p)\,.$$

Remarque 2. La matrice de la forme bilinéaire $d^2 f(p)$ dans la base canonique de $\mathbb{R}^n$ s'appelle la **matrice hessienne de f au point** p, et on la notera $H_f(p)$. Ainsi, on a

$$H_f(p) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(p)\right)_{i,j=1}^{n}.$$

Exercice. Pour $h, h' \in \mathbb{R}^n$, exprimer $d^2 f(p)(h, h')$ à l'aide de $H_f(p)$ et du produit scalaire usuel de $\mathbb{R}^n$.

PROPOSITION 5.3. *Si* $f : \Omega \to \mathbb{R}$ *est deux fois différentiable en un point* $p \in \Omega$*, alors la "dérivée directionnelle itérée"* $\partial_h \partial_{h'} f(p)$ *existe pour tous* $h, h' \in \mathbb{R}^n$*, et on a*

$$d^2 f(p)(h, h') = \partial_h \partial_{h'} f(p)\,.$$

DÉMONSTRATION. Fixons $h, h' \in \mathbb{R}^n$.

Comme f est différentiable sur un certain voisinage U de p, on sait que $\partial_{h'} f(u)$ existe en tout point $u \in U$, avec

$$\partial_{h'} f(u) = \sum_{j=1}^{n} \partial_j f(u)\, h'_j\,.$$

Comme f est deux fois différentiable en p, toutes les fonctions $\partial_j f$ sont différentiables en p. Donc $\partial_{h'} f$ est différentiable en p, et par conséquent $\partial_h(\partial_{h'} f)(p)$ existe avec

$$\begin{aligned} \partial_h(\partial_{h'} f)(p) &= d(\partial_{h'} f)(p)\, h \\ &= \sum_{j=1}^n h'_j\, d(\partial_j f)(p) h \\ &= \sum_{j=1}^n h'_j \left(\sum_{i=1}^n h_i\, \partial_i \partial_j f(p) \right) \\ &= d^2 f(p)(h,h')\,. \end{aligned}$$

□

5.2. Symétrie des dérivées “croisées”. Le théorème suivant signifie que l'ordre dans lequel on effectue des dérivations partielles n'a en fait aucune importance. C'est un résultat *a priori* surprenant, et bien entendu fondamental. Dans tout ce qui suit, Ω est comme d'habitude un ouvert de $\mathbb{R}^n$.

THÉORÈME 5.4. (théorème de Schwarz)
Si $f \in \mathcal{C}^2(\Omega)$, alors $\partial_i \partial_j f = \partial_j \partial_i f$ pour tous $i, j \in \{1, \ldots, n\}$.

DÉMONSTRATION. L'énoncé ne faisant intervenir que 2 variables à la fois, on peut se limiter au cas d'une fonction de 2 variables. Autrement dit, on se donne une fonction f de classe $\mathcal{C}^2$ sur $\Omega \subset \mathbb{R}^2$, et il s'agit de montrer qu'on a

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}\cdot$$

La preuve repose sur le lemme suivant. Rappelons que si $R = [a,b] \times [c,d]$ est un rectangle de $\mathbb{R}^2$ et si $\Phi : R \to \mathbb{R}$ est une fonction continue, alors on a

$$\int_a^b \left(\int_c^d \Phi(x,y)\, dy \right) dx = \int_a^b \left(\int_c^d \Phi(x,y)\, dy \right) dx\,,$$

et que la valeur commune de ces deux intégrales itérées est notée $\int_R \phi(x,y)\, dxdy$.

LEMME 5.5. *Soient Φ_1 et Φ_2 deux fonctions continues sur Ω (à valeurs réelles). On suppose qu'on a $\int_R \Phi_1(x,y)\, dxdy = \int_R \Phi_2(x,y)\, dxdy$ pour tout rectangle $R \subset \Omega$. Alors $\Phi_1 = \Phi_2$.*

DÉMONSTRATION DU LEMME. Soit $\Phi = \Phi_2 - \Phi_1$. On a $\int_R \Phi(x,y)\, dxdy = 0$ pour tout rectangle $R \subset \Omega$, et on veut montrer que $\Phi = 0$. Supposons que ce ne soit pas le cas. Alors on peut par exemple trouver un point $p_0 = (x_0, y_0) \in \Omega$ tel que $\Phi(p_0) > 0$. Soit α_0 tel que $0 < \alpha_0 < \Phi(p_0)$. Par continuité de Φ, on peut trouver $\delta > 0$ tels que $\|u - p_0\|_\infty \leq \delta \implies \Phi(u) \geq \alpha_0$ (prendre $\varepsilon = \Phi(p_0) - \alpha_0$ dans la définition de la continuité ...). La boule fermée $\overline{B}(p_0, \delta)$ est le carré $R = [x_0 - \delta, x_0 + \delta] \times [y_0 - \delta, y_0 + \delta]$. Par le choix de δ, on a

$$\int_R \Phi(x,y)\, dxdy \geq \int_R \alpha_0\, dxdy = \alpha_0 \times (2\delta)^2 > 0\,,$$

ce qui contredit l'hypothèse faite sur Φ.

□

On applique le lemme avec $\Phi_1 = \frac{\partial^2 f}{\partial x \partial y}$ et $\Phi_2 = \frac{\partial^2 f}{\partial y \partial x}$. Si $R = [a,b] \times [c,d]$ est un rectangle contenu dans Ω, alors

$$\begin{aligned}
\int_R \frac{\partial^2 f}{\partial x \partial y}\, dxdy &= \int_c^d \left(\int_a^b \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y}(x,y) \right) dx \right) dy \\
&= \int_c^d \left[\frac{\partial f}{\partial y}(b,y) - \frac{\partial f}{\partial y}(a,y) \right] dy \\
&= \big[f(b,d) - f(b,c) \big] - \big[f(a,d) - f(a,c) \big] .
\end{aligned}$$

De même, on trouve

$$\begin{aligned}
\int_R \frac{\partial^2 f}{\partial y \partial x}\, dxdy &= \int_a^b \left(\int_c^d \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x}(x,y) \right) dy \right) dx \\
&= \big[f(b,d) - f(a,d) \big] - \big[f(b,c) - f(a,c) \big] .
\end{aligned}$$

Ainsi, on a $\int_R \frac{\partial^2 f}{\partial x \partial y}\, dxdy = \int_R \frac{\partial^2 f}{\partial y \partial x}\, dxdy$ pour tout rectangle $R \subset \Omega$, et donc $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$. □

COROLLAIRE 5.6. *Si $f \in \mathcal{C}^2(\Omega)$, alors $d^2 f(p)$ est une forme bilinéaire* symétrique *(pour tout $p \in \Omega$). Pour $h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \in \mathbb{R}^n$, on a donc*

$$d^2 f(p)(h,h) = \sum_{j=1}^n \frac{\partial^2 f}{\partial x_j^2}\, h_j^2 + 2 \sum_{i<j} \frac{\partial^2 f}{\partial x_i \partial x_j}\, h_i h_j \, .$$

Remarque. On a énoncé le théorème de Schwarz pour une fonction de classe $\mathcal{C}^2$, mais le résultat est en fait valable sous une hypothèse beaucoup plus faible (et plus naturelle) : on a $\partial_i \partial_j f(p) = \partial_i \partial_j f(p)$ dès lors que f est *deux fois différentiable* en un point p. La preuve de ce résultat étant un peu moins transparente, on a préféré se limiter au cas $\mathcal{C}^2$.

5.3. La formule de Taylor à l'ordre 2. Dans cette section, on étend la formule de Taylor (théorème 3.1 du chapitre 2) au cas des fonctions de plusieurs variables. Comme on n'a pas parlé de différentiabilité à un ordre supérieur à 2, on ne peut énoncer cette formule qu'à l'ordre 2. Pour $h \in \mathbb{R}^n$, on posera

$$h^{[2]} = (h,h) \, .$$

PROPOSITION 5.7. (formule de Taylor à l'ordre 2)
Soit $f \in \mathcal{C}^2(\Omega)$. Si $p \in \Omega$ et si $h \in \mathbb{R}^n$ est tel que $[p, p+h] \subset \Omega$, alors

$$f(p+h) = f(p) + df(p)h + \int_0^1 (1-t)\, d^2 f(p+th) h^{[2]}\, dt \, .$$

La démonstration utilise le lemme suivant.

LEMME 5.8. *Soit $f \in \mathcal{C}^2(\Omega)$, et soient p, h tels que $[p, p+h] \subset \Omega$. On définit $\varphi : [0,1] \to \mathbb{R}$ par $\varphi(t) = f(p+th)$. Alors φ est de classe $\mathcal{C}^2$, et on a*

$$\begin{aligned}
\varphi'(t) &= df(p+th)h \, , \\
\varphi''(t) &= d^2 f(p+th) h^{[2]} \, .
\end{aligned}$$

Démonstration. Comme f est de classe $\mathcal{C}^1$, la fonction φ est $\mathcal{C}^1$ par composition, avec $\varphi'(t) = df(p+th)h = \partial_h f(p+th)$. Mais f est en fait $\mathcal{C}^2$, donc la fonction $g = \partial_h f$ est $\mathcal{C}^1$. Par conséquent, φ' est de classe $\mathcal{C}^1$, autrement dit φ est $\mathcal{C}^2$, avec

$$\begin{aligned} \varphi''(t) &= d(\partial_h f)(p+th)h \\ &= \partial_h \partial_h f(p+th) \\ &= d^2 f(p+th)h^{[2]} \end{aligned}$$

d'après la proposition 5.3. □

Démonstration de la proposition 5.7. Il suffit maintenant d'appliquer la formule de Taylor à l'ordre 2 pour les fonctions d'une variable à la fonction φ définie dans le lemme. □

Corollaire 5.9. (développement limité à l'ordre 2)
Soit $f \in \mathcal{C}^2(\Omega)$. Pour tout point $p \in \Omega$, on a le développement limité suivant quand $h \to 0$ (où $\|\cdot\|$ est une norme quelconque sur $\mathbb{R}^n$) :

$$f(p+h) = f(p) + df(p)h + \frac{1}{2}\, d^2 f(p)h^{[2]} + o(\|h\|^2)\,.$$

Démonstration. On munit par exemple $\mathbb{R}^n$ de la norme $\|\cdot\|_\infty$. Fixons $p \in \Omega$ et choisissons $r > 0$ tel que $\overline{B}(p,r) \subset \Omega$.

Par convexité de la boule $\overline{B}(p,r)$, on a $[p, p+h] \subset \Omega$ pour tout $h \in \mathbb{R}^n$ vérifiant $\|h\|_\infty \le r$. Pour un tel $h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$, la formule de Taylor s'écrit

$$\begin{aligned} f(p+h) &= f(p) + df(p)h + \int_0^1 (1-t) d^2 f(p+th)h^{[2]}\, dt \\ &= f(p) + df(p)h + \int_0^1 (1-t) \left(\sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(p+th)\, h_i h_j \right) dt. \end{aligned}$$

De plus, comme f est de classe $\mathcal{C}^2$ on peut écrire

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(p+u) = \frac{\partial^2 f}{\partial x_i \partial x_j}(p) + \varepsilon_{ij}(u)\,,$$

où $\varepsilon_{ij}(u) \to 0$ quand $u \to 0$. On obtient alors

$$\begin{aligned} f(p+h) &= f(p) + df(p)h + \sum_{i,j=1}^n h_i h_j \frac{\partial^2 f}{\partial x_i \partial x_j}(p) \times \underbrace{\int_0^1 (1-t)\, dt}_{1/2} + R(h) \\ &= f(p) + df(p)h + \frac{1}{2} d^2 f(p)h^{[2]} + R(h)\,, \end{aligned}$$

où le "reste" $R(h)$ est donné par la formule

$$R(h) = \sum_{i,j=1}^n \left[\int_0^1 (1-t)\varepsilon_{ij}(th)\, dt \right] h_i h_j\,.$$

Quand h tend vers 0, chaque $\varepsilon_{ij}(th)$ tend vers 0 *uniformément* sur $[0,1]$ (car $th \to 0$ uniformément), et donc $\int_0^1 (1-t)\varepsilon_{ij}(th)dt \to 0$. Par conséquent, le terme d'indice (i,j)

dans la somme précédente est un $o(|h_i h_j|)$. Comme $|h_i h_j| \leq \|h\|_\infty^2$, on en déduit

$$R(h) = o(\|h\|_\infty^2)\,,$$

ce qui termine la démonstration. □

Remarque 1. En fait, on peut montrer que f admet un DL d'ordre 2 au point p en supposant seulement que f est 2 fois différentiable en p.

Remarque 2. Pour déterminer le développement limité à l'ordre 2 d'une fonction de plusieurs variables en un point donné, on peut en théorie toujours appliquer la formule du corollaire précédent ; mais dans la pratique, cela devient rapidement pénible pour une fonction un peu compliquée, car cela suppose de calculer des dérivées partielles secondes. Il est souvent plus rapide de procéder comme pour les fonctions d'une variable, i.e. en "composant" des DL connus *de fonctions d'une variable.*

Exercice. Déterminer le développement limité à l'ordre 2 en $(0,0)$ de la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par $f(x,y) = \cos(x+3y^2) + e^{x^2-4y}$.

CHAPITRE 4

Extrema

1. Introduction

Soit A une partie d'un evn E, et soit $f : A \to \mathbb{R}$. Les deux questions suivantes sont assurément "naturelles" :

- La fonction f possède-t-elle un maximum ou un minimum ?
- Si oui, comment comment le calculer, et comment déterminer le ou les points où il est atteint ?

Dans ce chapitre, on donne quelques réponses à ces questions. La terminologie suivante sera constamment utilisée :

DÉFINITION 1.1. *Soit $f : A \to \mathbb{R}$, et soit $a \in A$.*

(1) *On dit que f possède*
  – ***un maximum global en*** *a si $\forall u \in A\ :\ f(u) \leq f(a)$;*
  – ***un maximum local en*** *a si on peut trouver un voisinage ouvert V de a tel que $\forall u \in V \cap A\ :\ f(u) \leq f(a)$;*

(2) *On définit de même les notions de minimum global et de minimum local.*

(3) *On dit que f possède un* ***extremum*** *(global ou local) en a si f possède un maximum ou un minimum (global ou local) en a.*

Remarque 1. Évidemment, tout extremum global est un extremum local.

Remarque 2. Le pluriel de "extremum" n'est pas "extremums" mais "extrema". De même pour maximum ou minimum.

Exercice. Dessiner le graphe d'une fonction $f : \mathbb{R} \to \mathbb{R}$ possédant 2 mimina locaux, 3 maxima locaux, 1 minimum global et aucun maximum global.

2. Compacité et existence d'extrema

THÉORÈME 2.1. *Soit A une partie d'un evn E, et soit $f : A \to \mathbb{R}$. Pour $\alpha \in \mathbb{R}$, notons $\{f \leq \alpha\}$ et $\{f \geq \alpha\}$ les ensembles $\{u \in A;\ f(u) \leq \alpha\}$ et $\{u \in A;\ f(u) \geq \alpha\}$.*

(1) *S'il existe $\alpha_0 \in \mathbb{R}$ tel que $\{f \leq \alpha_0\} \neq \emptyset$ et $\{f \leq \alpha\}$ est* compact *pour tout $\alpha \leq \alpha_0$, alors f possède un minimum global.*

(2) *S'il existe $\alpha_0 \in \mathbb{R}$ tel que $\{f \geq \alpha_0\} \neq \emptyset$ et $\{f \geq \alpha\}$ est compact pour tout $\alpha \geq \alpha_0$, alors f possède un maximum global.*

La démonstration repose sur le lemme suivant, très important par ailleurs.

LEMME 2.2. (théorème des compacts emboités)

Soit $(K_n)_{n\in\mathbb{N}}$ une suite de parties compactes d'un evn E. On suppose que les K_n sont tous non vides, et que la suite (K_n) est décroissante *($K_{n+1} \subset K_n$ pour tout n). Alors $\bigcap_{n\in\mathbb{N}} K_n \neq \emptyset$.*

Démonstration. Choisissons pour tout n un point $x_n \in K_n$. Comme les x_n appartiennent tous au compact K_0, la suite (x_n) possède une sous-suite $(x_{n_k})_{k\in\mathbb{N}}$ qui converge vers un point $a \in E$. Si $n \in \mathbb{N}$ est fixé, alors $n_k \geq n$ pour tout k assez grand, et donc $x_{n_k} \in K_n$ puisque $K_{n_k} \subset K_n$. Comme K_n est *fermé* dans E et $x_{n_k} \to a$, on en déduit que $a \in K_n$, pour tout $n \in \mathbb{N}$. Ainsi, $\bigcap_n K_n \neq \emptyset$.

□

Preuve du théorème. Il suffit de démontrer (1), car on obtient ensuite (2) en appliquant (1) à $-f$. Posons $m = \inf_A f \in [-\infty, +\infty[$ (par convention, $m = -\infty$ si f n'est pas minorée). Par définition de m, il suffit de montrer qu'il existe un point $a \in A$ tel que $f(a) \leq m$.

Si $\alpha_0 = m$, il n'y a rien à démontrer. Si $\alpha_0 > m$, choisissons une suite décroissante de nombres réels $(\alpha_n)_{n\geq 1}$ telle que $m < \alpha_n \leq \alpha_0$ pour tout n et $\alpha_n \to m$. Par définition de m, les ensembles $K_n = \{f \leq \alpha_n\}$ sont tous non vides. De plus, les K_n sont compacts par hypothèse, et la suite (K_n) est décroissante puisque la suite (α_n) l'est. D'après le théorème des compacts emboités, on a donc $\bigcap_{n\in\mathbb{N}} K_n \neq \emptyset$; soit $a \in \bigcap_n K_n$. On a $f(a) \leq \alpha_n$ pour tout n, donc $f(a) \leq m$ puisque $\alpha_n \to m$. □

Corollaire 2.3. *Supposons f* continue. *S'il existe $\alpha_0 \in \mathbb{R}$ tel que $\{f \leq \alpha_0\}$ est non-vide et compact, alors f possède un minimum global. De même, s'il existe $\alpha_0 \in \mathbb{R}$ tel que $\{f \geq \alpha_0\}$ est non-vide et compact, alors f possède un maximum global.*

Démonstration. Par continuité de f, tous les ensembles $\{f \leq \alpha\}$ pour $\alpha \leq \alpha_0$ sont fermés dans le compact $\{f \leq \alpha_0\}$, donc compacts.

□

Corollaire 2.4. *Si K est un compact d'un evn E, alors toute fonction continue sur K possède un maximum et un minimum (globaux).*

Démonstration. Par continuité de f, tous les ensembles $\{f \leq \alpha\}$ sont fermés dans K, donc compacts ; et il y en a certainement un qui est non-vide. □

Corollaire 2.5. *Soit E un evn* de dimension finie*, et soit M un* fermé *de E. Si $f : M \to \mathbb{R}$ est continue et vérifie $\lim_{\|x\|\to\infty} f(x) = +\infty$, alors f possède un minimum global.*

Démonstration. Si $\alpha \in \mathbb{R}$, alors $\{f \leq \alpha\}$ est un fermé de M par continuité de f, donc un fermé de E car M est fermé ; et $\{f \leq \alpha\}$ est également *borné* par hypothèse sur f. Donc $\{f \leq \alpha\}$ est compact pour tout $\alpha \in \mathbb{R}$ car E est de dimension finie. □

Exemple. Soit X est un evn, et soit $E \subset X$ est un sous-espace de dimension finie. En appliquant ce dernier corollaire avec $M = E$ et $f(u) = \|u - a\|$, on obtient le résultat suivant : *pour tout $a \in X$, il existe au moins un point $p \in E$ tel que* $\|p - a\| = \operatorname{dist}(a, E) := \inf\{\|u - a\|;\ u \in E\}$.

3. Conditions sur les différentielles première et seconde

3.1. Fonctions d'une variable. Le lemme suivant est très élémentaire, mais néammoins fort important.

Lemme 3.1. *Soit I un intervalle de $\mathbb{R}$, et soit $\varphi : I \to \mathbb{R}$.*

(1) *Si φ admet un extremum local en un point a* intérieur à *I et si φ est dérivable en a, alors $\varphi'(a) = 0$.*

(2) *Soit $a \in \mathring{I}$ tel que φ est deux fois dérivable en a.*

(i) *Si φ admet un maximum local en a, alors $\varphi''(a) \leq 0$.*
(ii) *Si φ admet un minimum local en a, alors $\varphi''(a) \geq 0$.*

(3) *Soit toujours $a \in \mathring{I}$ tel que φ est deux fois dérivable en a.*
(i) *Si $\varphi'(a) = 0$ et $\varphi''(a) < 0$, alors φ admet un maximum local en a.*
(ii) *Si $\varphi'(a) = 0$ et $\varphi''(a) > 0$, alors φ admet un minimum local en a.*

Remarque 1. Dans (1), on ne peut pas conclure que $\varphi'(a) = 0$ si a n'est pas intérieur à I, i.e. si a est une des extrémités de l'intervalle I. Par exemple, la fonction φ définie sur $[0, 1]$ par $\varphi(t) = t$ possède un minimum (global) en $t = 0$ et un maximum en $t = 1$, mais φ' ne s'annule jamais.

Remarque 2. La condition $\varphi'(a) = 0$ et les conditions (i) et/ou (ii) de (2) ne suffisent pas à assurer l'existence d'un extremum local. Par exemple, $\varphi(t) = t^3$ ne possède pas d'extremum local, et cependant $\varphi'(0) = 0$ et $\varphi''(0) = 0$.

Remarque 3. Sous les hypothèses de (3), la fonction φ ne possède pas nécessairement de maximum ou de minimum *global* en a. Par exemple, si $\varphi : \mathbb{R} \to \mathbb{R}$ est définie par $\varphi(t) = t^3 + t^2$, alors $\varphi'(0) = 0$ et $\varphi''(0) = 2 > 0$, mais φ ne possède pas de minimum global car elle n'est même pas minorée (elle tend vers $-\infty$ en $-\infty$). D'autre part, les hypothèses de (3) ne sont pas *nécessaires* pour avoir un extremum local. Par exemple, $\varphi(t) = t^4$ possède un minimum (global) en 0 mais $\varphi''(t) = 0$.

Démonstration du lemme. (1) Supposons par exemple que φ possède un maximum local en a. Comme $a \in \mathring{I}$, on peut alors trouver $\delta > 0$ tel que $]a - \delta, a + \delta[\subset I$ et $\varphi(t) \leq \varphi(a)$ pour tout $t \in]a - \delta, a + \delta[$. alors $\varphi(a + h)$ et $\varphi(a - h)$ sont définis pour tout h vérifiant $0 < h < \delta$, et on a

$$\frac{\varphi(a+h) - \varphi(a)}{h} \leq 0 \leq \frac{\varphi(a-h) - \varphi(a)}{-h}\,;$$

d'où $\varphi'(a) \leq 0 \leq \varphi'(a)$ en faisant tendre h vers 0^+.

(2) Supposons que a possède un extremum local en a. Alors $\varphi'(a) = 0$ par (1), donc le développement limité à l'ordre 2 de φ au point a s'écrit

$$\varphi(a + h) = \varphi(a) + 0 + \frac{h^2}{2}\varphi''(a) + o(h^2)\,.$$

On en déduit

$$\varphi''(a) = 2 \lim_{h \to 0} \frac{\varphi(a+h) - \varphi(a)}{h^2}\,.$$

Si φ possède un maximum local en a, le quotient du membre de droite est négatif ou nul pour h assez petit, et donc $\varphi''(a) \leq 0$; et de même, on obtient $\varphi''(a) \geq 0$ si φ possède un minimum local en a.

(3) Supposons que $\varphi'(a) = 0$. D'après la preuve de (2), on a alors

$$\lim_{h \to 0} \frac{\varphi(a+h) - \varphi(a)}{h^2} = \frac{1}{2}\,\varphi''(a)\,.$$

Si $\varphi''(a) < 0$, on en déduit qu'on a $\frac{\varphi(a+h)-\varphi(a)}{h^2} < 0$ pour h suffisamment proche de 0, i.e. $\varphi(t) < \varphi(a)$ pour $t \neq a$ suffisamment proche de a : la fonction φ possède donc un maximum local en a. De même, φ possède un minimum local en a si $\varphi''(a) > 0$. $\square$

3.2. Condition du premier ordre. Le résultat suivant est la généralisation "évidente" de la partie (1) du lemme 3.1.

THÉORÈME 3.2. *Soit A une partie d'un evn E, et soit $f : A \to \mathbb{R}$. Si f admet un extremum local en un point a* intérieur à *A et si f est différentiable en a, alors $df(a) = 0$.*

DÉMONSTRATION. Comme $a \in \mathring{A}$, on peut fixer $r > 0$ tel que $B(a, r) \subset A$.

Soit $h \in E$ quelconque, et choisissons $\delta > 0$ tel que $\delta \times \|h\| < r$. Alors $\varphi(t) = f(a + th)$ est bien défini pour $t \in]-\delta, \delta[$ (car $\|th\| < r$) et la fonction φ est dérivable en 0 par composition, avec $\varphi'(0) = df(a)h$. De plus, φ possède un extremum local en 0 par hypothèse sur f, et donc $\varphi'(0) = 0$ d'après le lemme 3.1. Ainsi, on a $df(a)h = 0$ pour tout $h \in E$. $\square$

Remarque 1. Un point pour lequel $df(a) = 0$ s'appelle un **point critique** pour la fonction f. On a donc montré que tout extremum local pour une fonction f définie et différentiable sur un ouvert de E est un point critique pour f.

Remarque 2. Si $E = \mathbb{R}^n$, la condition $df(a) = 0$ équivaut au système de n équations $\frac{\partial f}{\partial x_1}(a) = \cdots = \frac{\partial f}{\partial x_n}(a) = 0$.

EXEMPLE. Soit H un espace euclidien, et soit E un sous-espace vectoriel de H. Soit aussi $a \in H$. D'après le corollaire 2.5, on sait qu'il existe au moins un point $p \in E$ tel que $\|p - a\| = \text{dist}(a, E)$. Si on introduit la fonction $f : E \to \mathbb{R}$ définie par $f(u) = \|u - a\|^2$, cela signifie que f possède un minimum (global).

Comme le produit scalaire est une application bilinéaire continue, la fonction $\Phi : H \to \mathbb{R}$ définie par $\Phi(x) = \|x\|^2 = \langle x, x\rangle$ est différentiable sur H avec $D\Phi(x)h = \langle x, h\rangle + \langle h, x\rangle = 2\langle x, h\rangle$; donc f est différentiable sur E avec $df(p)h = 2\langle u - a, h\rangle$ pour tous $p, h \in E$.

Les points critiques de f sont les points $p \in E$ tels que $\langle p - a, h\rangle = 0$ pour tout $h \in E$. Il y a au plus 1 point $p \in E$ vérifiant cette propriété : en effet si p et p' sont deux tels points, alors $\langle p - p', h\rangle\langle p - a, h\rangle - \langle a - p', h\rangle = 0$ pour tout $h \in E$ et donc $\|p - p'\|^2 = 0$ en prenant $h = p - p'$. Mais on sait également qu'il y a au moins 1 point critique pour f puisque f possède un minimum. En conclusion, on a montré le résultat suivant : *il existe un unique point $p \in E$ tel $(p - a) \perp E$, et ce point est également l'unique point de E tel que $\|p - a\| = \text{dist}\,(a, E)$.* On a bien sûr reconnu le **projeté orthogonal** de a sur E. Ce que dit cet exemple est que le calcul différentiel permet de prouver l'existence du projeté orthogonal.

Exercice. Montrer directement qu'on a $(p-a) \perp E$ si et seulement si $\|p-a\| \leq \|u-a\|$ pour tout $u \in E$.

3.3. Conditions du deuxième ordre. Dans cette section, on va démontrer l'analogue des parties (2) et (3) du lemme 3.1 pour les fonctions de plusieurs variables.

3.3.1. *Formes bilinéaires positives, matrices positives.*

DÉFINITION 3.3. *On dit qu'une forme bilinéaire symétrique $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ est* ***positive*** *si on a $B(h, h) \geq 0$ pour tout $h \in \mathbb{R}^n$, et que B est* ***définie positive*** *si on a $B(h, h) > 0$ pour tout $h \neq 0$. On dit qu'une* matrice *$M \in M_n(\mathbb{R})$ est positive ou définie positive si M est symétrique et si la forme bilinéaire B définie par $B(h, h') = \langle Mh, h'\rangle$ est positive ou définie positive (où on a noté $\langle\ ,\ \rangle$ le produit scalaire usuel sur $\mathbb{R}^n$).*

Remarque 1. On définit de même les formes bilinéaires **négatives** et les matrices négatives. Évidemment, B ou M est négative si et seulement si $-B$ ou $-M$ est positive.

Remarque 2. En utilisant la formule de changement de base pour les formes bilinéaires ("$M' = {}^tPMP$"), on vérifie facilement les équivalences suivantes pour une forme bilinéaire symétrique $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$:

$$\begin{aligned} B \text{ est positive } &\iff \text{sa matrice dans toute base de } \mathbb{R}^n \text{ est positive} \\ &\iff \text{sa matrice dans la base canonique est positive} \\ &\iff \text{sa matrice dans au moins une base de } \mathbb{R}^n \text{ est positive} \end{aligned}$$

et les mêmes équivalences en remplaçant "positive" par "définie positive".

Le théorème suivant est bien connu et très important :

Théorème 3.4. *Si $M \in M_n(\mathbb{R})$ est une matrice symétrique, alors M est diagonalisable en base orthonormée. La matrice M est positve si et seulement si toutes ses valeurs propres sont positives, et définie positive si et seulement si toutes ses valeurs propres sont strictement positives.*

Démonstration. En identifiant une matrice à une application linéaire, il s'agit de démontrer le résultat suivant : *si E est un espace euclidien et si $M \in \mathcal{L}(E)$ est un endomorphisme* ***auto-adjoint****, i.e. $\langle Mx, y\rangle = \langle x, My\rangle$ pour tous $x, y \in E$, alors il existe une base orthonormée de E formée de vecteurs propres pour M.* On va le faire par récurrence sur $n = \dim(E)$.

Pour $n = 1$, le résultat est évident. Supposons le résultat établi pour tous les espaces euclidiens de dimension $k \leq n$, et soit $M \in \mathcal{L}(E)$ auto-adjoint, avec $\dim(E) = n + 1$. Admettons provisoirement avoir démontré que M possède au moins une valeur propre réelle λ_0, et posons $E_0 = \ker(M - \lambda_0 Id)$. Si $E_0 = E$, alors $M = \lambda_0 Id$ et il n'y a rien à démontrer. Si $E_0 \neq E$, alors $\dim(E_0) < \dim(E)$, et on a aussi $\dim\left(E_0^\perp\right) < \dim(E)$ puisque $E_0 \neq \{0\}$. Le sous-espace E_0 est évidemment stable par M (i.e. $M(E_0) \subset E_0$), et $E_0^\perp$ est également stable par M car M est auto-adjoint (*exercice*). D'après l'hypothèse de récurrence appliquée à E_0 et à $E_0^\perp$, les restrictions de M à E_0 et à $E_0^\perp$ sont diagonalisables en base orthonormée ; et comme $E = E_0 \oplus E_0^\perp$, on obtient une base orthonormée de M en mettant bout à bout deux bases de diagonalisation pour $M_{|E_0}$ et pour $M_{|E_0^\perp}$. Ainsi, on voit qu'il suffit de démontrer que M possède au moins une valeur propre réelle.

Soit $\lambda_0 = \sup\{\langle Mx, x\rangle;\ x \in S_E\}$, où $S_E = \{x \in E;\ \|x\| = 1\}$. Comme S_E est un fermé borné de E, donc un *compact* (car $\dim(E) < \infty$), on peut trouver $x_0 \in S_E$ tel que $\langle Mx_0, x_0\rangle = \lambda_0$. Soit $B : E \times E \to \mathbb{R}$ la forme bilinéaire définie par

$$B(x, y) = \langle \lambda_0 x - Mx, y\rangle = \langle (\lambda_0 Id - M)x, y\rangle .$$

Comme M est autoadjoint, B est symétrique. De plus, B est également *positive* par définition de λ_0. En effet, si $x \in S_E$, alors $B(x, x) = \lambda_0\|x\|^2 - \langle Mx, x\rangle = \lambda_0 - \langle Mx, x\rangle \geq 0$; et on en déduit qu'on a $B(u, u) \geq 0$ pour tout $u \in E$ en appliquant cela à $x = \frac{u}{\|u\|}$. D'après *l'inégalité de Cauchy-Schwarz* appliquée à B, on a donc

$$0 \leq B(x, y)^2 \leq B(x, x)\, B(y, y)$$

pour tous $x, y \in E$. Mais par le choix de x_0, on a $B(x_0, x_0) = \lambda_0 - \langle Mx_0, x_0\rangle = 0$. On en déduit $B(x_0, y) = 0$, autrement dit

$$\langle Mx_0 - \lambda_0, y\rangle = 0$$

pour tout $y \in E$. Ainsi, $Mx_0 - \lambda_0 x_0$ est orthogonal à tous les vecteurs de E, et donc $Mx_0 - \lambda_0 x_0 = 0$. Donc x_0 est un vecteur propre pour M_0 avec pour valeur propre associée λ_0.

Notons $\lambda_1, \dots, \lambda_n$ les valeurs propres de M (répétées selon leur multiplicité), et soit $(f_1, \dots, f_n)$ une base orthonormée de vecteurs propres pour M, avec $Mf_i = \lambda_i e_i$. Si $x = \sum_1^n x_i f_i \in E$, alors $Mx = \sum_1^n \lambda_i x_i f_i$ et donc (par orthonormalité)

$$\langle Mx, x\rangle = \sum_{i=1}^n \lambda_i\, x_i^2\, .$$

On en déduit très facilement que M est positive si et seulement si les valeurs propres λ_i sont toutes positives, et définie positive si et seulement si les λ_i sont toutes strictement positives.

□

Corollaire 3.5. (cas des matrices 2×2)
Une matrice symétrique $M \in M_2(\mathbb{R})$ *est positive si et seulement si* $\det(M) \geq 0$ *et* $\mathrm{tr}(M) \geq 0$*, et définie positive si et seulement si* $\det(M) > 0$ *et* $\mathrm{tr}(M) > 0$*.*

Démonstration. D'après le théorème, il est clair que si M est positive alors $\det(M) \geq 0$ et $\mathrm{tr}(M) \geq 0$, puisque $\det(M)$ est le produit des valeurs propres de M et $\mathrm{tr}(M)$ est la somme des valeurs propres.. Inversement, si $\det(M) \geq 0$ alors les deux valeurs propres de M sont de même signe (au sens large), et si de plus $\mathrm{tr}(M) \geq 0$ ce signe doit être positif. Le raisonnement est identique dans le cas "défini positif". □

Remarque 1. En changeant M en $-M$ et en remarquant que $\det(-M) = \det(M)$ car la dimension est paire (!), on voit qu'une matrice symétrique $M \in M_2(\mathbb{R})$ est négative si et seulement si $\det(M) \geq 0$ et $\mathrm{tr}(M) \leq 0$; et définie négative si et seulement si $\det(M) > 0$ et $\mathrm{tr}(M) < 0$.

Remarque 2. Ces résultats ne sont valables que pour des matrices 2×2 !

Exercice. Trouver une matrice symétrique M de taille 3 qui vérifie $\det(M) > 0$ et $\mathrm{tr}(M) > 0$, mais qui ne soit ni positive ni négative.

Corollaire 3.6. *Soit* $M = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$ *une matrice symétrique* 2×2*. Alors* M *est définie positive si et seulement si* $a > 0$ *et* $\det(M) > 0$ *; et* M *est définie négative si et seulement si* $a < 0$ *et* $\det(M) > 0$*.*

Démonstration. Comme plus haut, le cas "défini négatif" s'obtient en changeant M en $-M$.

Si M est définie positive alors $\det(M) > 0$ et $a = \langle Me_1, e_1\rangle > 0$. Inversement, si $\det(M) > 0$, alors les 2 valeurs propres de M sont non nulles et de même signe, donc M est soit définie positive, soit définie négative. Comme $ac - b^2 > 0$, on voit que a et c sont de même signe, et donc tous les deux strictement positifs puisque $a > 0$. On a donc $\mathrm{tr}(M) = a + c > 0$, et donc M est définie positive. □

Remarque. Ce dernier résultat concerne uniquement les matrices *définies* positives ou négatives. Il n'est *pas vrai* en général qu'une matrice $M = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$ est positive si et seulement si $a \geq 0$ et $\det M \geq 0$. Par exemple, la matrice $M = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}$ n'est pas positive, et pourtant $a = 0$ et $\det(M) = 0$.

L'inconvénient "pratique" du théorème 3.4 est qu'au delà de la dimension 2 ou 3, il est souvent très difficile de déterminer les valeurs propres d'une matrice. La proposition suivante (qui est une généralisation naturelle du corollaire 3.6) donne un moyen de vérifier qu'une matrice symétrique est définie positive sans avoir à trouver ses valeurs propres.

PROPOSITION 3.7. *Soit* $M = (a_{ij})_{i,j=1}^n$ *une matrice symétrique réelle. Pour* $k \in \{1, \dots, n\}$*, posons* $\Delta_k = \det(a_{ij})_{i,j=1}^k$*. Alors* M *est* définie positive *si et seulement si tous les "mineurs principaux"* $\Delta_1, \dots, \Delta_n$ *sont* strictement positifs.

DÉMONSTRATION. Notons B la forme bilinéaire symétrique associée à M, i.e. $B(h, h') = \langle Mh, h'\rangle$. Notons aussi $(e_1, \dots, e_n)$ la base canonique de $\mathbb{R}^n$, et pour $k \in \{1, \dots, n\}$, posons $E_k = \operatorname{vect}(e_1, \dots, e_k)$. Alors la matrice $M_k = (a_{ij})_{i,j=1}^k$ est la matrice de la restriction de B à $E_k \times E_k$ dans la base $(e_1, \dots, e_k)$, et elle est donc définie positive si B est définie positive. Par conséquent, tous les $\Delta_k = \det(M_k)$ doivent être strictement positifs si M est définie positive.

Pour la réciproque, on procède par récurrence sur n, le résultat étant évident pour $n = 1$. Supposons avoir établi la propriété voulue pour un certain $n \geq 1$, et soit $M = (a_{ij})_{i,j=1}^{n+1}$ une matrice symétrique de taille $n+1$ dont tous les mineurs principaux sont strictement positifs. On notera comme plus haut B la forme bilinéaire symétrique associée à M. Soit également $(e_1, \dots, e_{n+1})$ la base canonique de $\mathbb{R}^{n+1}$, et soit $E_n = \operatorname{vect}(e_1, \dots, e_n)$.

On remarque d'abord que la matrice de la restriction de B à $E_n \times E_n$ est $M_n = (a_{ij})_{i,j=1}^n$, donc $B_{|E_n \times E_n}$ est définie positive par hypothèse de récurrence.

Choisissons maintenant un vecteur $f \in \mathbb{R}^{n+1} \setminus \{0\}$ tel que $f \perp \operatorname{vect}(Me_1, \dots, Me_n)$. Alors $B(f, u) = \langle Mf, u\rangle = \langle f, Mu\rangle = 0$ pour tout $u \in E_n$, et comme $B_{|E_n \times E_n}$ est définie positive et $f \neq 0$, on en déduit que $f \notin E_n$. Par conséquent, $(e_1, \dots, e_n, f)$ est une base de $\mathbb{R}^{n+1}$.

Par le choix de f, la matrice de B dans la base $(e_1, \dots, e_n, f)$ est

$$M' = \begin{pmatrix} M_n & 0 \\ 0 & \alpha \end{pmatrix}$$

où $\alpha = \langle Mf, f\rangle$. De plus α est *strictement positif*. En effet, d'après la formule de changement de base pour les formes bilinéaires, on sait qu'il existe une matrice inversible P telle que $M' = {}^tPMP$. On a donc $\det(M') = \det(P)^2 \det(M) = \det(P)^2 \Delta_{n+1} > 0$, d'où $\alpha > 0$ puisque $\det(M') = \alpha \times \det(M_n)$ et $\det(M_n) = \Delta_n > 0$.

Si maintenant h est un vecteur quelconque de $\mathbb{R}^{n+1}$, on peut écrire $h = u + \lambda f$ avec $u \in E_n$ et $\lambda \in \mathbb{R}$, et comme $B(f, u) = 0 = B(u, f)$ on en déduit

$$\begin{aligned} B(h, h) &= B(u, u) + \lambda^2 B(f, f) \\ &= B(u, u) + \lambda^2 \alpha\,. \end{aligned}$$

Comme u ou λ est non-nul si $h \neq 0$ et comme $B_{|E_n \times E_n}$ est définie positive, on voit donc qu'on a $B(h, h) > 0$ si $h \neq 0$. Ainsi, B est définie positive. □

COROLLAIRE 3.8. *Avec les notations précédentes, la matrice* M *est définie négative si et seulement si tous les* Δ_k *pour* k *impair sont strictement négatifs et tous les* Δ_k *pour* k *pair sont strictement positifs.*

DÉMONSTRATION. C'est immédiat en appliquant la proposition à $-M$, puisque $\Delta_k(-M) = (-1)^k \Delta_k(M)$ pour tout k. □

3.3.2. *Extrema et différentielle seconde.* Les deux théorèmes suivants donnent des conditions nécessaires et des conditions suffisantes pour l'existence d'extrema locaux portant sur la différentielle seconde. Ce sont les géralisations naturelles des parties (2) et (3) du lemme 3.1.

THÉORÈME 3.9. *Soit Ω un ouvert de $\mathbb{R}^n$, et soit $f \in \mathcal{C}^2(\Omega)$.*

(1) *Si f possède un minimum local en un point $a \in \Omega$, alors $d^2 f(a)$ est positive ; autrement dit la matrice hessienne $H_f(a)$ est positive.*
(2) *Si f possède un maximum local en un point $a \in \Omega$, alors $H_f(a)$ est négative.*

DÉMONSTRATION. Il suffit évidemment de démontrer (1). Supposons que f possède un minimum local en a, et fixons $h \in \mathbb{R}^n$ quelconque. Alors la fonction φ définie par $\varphi(t) = f(a+th)$ est de classe $\mathcal{C}^2$ au voisinage de 0 avec $\varphi''(t) = d^2 f(a+th)h^{[2]}$ (voir le lemme 5.8 du chapitre 3) et φ possède un minimum local en $t = 0$. D'après le lemme 3.1, on a donc $\varphi''(0) \geq 0$, autrement dit $d^2 f(a)(h,h) \geq 0$. □

THÉORÈME 3.10. *Soit Ω un ouvert de $\mathbb{R}^n$, et soit $f \in \mathcal{C}^2(\Omega)$. Soit également $a \in \Omega$.*

(1) *Si $df(a) = 0$ et si $d^2 f(a)$ est* définie *positive, alors f possède un minimum local en a.*
(2) *Si $df(a) = 0$ et si $d^2 f(a)$ est définie positive, alors f possède un minimum local en a.*

La démonstration utilise le lemme suivant.

LEMME 3.11. *Soit $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ une forme bilinéaire symétrique définie positive, et soit $\|\cdot\|$ une norme quelconque sur $\mathbb{R}^n$. Alors il existe une constante $c > 0$ telle que $\forall h \in \mathbb{R}^n \ : \ B(h,h) \geq c\,\|h\|^2$.*

DÉMONSTRATION. Soit $S = \{h \in \mathbb{R}^n;\ \|h\| = 1\}$. Par définition, S est une partie fermée bornée de $\mathbb{R}^n$, donc un *compact*. Comme la forme bilinéaire B est continue (on est en dimension finie), la fonction $f : S \to \mathbb{R}$ définie par $f(u) = B(u,u)$ est continue et possède donc un minimum global ; et comme $B(u,u) > 0$ pour tout $u \neq 0$, on en déduit que $c = \inf\{B(u,u);\ \|u\| = 1\}$ est *strictement* positif. Si maintenant $h \in \mathbb{R}^n$ est quelconque, $h \neq 0$, on a $B\left(\frac{h}{\|h\|}, \frac{h}{\|h\|}\right) \geq c$ par définition de c, d'où $B(h,h) \geq c\,\|h\|^2$ par bilinéarité ; et ceci est évidemment encore vrai pour $h = 0$. □

PREUVE DU THÉORÈME 3.10. Bien entendu, on se contente de démontrer (1) ; supposons donc que a soit un point critique de f et que $d^2 f(a)$ soit définie positive.

Soit $\|\cdot\|$ une norme quelconque sur $\mathbb{R}^n$. Comme $df(a) = 0$, le développement limité à l'ordre 2 de f au point a s'écrit

$$f(a+h) = f(a) + \frac{1}{2}\, d^2 f(a)h^{[2]} + R(h)\,,$$

où $R(h) = o(\|h\|^2)$.

D'après le lemme 3.11, on peut trouver une constante $c > 0$ telle que $d^2 f(a)h^{[2]} \geq c\,\|h\|^2$ pour tout $h \in \mathbb{R}^n$; et par définition d'un "o", peut choisir $\delta > 0$ tel que $\|h\| < \delta \Rightarrow |R(h)| \leq (c/4)\|h\|^2$. On a alors

$$\frac{1}{2}\, d^2 f(a)h^{[2]} + R(h) \geq \frac{c}{2}\,\|h\|^2 - \frac{c}{4}\|h\|^2 = \frac{c}{4}\,\|h\|^2$$

pour tout h vérifiant $\|h\| < \delta$, et donc

$$\forall u \in B(a,\delta) \ : \ f(u) \geq f(a) + \frac{c}{4}\,\|u-a\|^2 \geq f(a)\,.$$

Ainsi, f possède un minimum local en a. $\square$

Remarque. Sous l'hypothèse de (1), on a en fait montré que f possède un minimum local **strict** au point a, i.e. $f(u) > f(a)$ pour tout $u \neq a$ suffisamment proche de a.

Exercice. Déterminer les extrema locaux de la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par $f(x, y) = x^4 + y^4 - (x + y)^2$.

4. Le cas des fonctions convexes

Les fonctions convexes et concaves d'une variable ont été définies au chapitre 2. En fait, la définition a un sens dans un cadre beaucoup plus général :

DÉFINITION 4.1. *Soit C une partie convexe d'un espace vectoriel E. Une fonction $f : C \to \mathbb{R}$ est dite **convexe** sur C si*

$$\forall u, v \in C\ \forall \lambda \in [0, 1]\ :\ f((1 - \lambda)u + \lambda v) \leq (1 - \lambda)f(u) + \lambda f(v)\,. \tag{4.1}$$

*La fonction f est dite **concave** si $-f$ est convexe ; autrement dit, si l'inégalité est inversée dans (4.1).*

Exemple 1. Soit I un intervalle de $\mathbb{R}$ et soit $\varphi : I \to \mathbb{R}$ une fonction continue sur I et dérivable au moins sur $\mathring{I}$. Alors φ est convexe si et seulement si φ' est croissante. Donc, une fonction φ continue sur I et deux fois dérivable au moins sur $\mathring{I}$ est convexe si et seulement si $\varphi'' \geq 0$.

Cas particuliers. La fonction $t \mapsto t^\alpha$ est convexe sur $[0, \infty[$ si $\alpha \geq 1$, et concave si $\alpha \in [0, 1]$. La fonction exponentielle est convexe sur $\mathbb{R}$ et la fonction logarithme est concave sur $]0, \infty[$.

Exemple 2. Soit E un evn. Pour tout $a \in E$, la fonction $u \mapsto \|u - a\|$ est convexe sur E. En particulier, la fonction $u \mapsto \|u\|$ est convexe.

DÉMONSTRATION. Posons $f(u) = \|u - a\|$. Si $u, v \in E$ et $\lambda \in [0, 1]$ alors, en écrivant $a = (1 - \lambda)a + \lambda a$, on obtient

$$\begin{aligned} f((1 - \lambda)u + \lambda v)) &= \|(1 - \lambda)u + \lambda v - a\| \\ &= \|(1 - \lambda)(u - a) + \lambda(v - a)\| \\ &\leq \|(1 - \lambda)u - a\| + \|\lambda(v - a)\| \\ &= (1 - \lambda)f(u) + \lambda f(v)\,. \end{aligned}$$

$\square$

Les fonctions convexes possèdent des propriétés de "stabilité" intéressantes et utiles, qui se vérifient sans difficulté en écrivant la définition :

Exercice. (propriétés de stabilité)

- La somme de deux fonctions convexes est convexe.
- Si f est une fonction convexe, alors αf est convexe pour tout $\alpha \geq 0$.
- Si $f : C \to \mathbb{R}$ est convexe et si $\varphi : I \to \mathbb{R}$ est une fonction convexe croissante sur un intervalle I contenant $f(C)$, alors $\varphi \circ f$ est convexe.

Le lemme suivant est presque évident, mais très important car il permet de se ramener à étudier des fonctions convexes d'*une variable.*

LEMME 4.2. *Soit C une partie convexe d'un espace vectoriel E. Pour une fonction $f : C \to \mathbb{R}$, les propriétés suivantes sont équivalentes :*

(i) *f est convexe sur C ;*

(ii) *pour tout $x \in C$ et pour tout $h \in E$ tel que $x+h \in C$, la fonction $t \mapsto f(x+th)$ est convexe sur $[0,1]$.*

DÉMONSTRATION. Supposons f convexe et fixons x, h comme dans (ii). Notons $\varphi_{x,h}$ la fonction $t \mapsto f(x+th)$. On a $\varphi_{x,h}(t) = f(\gamma(t))$, où $\gamma(t) = a + th$ est une fonction **affine**, i.e. vérifiant $\gamma((1-\lambda)s + \lambda t) = (1-\lambda)\gamma(s) + \lambda\gamma(t)$ pour $s, t \in [0,1]$ et $0 \leq \lambda \leq 1$ (*exercice*). On a donc

$$\begin{aligned} \varphi_{x,h}((1-\lambda)s + \lambda t) &= f((1-\lambda)\gamma(s) + \lambda\gamma(t)) \\ &\leq (1-\lambda)f(\gamma(s)) + \lambda f(\gamma(t)) \\ &= (1-\lambda)\varphi_{x,h}(s) + \lambda\varphi_{x,h}(t) \end{aligned}$$

pour tous $s, t \in [0,1]$ et $0 \leq \lambda \leq 1$, ce qui prouve que $\varphi_{x,h}$ est convexe.

Inversement, en appliquant (ii) avec $x = u$ et $h = v - u$ et en écrivant $\lambda = (1-\lambda) \times 0 + \lambda \times 1$, on obtient

$$\begin{aligned} f((1-\lambda)u + \lambda v) &= f(u + \lambda(v-u)) \\ &= \varphi_{u,v-u}(\lambda) \\ &\leq (1-\lambda)\varphi_{u,v-u}(0) + \lambda\varphi_{u,v-u}(1) \\ &= (1-\lambda)f(u) + \lambda f(v) \end{aligned}$$

pour tout $\lambda \in [0,1]$, d'où la convexité de f.

□

En appliquant ce lemme et les critères de convexité pour les fonctions d'une variable rappelés dans l'exemple 1, on obtient les résultats suivants.

PROPOSITION 4.3. *Soit Ω un ouvert convexe d'un evn E, et soit $f : \Omega \to \mathbb{R}$.*

(1) *On suppose que f est différentiable sur Ω. Alors f est convexe sur Ω si et seulement si*

$$\forall u, v \in \Omega \ : \ (df(v) - df(u))(v-u) \geq 0\,. \tag{4.2}$$

(2) *On suppose que Ω est un ouvert de $\mathbb{R}^n$ et que f est de classe $\mathcal{C}^2$. Alors f est convexe si et seulement si $d^2f(u)$ est positive pour tout $u \in \Omega$.*

Remarque. Si Ω est un ouvert de $\mathbb{R}^n$, la condition (4.2) s'écrit

$$\langle \nabla f(v) - \nabla f(u), v - u \rangle \geq 0\,.$$

PREUVE DE LA PROPOSITION. Pour $x \in \Omega$ et $h \in E$ tel que $x + h \in \Omega$, on notera $\varphi_{x,h} : [0,1] \to \mathbb{R}$ la fonction définie par $\varphi_{x,h}(t) = f(x+th)$. D'après le lemme 4.2, f est convexe si et seulement si toutes les fonctions $\varphi_{x,h}$ le sont.

(1) Comme f est différentiable, les fonctions $\varphi_{x,h}$ sont dérivables sur $[0,1]$ avec $\varphi'_{x,h}(t) = df(x+th)h$. Si f est convexe, alors les $\varphi'_{x,h}$ sont croissantes et donc $\varphi'_{x,h}(1) \geq \varphi'_{x,h}(0)$, autrement dit $(df(x+h) - df(x))h \geq 0$ pour tous x, h tels que $x, x+h \in \Omega$. En prenant $x = u$ et $h = v - u$, on obtient donc (4.2). Inversement, supposons (4.2) vérifiée. Si $s, t \in [0,1]$ alors, en appliquant (4.2) avec $u = x + sh$ et $v = x + th$, on obtient $(df(x+th) - df(x+sh))((t-s)h) \geq 0$, autrement dit

$$(t-s) \times (\varphi'_{x,h}(t) - \varphi'_{x,h}(s)) \geq 0\,,$$

pour tous $s, t \in [0,1]$. Ainsi, $\varphi'_{x,h}(t) - \varphi'_{x,h}(s)$ est toujours du signe de $t - s$, ce qui signifie exactement que toutes les fonctions $\varphi'_{x,h}$ sont croissantes, et donc que f est convexe.

(2) Comme f est $\mathcal{C}^2$, toutes les fonctions $\varphi_{x,h}$ sont de classe $\mathcal{C}^2$, avec $\varphi''_{x,h}(t) = d^2 f(x + th)h^{[2]}$. On en déduit immédiatement que si $d^2 f(u)$ est positive pour tout $u \in \Omega$, alors toutes les fonctions $\varphi_{x,h}$ sont convexes. Inversement supposons que toutes les $\varphi_{x,h}$ soient convexes. Si $u \in \Omega$ et si $h \in \mathbb{R}^n$ est quelconque, on peut trouver $\delta > 0$ tel que $u+\delta h \in \Omega$. Comme $\varphi_{u,\delta h}$ est convexe, on a $d^2 f(u)(\delta h)^{[2]} = \varphi''_{u,\delta h}(0) \geq 0$, autrement dit $\delta^2 \times d^2 f(u)h^{[2]} \geq 0$ et donc $d^2 f(u)h^{[2]} \geq 0$. Comme $h \in \mathbb{R}^n$ est quelconque, cela montre que $d^2 f(u)$ est positive, pour tout $u \in \Omega$. □

Ce qu'on a raconté pour le moment sur les fonctions convexes n'a pas grand chose à voir avec les extrema. Ce n'est pas le cas de la proposition suivante.

PROPOSITION 4.4. *Soit Ω un ouvert convexe d'un evn E et soit $f : \Omega \to \mathbb{R}$ une fonction convexe différentiable. Alors, pour tout $a \in \Omega$ et pour tout $h \in E$ tel que $a + h \in \Omega$, on a*

$$f(a + h) \geq f(a) + df(a)h\,. \tag{4.3}$$

DÉMONSTRATION. Comme toujours, on regarde la fonction $\varphi = \varphi_{a,h} : [0,1] \to \mathbb{R}$ définie par $\varphi(t) = f(a + th)$. D'après le lemme 4.2, φ est convexe et dérivable, avec $\varphi'(t) = df(a + th)h$. Le graphe de φ est donc "au dessus de toutes ses tangentes" ; et en particulier, on a $\varphi(1) - \varphi(0) \geq \varphi'(0) \times (1 - 0)$, autrement dit $f(a + h) - f(a) \geq \varphi'(0) = df(a)h$. □

COROLLAIRE 4.5. *Si $f : \Omega \to \mathbb{R}$ est convexe différentiable et si $a \in \Omega$ est un point critique de f, alors f possède un minimum global en a.*

DÉMONSTRATION. C'est évident d'après la proposition. □

Exercice 1. La réciproque de la proposition 4.4 est-elle vraie ? Autrement dit, la validité de (4.3) entraîne-t-elle la convexité de f ?

Exercice 2. Que peut-on dire d'une fonction convexe $f : \Omega \to \mathbb{R}$ admettant un *maximum* local en un point de Ω ?

Pour conclure cette section, voici un "exemple récapitulatif".

Exemple. Soit E un espace **préhilbertien**, i.e. un evn dont la norme dérive d'un produit scalaire (on ne suppose pas que E est de dimension finie). Si $a_1, \dots, a_n \in E$, il existe un unique point $u \in E$ minimisant la quantité $\|u - a_1\|^2 + \cdots + \|u - a_n\|^2$: c'est l'isobarycentre des points $a_1, \dots, a_n$.

DÉMONSTRATION. La fonction $f : E \to \mathbb{R}$ définie par $f(u) = \sum_1^n \|u - a_i\|^2$ est convexe car c'est une somme de carrés de fonctions convexes (cf les propriétés de stabilité). De plus, f est différentiable sur E avec

$$df(u)h = 2\sum_{i=1}^{n} \langle u - a_i, h\rangle\,.$$

On a donc $df(u) = 0$ si et seulement si $\sum_1^n \langle u - a_i, h\rangle = 0$ pour tout $h \in E$, autrement dit $\sum_1^n (u - a_i) = 0$. Cette équation possède une unique solution :

$$u = \frac{1}{n}\sum_{i=1}^{n} a_i\,,$$

c'est-à-dire l'isobarycentre des points $a_1, \dots, a_n$. Ainsi, f possède un unique point critique, et d'après le corollaire précédent on sait que f admet un minimum global en u. □

5. Extrema liés

Dans cette section, on considère des problèmes du type suivant. Soit $f : \mathbb{R}^n \to \mathbb{R}$, et soient $C_1, \dots, C_m$ des fonctions à valeurs réelles définies sur $\mathbb{R}^n$. Le problème consiste à maximiser ou minimiser $f(x)$ "sous les contraintes $C_1(x) = 0, \dots, C_m(x) = 0$". Autrement dit, en posant $\Sigma := \{x \in \mathbb{R}^n;\ C_1(x) = \cdots = C_m(x) = 0\}$, il s'agit de déterminer les extrema de la fonction $f_{|\Sigma}$, restriction de f à l'ensemble Σ.

5.1. Le "théorème des extrema liés", et quelques exemples. Bien que cela ne soit pas immédiatement apparent, le résultat suivant est l'analogue du théorème 3.2.

THÉORÈME 5.1. (théorème des extrema liés)
Soit Ω un ouvert de $\mathbb{R}^n$, et soient $C_1, \dots, C_m$ des fonctions à valeurs réelles définies sur $\overline{\Omega}$, de classe $\mathcal{C}^1$ dans Ω. On pose

$$\Sigma = \{x \in \overline{\Omega};\ C_1(x) = \cdots = C_m(x) = 0\}\,.$$

Soit également $f : \mathbb{R}^n \to \mathbb{R}$, différentiable en tout point de Ω. Enfin, soit $a \in \Sigma$. On suppose que

- *$f_{|\Sigma}$ possède un extremum local en a* ***et de plus*** *$a \in \Omega$;*
- *les vecteurs $\nabla C_1(a), \dots, \nabla C_m(a)$ sont* linéairement indépendants.

Alors $\nabla f(a)$ est combinaison linéaires de $\nabla C_1(a), \dots, \nabla C_m(a)$. Autrement dit, il existe m nombres réels $\lambda_1, \dots, \lambda_m$, qu'on appelle des ***multiplicateurs de Lagrange****, tels que*

$$\nabla f(a) = \lambda_1 \nabla C_1(a) + \cdots + \lambda_m \nabla C_m(a)\,;$$

ce qui s'écrit encore

$$df(a) = \sum_{i=1}^{n} \lambda_i dC_i(a)\,.$$

Lorsqu'il n'y a qu'une seule "fonction contrainte" C, le théorème prend la forme suivante :

COROLLAIRE 5.2. *Soit Ω un ouvert de $\mathbb{R}^n$, et soit $C : \overline{\Omega} \to \mathbb{R}$, de classe $\mathcal{C}^1$ dans Ω. On pose $\Sigma = \{x \in \overline{\Omega};\ C(x) = 0\}$. Soit également $f : \mathbb{R}^n \to \mathbb{R}$ différentiable dans Ω, et soit $a \in \Sigma$. On suppose que*

- *$f_{|\Sigma}$ possède un extremum local en a et $a \in \Omega$;*
- *$\frac{\partial C}{\partial x_j}(a) \neq 0$ pour au moins un $j \in \{1, \dots, n\}$.*

Alors il existe un nombre réel λ tel que $df(a) = \lambda\, dC(a)$, ce qui s'écrit

$$\begin{cases} \frac{\partial f}{\partial x_1}(a) = \lambda \frac{\partial C}{\partial x_1}(a) \\ \quad\vdots \\ \frac{\partial f}{\partial x_1}(a) = \lambda \frac{\partial C}{\partial x_1}(a) \end{cases}$$

Remarque. Dans le théorème des extrema liés, il suffit évidemment de supposer que f est définie sur $\overline{\Omega}$.

On va donner plus bas deux démonstrations du théorème des extrema liés. Pour le moment, voici quelques exemples d'utilisation.

Exemple 1. On veut construire une boite en carton parallèlépipèdique sans couvercle de volume 4 litres, en utilisant le moins de carton possible. Il s'agit de déterminer les dimensions de la boite.

Notons x, y et z les dimensions de la boite, exprimés en dm (ce qui est naturel puisqu'un litre vaut 1dm^3), z étant la “hauteur”. Bien entendu x, y et z doivent être positifs. Comme la boite n'a pas de couvercle, la surface de carton utilisé est égale à $xy + 2xz + 2yz$. Le volume vaut xyz, et doit être égal à 4. Il s'agit donc de minimiser la fonction $f(x, y, z) = xy + 2xz + 2yz$ sur l'ensemble

$$\Sigma = \{(x, y, z) \in \mathbb{R}^3;\ x \geq 0, y \geq 0, y \geq 0 \text{ et } xyz = 4\}.$$

En posant $\Omega = \{(x, y, z);\ x > 0, y > 0, z > 0\}$ et $C(x, y, z) = xyz - 4$, on a $\Sigma = \{(x, y, z) \in \overline{\Omega};\ C(x, y, z) = 0\}$. De plus, tous les points de Σ appartiennent en fait à Ω car la contrainte impose $xyz \neq 0$.

Admettons provisoirement que la fonction f possède un minimum global sur Σ, atteint en un point (a, b, c). Comme $(a, b, c) \in \Omega$ et $\nabla C(a, b, c) = \left(\begin{smallmatrix} bc \\ ac \\ ab \end{smallmatrix}\right) \neq 0$, on peut appliquer le théorème des extrema liés : il existe un multiplicateur λ tel que $\nabla f(a, b, c) = \lambda\, \nabla C(a, b, c)$, ce qui s'écrit

$$\begin{cases} b + 2c = \lambda\, bc \\ a + 2c = \lambda\, ac \\ 2a + 2b = \lambda\, ab \end{cases}$$

En multipliant la première équation par a, la deuxième par b et la troisième par c, on obtient $\lambda\, abc = ab + 2ac = ab + 2bc = 2ac + 2bc$. On en tire $ac = bc$ et $ab = 2ac$, autrement dit $a = b$ et $b = 2c$. Comme de plus $abc = 4$, on en déduit $c^3 = 1$, d'où finalement $c = 1$ et $a = b = 2$. Ainsi la boite a une base carrée de 20cm de côté et une hauteur de 10cm.

Montrons maintenent que la fonction f possède effectivement un minimum global sur Σ. D'après le théorème 2.1, il suffit de vérifier que pour tout $\alpha \in \mathbb{R}$, l'ensemble $\{f \leq \alpha\} = \{(x, y, z) \in \Sigma;\ f(x, y, z) \leq \alpha\}$ est compact. On a visiblement $\{f \leq \alpha\} = \emptyset$ si $\alpha \leq 0$, donc on peut suposer $\alpha > 0$. Comme Σ est un fermé de $\mathbb{R}^3$ et comme f est continue, $\{f \leq \alpha\}$ est un fermé de $\mathbb{R}^3$. Si $(x, y, z) \in \{f \leq \alpha\}$, alors $xy \leq \alpha$ et $xyz = 4$, donc $z \geq 4/\alpha$; mais on a aussi $2xz \leq \alpha$, donc $x \leq \alpha^2/8$ (et bien sûr $x \geq 0$). De même, on trouve $0 \leq y \leq \alpha^2/8$ et $0 \leq z \leq \alpha^2/16$. Ainsi $\{f \leq \alpha\}$ est une partie *bornée* de $\mathbb{R}^3$. Au total, $\{f \leq \alpha\}$ est bien compact.

Exemple 2. Diagonalisation des matrices symétriques réelles.

Soit $M \in M_n(\mathbb{R})$ une matrice symétrique, et soit $\Sigma = \{x \in \mathbb{R}^n;\ \|x\| = 1\}$, où $\|\cdot\|$ est la norme euclidienne. Comme Σ est un fermé borné de $\mathbb{R}^n$, donc un compact, on peut trouver un point $a \in \Sigma$ qui maximise $f(x) = \langle Mx, x\rangle$ sur Σ. Comme $\Sigma = \{x \in \mathbb{R}^n;\ C(x) = 0\}$ avec $C(x) = \|x\|^2 - 1$ et comme $\nabla C(a) = 2a \neq 0$ (car $\|a\| = 1$), on peut appliquer le théorème des extrema liés à f et C (avec $\Omega = \mathbb{R}^n$) : il existe un “multiplicateur” $\lambda \in \mathbb{R}$ tel que $\nabla f(a) = \lambda \nabla C(a) = 2\lambda a$. Mais $df(x)h = \langle Mx, h\rangle + \langle Mh, x\rangle = 2\langle Mx, h\rangle$, donc $\nabla f(x) = 2Mx$ pour tout $x \in \mathbb{R}^n$. On obtient donc $Ma = \lambda a$, ce qui prouve que λ est une valeur propre de M et a un vecteur propre associé. Ainsi, on vient de retrouver le fait que *toute matrice symétrique $M \in M_n(\mathbb{R})$ possède au moins une valeur propre réelle.* Et on sait que de là, il n'est pas difficile de montrer que toute matrice symétrique est en fait diagonalisable en base orthonormée.

Exemple 3. Inégalité de Hölder.

Soient p et q deux nombre réels strictement plus grands que 1 et tels que $\frac{1}{p}+\frac{1}{q}=1$. On va utiliser le théorème des extrema liés pour montrer que si $a_1,\dots,a_n$ et $b_1,\dots,b_n$ sont des nombres réels strictement positifs, alors

$$\sum_{i=1}^n a_ib_i \leq \left(\sum_{i=1}^n a_i^p\right)^{1/p} \times \left(\sum_{i=1}^n b_i^q\right)^{1/q} .$$

Fixons $a_1,\dots,a_n>0$, posons

$$\Sigma = \left\{(x_1,\dots,x_n)\in\mathbb{R}^n;\ x_1\geq 0,\dots,x_n\geq 0 \text{ et} \sum_{i=1}^n x_i^q = 1\right\},$$

et notons $f:\mathbb{R}^n\to\mathbb{R}$ la fonction définie par

$$f(x_1,\dots,x_n)=\sum_{i=1}^n a_ix_i\,.$$

L'ensemble Σ est visiblement un fermé de $\mathbb{R}^n$, et il est également borné car $\|x\|_\infty \leq 1$ pour tout $x\in\Sigma$. Donc Σ est compact, et par conséquent la fonction f, qui est continue, possède un maximum global sur Σ. Dans la suite, on posera

$$M=\max_\Sigma f = \max\left\{\sum_{i=1}^n a_ix_i;\ x_i\geq 0,\ \sum_{i=1}^n x_i^q=1\right\}.$$

Soit $\xi=(\xi_1,\dots,\xi_n)$ un point de Σ tel que $f(\xi)=M$, et admettons provisoirement que toutes les coordonnées de ξ sont *strictement* positives. On peut alors appliquer le théorème des extrema liés avec

$$\Omega=\{(x_1,\dots,x_n)\in\mathbb{R}^n;\ x_1>0,\dots,x_n>0\}$$

et $C(x_1,\dots,x_n)=\sum_1^n x_i^q-1$, car $\Sigma=\{x\in\overline{\Omega};\ C(x_1,\dots,x_n)=0\}$ et $\xi\in\Omega$. Il existe ainsi un multiplicateur de Lagrange λ tel que $\nabla f(\xi_1,\dots,\xi_n)=\lambda\nabla C(\xi_1,\dots,\xi_n)$, ce qui s'écrit

$$\begin{cases} a_1=\lambda\, q\xi_1^{q-1} \\ \qquad\vdots \\ a_n=\lambda\, q\xi_1^{q-1}\end{cases}$$

En posant $\mu=\lambda q$, on a $\xi=\left(\frac{a_i}{\mu}\right)^{\frac{1}{q-1}}$ pour tout $i\in\{1,\dots,n\}$, et comme $\sum_i \xi_i^q=1$ on en déduit $\sum_i\left(\frac{a_i}{\mu}\right)^{\frac{q}{q-1}}=1$. Comme de plus $\frac{q}{q-1}=p$, on en tire $\mu=(\sum_i a_i^p)^{1/p}$.

D'autre part, les équations précédentes donnent aussi $a_i\xi_i=\mu\,\xi_i^q$ pour tout i et donc $\sum_i a_i\xi_i=\mu\sum_i\xi_i^q$, autrement dit $M=\mu$. Au total, on obtient ainsi

$$M=\left(\sum_{i=1}^n a_i^p\right)^{1/p}.$$

Cela signifie que pour des nombres réels positifs $x_1,\dots,x_n$, l'implication suivante a lieu :

$$\sum_{i=1}^n x_i^q=1 \implies \sum_{i=1}^n a_ix_i \leq \left(\sum_{i=1}^n a_i^p\right)^{1/p}. \tag{5.1}$$

Si maintenant $b_1, \dots, b_n$ sont des réels strictement positifs quelconques et si on applique (5.1) avec $x_i = \frac{b_i}{(\sum_{j=1}^n b_j^q)^{1/q}}$, on obtient le résultat souhaité :

$$\sum_{i=1}^n a_i b_i \leq \left(\sum_{i=1}^n a_i^p\right)^{1/p} \times \left(\sum_{j=1}^n b_j^q\right)^{1/q} .$$

Pour que la preuve de l'inégalité de Hölder soit complète, il reste à vérifier que le point ξ introduit plus haut a toutes ses coordonnées strictement positives.

Supposons par exemple qu'on ait $\xi_1 = 0$. Comme $\xi \neq 0$, on peut également supposer, quitte à permuter les coordonnées, qu'on a $\xi_2 > 0$. Pour $\varepsilon \in]0, \xi_2[$, on peut alors introduire le point

$$x(\varepsilon) = (\varepsilon, (\xi_2^q - \varepsilon^q)^{1/q}, \xi_3, \dots, \xi_n) .$$

Il est clair que $x(\varepsilon)$ appartient à Σ. De plus, comme $(\xi_2^q - \varepsilon^q)^{1/q} = \xi_2 \left(1 - \frac{\varepsilon^q}{\xi_2^q}\right)^{1/q} = \xi_2 + 0(\varepsilon^q)$ quand $\varepsilon \to 0^+$, on a

$$\begin{aligned} f(x(\varepsilon)) &= a_1\varepsilon + a_2(\xi_2 + O(\varepsilon^q)) + \sum_{i=3}^n a_i \xi_i \\ &= a_1 \varepsilon + \sum_{i=2}^n a_i \xi_i + 0(\varepsilon^q) \\ &= f(\xi) + a_1 \varepsilon + 0(\varepsilon^q) , \end{aligned}$$

où on a utilisé le fait que $\xi_1 = 0$.

Comme $q > 1$ on a $\varepsilon^q = o(\varepsilon)$, et on en déduit que si $\varepsilon > 0$ est assez petit, alors $f(x(\varepsilon)) > f(\xi)$. Mais ceci contredit le choix de ξ.

5.2. Première démonstration du théorème 5.1. On donne ici la preuve "classique" du théorème des extrema liés, qui repose sur le *théorème des fonctions implicites* (voir le chapitre 5).

On aura besoin de la définition suivante.

DÉFINITION 5.3. *Soit $M \subset \mathbb{R}^n$, et soit $a \in V$. On dit qu'un vecteur $h \in \mathbb{R}^n$ est **tangent à M au point** a s'il existe $\delta > 0$ et une application $\gamma :]-\delta, \delta[\to \mathbb{R}^n$ de classe $\mathcal{C}^1$ telle que*

(i) *$\gamma(t) \in M$ pour tout $t \in]-\delta, \delta[$;*
(ii) *$\gamma(0) = a$ et $\gamma'(0) = h$.*

On note $T_a(M)$ l'ensemble des vecteurs tangents à M au point a.

Exemple. Si le point a est intérieur à M, alors $T_a(M) = \mathbb{R}^n$.

DÉMONSTRATION. Supposons $a \in \mathring{M}$, fixons une norme sur $\mathbb{R}^n$ et choisissons $r > 0$ tel que $B(a, r) \subset M$. Si $h \in \mathbb{R}^n$ est quelconque, on peut trouver $\delta > 0$ tel que $\delta \times \|h\| < r$, et on a alors $a + th \in M$ pour tout $t \in]-\delta, \delta[$. Il suffit donc de poser $\gamma(t) = a + th$ pour voir que $h \in T_a(M)$. □

Le lemme suivant est la généralisation naturelle du théorème 3.2.

LEMME 5.4. *Soit $f : \mathbb{R}^n \to \mathbb{R}$, et soit $M \subset \mathbb{R}^n$. Si $f_{|M}$ possède un extremum local en un point $a \in M$ et si f est différentiable en a, alors $df(a)h = 0$ pour tout $h \in T_a(M)$.*

Démonstration. Soit $h \in T_a(M)$ quelconque, et choisissons $\gamma :]-\delta;\delta[\to \mathbb{R}^n$ comme dans la définitions 5.3. Alors la fonction φ définie par $\varphi(t) = f(\gamma(t))$ est dérivable en $t=0$ avec $\varphi'(0) = df(\gamma(0))\gamma'(0) = df(a)h$, et φ possède un extremum local en 0. On a donc $\varphi'(0) = 0$, autrement dit $df(a)h = 0$. □

Pour utiliser ce lemme, on a besoin d'un résultat donnant une autre description de l'espace tangent. La preuve est tout-à-fait non-triviale car elle repose sur le théorème des fonctions implicites.

Lemme 5.5. *Soit Ω un ouvert de $\mathbb{R}^n$, et soit $\mathbf{C} : \overline{\Omega} \to \mathbb{R}^m$, de classe $\mathcal{C}^1$ dans Ω. On pose $\Sigma = \{x \in \overline{\Omega};\ \mathbf{C}(x) = 0\}$. Si $a \in \Omega$ et si $D\mathbf{C}(a) : \mathbb{R}^n \to \mathbb{R}^m$ est* surjective*, alors $T_a(\Sigma) = \ker D\mathbf{C}(a)$.*

Démonstration. (i) Soit $h \in T_a(\Sigma)$ et soit $\gamma :]-\delta;\delta[\to \Sigma$ de classe $\mathcal{C}^1$ telle que $\gamma(0) = a$ et $\gamma'(0) = h$. Par définition de Σ, on a $\mathbf{C}(\gamma(t)) \equiv 0$; d'où en dérivant : $D\mathbf{C}(\gamma(t))\gamma'(t) = 0$. Donc $D\mathbf{C}(a)h = D\mathbf{C}(\gamma(0))\gamma'(0) = 0$.

(ii) Inversement, fixons un vecteur $h \in \ker D\mathbf{C}(a)$. Il s'agit de trouver une “courbe” $\gamma :]-\delta;\delta[\to \Sigma$ telle que $\gamma(0) = a$ et $\gamma'(0) = h$.

Remarquons que l'hypothèse faite sur $D\mathbf{C}(a)$ oblige à avoir $m \leq n$. Soit E un supplémentaire de $\ker D\mathbf{C}(a)$ dans $\mathbb{R}^n$, i.e.

$$\mathbb{R}^n = \ker D\mathbf{C}(a) \oplus E\,.$$

On identifie $\mathbb{R}^n$ à $\ker D\mathbf{C}(a) \times E$, et on écrit tout vecteur $x \in \mathbb{R}^n$ sous la forme

$$x = (\lambda, u) = \lambda \oplus u\,.$$

En particulier, le point a s'écrit $a = (\lambda_0, u_0)$. Avec ces notations, on a

$$\Sigma = \{(\lambda, u) \in \overline{\Omega};\ \mathbf{C}(\lambda, u) = 0\}\,. \tag{5.2}$$

Par définition de E et comme $D\mathbf{C}(a)$ est surjective, la restriction de $D\mathbf{C}(a)$ à E est un isomorphisme de E sur $\mathbb{R}^m$; autrement dit, la “différentielle partielle” $D_u\mathbf{C}(\lambda_0, u_0) : E \to \mathbb{R}^m$ est un isomorphisme de E sur $E' = \mathbb{R}^m$ (voir le lemme 4.6 du chapitre 6). D'après le théorème des fonctions implicites, l'équation $\mathbf{C}(\lambda, u) = 0$ définit donc implicitement u comme fonction de λ au voisinage de $a = (\lambda_0, u_0)$. Comme de plus $(\lambda_0, u_0) \in \Omega$, on peut donc trouver un voisinage V de λ_0 dans $\ker D\mathbf{C}(a)$ et une application $\lambda \mapsto u(\lambda)$ de classe $\mathcal{C}^1$ sur V telle que $u(\lambda_0) = u_0$ et

$$\forall \lambda \in V \quad : \quad (\lambda, u(\lambda)) \in \Omega \ \text{ et } \ \mathbf{C}(\lambda, u(\lambda)) = 0\,. \tag{5.3}$$

Comme $h \in \ker D\mathbf{C}(a)$ on peut choisir $\delta > 0$ tel que $\lambda_0 + th \in V$ pour tout $t \in]-\delta,\delta[$. On peut alors définir $\gamma :]-\delta,\delta[\to \mathbb{R}^n = \ker D\mathbf{C}(a) \times E$ par

$$\gamma(t) = (\lambda_0 + th, u(\lambda_0 + th))\,.$$

Il est clair que γ est de classe $\mathcal{C}^1$, avec $\gamma(0) = (\lambda_0, u_0) = a$. De plus, d'après (5.2) et (5.3) on a $\gamma(t) \in \Sigma$ pour tout $t \in]-\delta,\delta[$. Pour terminer la démonstration, il suffit donc de vérifier que $\gamma'(0) = h$.

Par définition de γ, on voit que $\gamma'(0) = (h, \tau) = h \oplus \tau$ pour un certain $\tau \in E$ (à savoir $\tau = DU(\lambda_0)h$). Mais d'autre part, on a aussi $\gamma'(0) \in \ker D\mathbf{C}(a)$ d'après la première partie de la démonstration, puisque $\gamma'(0) \in T_a(\Sigma)$. Par conséquent, on doit avoir $\tau = 0$ et $\gamma'(0) = h \oplus 0 = h$.

□

Preuve du Théorème 5.1. Soit $\mathbf{C} : \mathbb{R}^n \to \mathbb{R}^m$ l'application définie par

$$\mathbf{C}(u) = \begin{pmatrix} C_1(u) \\ \vdots \\ C_m(u) \end{pmatrix}.$$

Par définition, les lignes de la matrice jacobienne de $\mathbf{C}$ au point a sont les (transposées des) vecteurs $\nabla C_1(a), \dots, \nabla C_m(a)$, et elles sont donc linéairement indépendantes. Par conséquent, la matrice $\mathrm{Jac}_{\mathbf{C}}(a)$ est de rang m, i.e. $D\mathbf{C}(a) : \mathbb{R}^n \to \mathbb{R}^m$ est surjective. D'après les deux lemmes précédents, on a donc $\ker D\mathbf{C}(a) = T_a(\Sigma) \subset \ker df(a)$. Autrement dit, en posant $Z = \mathrm{vect}(\nabla C_1(a), \dots, \nabla C_m(a)) \subset \mathbb{R}^n$, on a $Z^\perp \subset \nabla f(a)^\perp$, i.e. $\nabla f(a) \in (Z^\perp)^\perp = Z$.

$\square$

Remarque 5.6. Le théorème des extrema liés peut s'énoncer comme suit. *Soit Ω un ouvert de $\mathbb{R}^n$ et soit $\mathbf{C} : \overline{\Omega} \to \mathbb{R}^m$ de classe $\mathcal{C}^1$ dans Ω. On pose $\Sigma = \{u \in \overline{\Omega};\ \mathbf{C}(u) = 0\}$. Si $f : \overline{\Omega} \to \mathbb{R}$ est de classe $\mathcal{C}^1$ dans Ω si $f_{|\Sigma}$ possède un extremum local en un point $a \in \Sigma \cap \Omega$ tel $D\mathbf{C}(a) : \mathbb{R}^n \to \mathbb{R}^m$ est surjective, alors il existe une forme linéaire (continue) $\Lambda : \mathbb{R}^m \to \mathbb{R}$ telle que $df(a) = \Lambda \circ D\mathbf{C}(a)$.* Sous cette forme, le théorème est en fait vrai en remplaçant $\mathbb{R}^n$ et $\mathbb{R}^m$ par des espaces de Banach E et F quelconques (éventuellement de dimension infinie), *à condition de supposer* que $\ker D\mathbf{C}(a)$ possède un supplémentaire fermé dans E. La démonstration est à peu près identique à celle qui a été donnée pour $E = \mathbb{R}^n$ et $F = \mathbb{R}^m$.

5.3. Deuxième démonstration du Théorème 5.1. On va maintenant donner une preuve du théorème des extrema liés qui n'utilise pas le théorème des fonctions implicites. Cette preuve a l'avantage d'être plus "élémentaire" que la première ; mais elle utilise de façon essentielle la dimension finie, et ne permet donc pas d'obtenir l'énoncé plus général de la Remarque 5.6.

Dans ce qui suit, on garde les notations du Théorème 5.1, et on suppose que $f_{|\Sigma}$ possède un *minimum* local en a. Par ailleurs, on note $\|\cdot\|$ la norme euclidienne sur $\mathbb{R}^n$, et on choisit $r_0 > 0$ tel que $\overline{B}(a, r_0) \subseteq \Omega$ et

$$\forall x \in \overline{B}(a, r_0) \cap \Sigma \ :\ f(a) \le f(x)\,. \tag{5.4}$$

On aura besoin de deux lemmes.

Lemme 5.7. *Soit $\Phi : \Omega \to \mathbb{R}$ une fonction continue, positive, telle que $\Phi(a) = 0$ et $\Phi(x) > 0$ en tout point de $\Omega \setminus \Sigma$. Pour $k \in \mathbb{N}$, posons*

$$\Psi_k(x) := f(x) + \|x - a\|^2 + k\,\Phi(x)\,.$$

Choisissons pour tout k un point $x_k \in \overline{B}(a, r_0)$ tel que Ψ_k atteigne son minimum sur $\overline{B}(a, r_0)$ au point x_k (ce qui est possible par compacité *de $\overline{B}(a, r_0)$). Alors $x_k \to a$ quand $k \to \infty$.*

Démonstration. Comme (x_k) vit dans le compact $\overline{B}(a, r_0)$, il suffit de montrer que la seule valeur d'adhérence possible pour (x_k) est le point a. Fixons une valeur d'adhérence b de (x_k) et une sous-suite (x_{k_i}) telle que $x_{k_i} \to b$.

Remarquons d'abord que $\Psi_{k_i}(a) = f(a)$ pour tout i (car $\Phi(a) = 0$) et donc, par définition de x_{k_i} :

$$f(x_{k_i}) + \|x_{k_i} - a\|^2 + k_i\,\Phi(x_{k_i}) \le f(a)\,. \tag{5.5}$$

Comme Φ est continue, $\Phi(x_{k_i})$ tend vers $\Phi(b)$, qui est ≥ 0. On ne peut pas avoir $\Phi(b) > 0$, car dans ce cas $k_i \Phi(x_{k_i})$ tendrait vers $+\infty$ et on obtiendrait une contradiction en passant à la limite dans (5.5). Donc $\Phi(b) = 0$, et donc $b \in \Sigma$ par hypothèse sur Φ. Comme $b \in \overline{B}(a, r_0)$, on en déduit

$$f(a) \leq f(b) .$$

Mais d'après (5.5) (et comme $\Phi(x_{k_i}) \geq 0$), on a aussi $f(x_{k_i}) + \|x_{k_i} - a\|^2 \leq f(a)$ pour tout i, d'où en passant à la limite :

$$f(b) + \|b - a\|^2 \leq f(a)$$

Ainsi $f(b) + \|b - a\|^2 \leq f(a) \leq f(b)$, et donc $\|b - a\| = 0$. □

On utilisera ce lemme *via* la conséquence suivante.

Corollaire 5.8. *Avec les notations du Lemme 5.7, supposons de plus que la fonction Φ soit* différentiable *sur Ω (de sorte que les Ψ_N le sont également). Alors il existe une suite (x_k) tendant vers a telle que $\nabla \Psi_k(x_k) = 0$ pour tout N.*

Démonstration. La suite (x_k) donnée par le lemme tend vers a, donc x_k appartient à la boule *ouverte* $B(a, r_0)$ pour k assez grand. Ainsi, Ψ_k possède un minimum local au point x_k (pour k assez grand), et donc $\nabla \Psi_k(x_k) = 0$. □

Lemme 5.9. *Soit E un espace vectoriel normé, et soit $\bar{u} = (u_1, \dots, u_m) \in E^m$. On suppose que les u_i sont* linéairement indépendants. *Pour tout $k \in \mathbb{N}$, soient également $\bar{u}^k = (u_1^k, \dots, u_m^k) \in E^m$ et $\bar{\lambda}^k = (\lambda_1^k, \dots, \lambda_m^k) \in \mathbb{R}^m$, et posons $v^k := \sum_{i=1}^m \lambda_i^k u_i^k$. On suppose que $\bar{u}^k$ tend vers $\bar{u}$ quand $k \to \infty$, et que la suite (v^k) est bornée dans E. Alors la suite $(\bar{\lambda}^k)$ est bornée dans $\mathbb{R}^m$.*

Démonstration. Supposons que $(\bar{\lambda}^k)$ ne soit pas bornée. Alors, quitte à extraire une sous-suite, on peut en fait supposer que $\|\bar{\lambda}^k\|_\infty$ tend vers l'infini. Comme la suite (v^k) est bornée, on en déduit que $\frac{1}{\|\bar{\lambda}^k\|_\infty} v_k$ tend vers 0 ; autrement dit

$$\sum_{i=1}^m \mu_i^k u_i^k \to 0 , \qquad \text{où} \quad \mu_i^k = \frac{\lambda_i^k}{\|\bar{\lambda}^k\|_\infty} \cdot \tag{5.6}$$

Si on pose $\bar{\mu}^k = (\mu_1^k, \dots, \mu_m^k)$, alors $\|\bar{\mu}^k\|_\infty = 1$ pour tout k. Comme la sphère unité de $\mathbb{R}^m$ est compacte, on peut donc supposer, quitte à extraire une sous-suite, que la suite $(\bar{\mu}^k)$ converge vers un certain $\bar{\mu} \in \mathbb{R}^m$ vérifiant $\|\bar{\mu}\|_\infty = 1$ (en particulier, $\bar{\mu} \neq 0$).

Comme $\bar{u}^k \to u$, on obtient alors, en passant à la limite dans (5.6) :

$$\sum_{i=1}^m \mu_i u_i = 0 ;$$

ce qui est absurde car les u_i sont linéairement indépendants et $\bar{\mu} \neq 0$. □

Preuve du Théorème 5.1. Il s'agit de montrer qu'il existe des nombres réels $\lambda_1, \dots, \lambda_m$ tels que

$$\nabla f(a) = \sum_{i=1}^m \lambda_i \nabla C_i(a) .$$

On va appliquer le Lemme 5.7 avec la fonction Φ définie par $\Phi(x) = \sum_{i=1}^m C_i(x)^2$ (qui vérifie bien les hypothèses du lemme). On a ainsi pour tout $k \in \mathbb{N}$:

$$\Psi_k(x) = f(x) + \|x-a\|^2 + k\sum_{i=1}^m C_i(x)^2\,.$$

La fonction Φ est différentiable sur Ω, avec $\nabla\Phi(x) = \sum_{i=1}^m C_i(x)\,\nabla C_i(x)$; donc

$$\nabla\Psi_k(x) = \nabla f(x) + 2(x-a) + 2k\sum_{i=1}^m C_i(x)\,\nabla C_i(x)\,.$$

D'après le Lemme 5.7, on peut trouver une suite (x_k) tendant vers a telle que $\nabla\Psi_k(x_k) = 0$ pour tout k, autrement dit

$$\nabla f(x_k) + 2(x_k - a) = \sum_{i=1}^m \underbrace{(-2kC_i(x_k))}_{\lambda_i^k}\,\nabla C_i(x_k) \tag{5.7}$$

Comme $x_k \to a$, le premier membre de (5.7) tend vers $\nabla f(a)$ quand $k \to \infty$. En particulier, ce premier membre reste borné, et donc $v_k := \sum_{i=1}^m \lambda_i^k\,\nabla C_i(x_k)$ reste borné. De plus, $\nabla C_i(x_k) \to \nabla C_i(a)$ pour $i = 1,\dots,m$, et par hypothèse les $\nabla C_i(a)$ sont linéairement indépendants dans $E = \mathbb{R}^n$. Si on pose $\bar\lambda^k := (\lambda_1^k,\dots,\lambda_m^k)$, on en déduit d'après le Lemme 5.9 que la suite $(\bar\lambda^k)$ est bornée dans $\mathbb{R}^m$. Quitte à extraire une sous-suite, on peut donc supposer que $(\bar\lambda^k)$ converge vers un certain $\bar\lambda = (\lambda_1,\dots,\lambda_m) \in \mathbb{R}^m$.

En passant à la limite dans (5.7), on obtient alors ce qu'on voulait :

$$\nabla f(a) = \sum_{i=1}^m \lambda_i\,\nabla C_i(a)\,.$$

$\square$

CHAPITRE 5

Inversion locale, fonctions implicites

1. Applications linéaires inversibles

DÉFINITION 1.1. *Soient E et E' deux evn. On dira qu'une application linéaire continue $L : E \to E'$ est* ***inversible*** *(ou encore que L est un* ***isomorphisme****) si L est bijective et si L^{-1} est continue. On notera $GL(E, E')$ l'ensemble des isomorphismes de E sur E', et $GL(E)$ si $E = E'$.*

Remarque 1. Si E et E' sont de dimension finie, la continuité de L^{-1} est automatique. Plus généralement, c'est encore vrai si E et E' sont *complets*, mais ceci est loin d'être évident. Il s'agit d'un théorème célèbre dû à Banach (le **théorème d'isomorphisme de Banach**).

Remarque 2. Il n'existe pas nécessairement d'isomorphisme entre E et E' ; s'il en existe, on dit que les evn E et E' sont **isomorphes**. Lorsque E et E' sont de dimension finie, ils sont isomorphes si et seulement si $\dim(E) = \dim(E')$.

LEMME 1.2. *Si E et E' sont deux evn isomorphes, alors l'application $X \mapsto X^{-1}$ est continue de $GL(E, E')$ dans $GL(E', E)$.*

PREUVE EN DIMENSION FINIE. Si E et E' sont de (même) dimension finie, on peut les identifier tous les deux à $\mathbb{R}^n$ et donc identifier $GL(E, E')$ à $GL_n(\mathbb{R})$, l'ensemble des matrices inversibles de taille n. Alors la célèbre formule

$$X^{-1} = \frac{1}{\det(X)} \, {}^t\mathrm{Com}(X)$$

montre que les coefficients de la matrice X^{-1} sont des fonctions rationnelles de ceux de X, d'où la continuité. $\square$

PREUVE DANS LE CAS GÉNÉRAL. Fixons $A \in GL(E, E')$. Le point clé (qu'on utilisera aussi plus loin) est l'identité suivante, valable pour tout $X \in GL(E, E')$:

$$X^{-1} - A^{-1} = X^{-1}(A - X)A^{-1} . \tag{1.1}$$

Cette identité se démontre en "réduisant au même dénominateur" ; de façon précise, en factorisant à gauche par X^{-1} et à droite par A^{-1} dans $X^{-1} - A^{-1}$.

On en déduit d'abord que si $\|X - A\| \leq 1/2\|A^{-1}\|$, alors $\|X^{-1}\| - \|A^{-1}\| \leq \|X^{-1}\|/2$, et donc $\|X^{-1}\| \leq 2\|A^{-1}\|$. En revenant à (1.1), on obtient ensuite

$$\|X^{-1} - A^{-1}\| \leq 2\|A^{-1}\|^2 \, \|X - A\|$$

pour tout $X \in GL(E)$ tel que $\|X - A\| \leq 1/2\|A^{-1}\|$; d'où la continuité. $\square$

LEMME 1.3. *Soit E un espace* de Banach. *Si $H \in \mathcal{L}(E)$ vérifie $\|H\| < 1$, alors $Id - H$ est inversible et on a*

$$(Id - H)^{-1} = \sum_{k=0}^{\infty} H^k.$$

Démonstration. Comme $\|H^k\| \le \|H\|^k$ pour tout k et comme $\|H\| < 1$, on voit que la série $\sum H^k$ est normalement convergente, et donc convergente dans $\mathcal{L}(E)$ puisque $\mathcal{L}(E)$ est complet (proposition 5.6 du chapitre 1). En posant $S = \sum_0^\infty H^k$, on a (par continuité du produit)

$$SH = \sum_{k=0}^{\infty} H^{k+1} = HS\,.$$

Comme $S = \sum_0^\infty H^k$, on en déduit $SH = S - H^0 = S - Id = HS$, autrement dit $S(Id - H) = Id = (Id - H)S$. Ainsi, $(Id - H)$ est inversible d'inverse S. □

Corollaire 1.4. *Si $L \in \mathcal{L}(E)$ vérifie $\|L - I\| < 1$, alors L est inversible.*

Démonstration. Il suffit d'écrire $L = I - H$, avec $H = I - L$. □

Théorème 1.5. *Si E est un espace de Banach, alors $GL(E)$ est ouvert dans $\mathcal{L}(E)$ et l'application $X \mapsto X^{-1}$ est de classe $\mathcal{C}^1$ sur $GL(E)$. En notant J cette application, la différentielle de J est donnée par*

$$DJ(X)H = -X^{-1}HX^{-1}\,.$$

Démonstration. Soit $A \in GL(E)$ quelconque, et posons $r_A = 1/\|A^{-1}\|$. Si $X \in \mathcal{L}(E)$ vérifie $\|X - A\| < r_A$, alors $\|A^{-1}X - I\| = \|A^{-1}(X - A)\| \le \|A^{-1}\|\,\|X - A\| < 1$; donc $A^{-1}X \in GL(E)$ d'après le lemme 1.3, et donc $X \in GL(E)$ également. Ainsi, la boule $B(A, r_A)$ est contenue dans $GL(E)$, ce qui montre que $GL(E)$ est ouvert dans $\mathcal{L}(E)$.

D'après l'identité (1.1) et comme l'application J est continue (lemme 1.2), on peut écrire

$$\begin{aligned} J(A+H) - J(A) &= -J(A+H)HJ(A) \\ &= -\big(J(A) + \varepsilon(H)\big)HJ(A) \\ &= -A^{-1}HA^{-1} - \varepsilon(H)HA^{-1}\,, \end{aligned}$$

où $\varepsilon(H)$ tend vers 0 quand H tend vers 0. Comme l'application $H \mapsto -A^{-1}HA^{-1}$ est linéaire continue sur $\mathcal{L}(E)$ et comme $\varepsilon(H)HA^{-1} = o(\|H\|)$ (car $\|\varepsilon(H)HA^{-1}\| \le \|A^{-1}\|\,\|\varepsilon(H)\|\,\|H\|$), on en déduit que J est différentiable en tout point $A \in GL(E)$, avec $DJ(A)H = -A^{-1}HA^{-1}$.

Notons $B : \mathcal{L}(E) \times \mathcal{L}(E) \to \mathcal{L}(\mathcal{L}(E))$ l'application définie par $B(U,V)(H) = UHV$. L'aplication B est clairement bilinéaire, et elle est également continue car $\|B(U,V)(H)\| \le \|U\|\,\|V\|\,\|H\|$ pour tout $H \in \mathcal{L}(E)$ et donc $\|B(U,V)\| \le \|U\|\,\|V\|$. De plus, on a $DJ(X) = -B(J(X), J(X))$ pour tout $X \in GL(E)$ d'après ce qu'on vient de voir. Comme J est continue sur $GL(E)$, on en déduit que DJ est continue et donc que J est de classe $\mathcal{C}^1$. □

2. Difféomorphismes

Définition 2.1. *Soient E et E' deux evn, Ω un ouvert de E et Ω' un ouvert de E'. On dit qu'une application $\Phi : \Omega \to \Omega'$ est un* ***difféomorphisme de Ω sur Ω'*** *si Φ est une bijection de Ω sur Ω', et si Φ et Φ^{-1} sont toutes les deux de classe $\mathcal{C}^1$.*

Remarque. Il est souvent utile de penser à un difféomorphisme comme à un “changement de variable(s)”, avec toutes les applications que cette terminologie est susceptible de faire venir à l'esprit.

Exemple 1. Si $L : E \to E'$ est une application linéaire (continue) inversible, alors L est un difféomorphisme de E sur E'.

Exemple 2. Soit I un intervalle ouvert de $\mathbb{R}$ et soit $\varphi : I \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$, avec $\varphi'(t) \neq 0$ pour tout $t \in I$. Alors $J = \varphi(I)$ est un intervalle ouvert et φ est un difféomorphisme de I sur J, avec

$$(\varphi^{-1})'(x) = \frac{1}{\varphi'(\varphi^{-1}(x))}\,\cdot$$

DÉMONSTRATION. Comme φ' est continue et ne s'annule jamais, elle garde un signe constant d'après le théorème des valeurs intermédiaires. Par conséquent φ est strictement monotone. D'après un théorème bien connu, on sait alors que $J = \varphi(I)$ est un intervalle ouvert, que φ est une bijection de I sur J et que φ^{-1} est *continue* sur J. En écrivant

$$\frac{\varphi^{-1}(y) - \varphi^{-1}(x)}{y - x} = \frac{1}{\dfrac{\varphi(v) - \varphi(u)}{v - u}}$$

où $u = \varphi^{-1}(x)$ et $v = \varphi^{-1}(y)$, on en déduit facilement que φ^{-1} est dérivable sur J avec la formule souhaitée pour $(\varphi^{-1})'$. Comme φ' et φ^{-1} sont continues, on voit alors que $(\varphi^{-1})'$ est continue, autrement dit que φ^{-1} est $\mathcal{C}^1$. $\square$

Exemple 3. Soit $\varphi : \mathbb{R} \to \mathbb{R}$ définie par $\varphi(t) = t^3$. Alors φ est une bijection $\mathcal{C}^1$ de $\mathbb{R}$ sur $\mathbb{R}$, mais φ n'est pas un difféomorphisme car $\varphi^{-1}(t) = \sqrt[3]{t}$ n'est pas dérivable en 0.

Exercice. Soit $\Omega = \{(x, y) \in \mathbb{R}^2;\ x < y\}$ et soit $\Phi : \Omega \to \mathbb{R}^2$ l'application définie par $\Phi(x, y) = (x + y, xy)$. Montrer que Φ est un difféomorphisme de Ω sur l'ouvert $\Omega' := \{(s, p) \in \mathbb{R}^2;\ s^2 - 4p > 0\}$, et déterminer Φ^{-1}.

PROPOSITION 2.2. *Soient E et E' deux evn. Si $\Phi : \Omega \to \Omega'$ est un difféomorphisme d'un ouvert $\Omega \subset E$ sur un ouvert $\Omega' \subset E'$, alors $D\Phi(u) : E \to E'$ est inversible pour tout $u \in \Omega$, avec $D\Phi(u)^{-1} = D\Phi^{-1}(\Phi(u))$.*

DÉMONSTRATION. On a $\Phi^{-1} \circ \Phi = (Id_E)_{|\Omega}$ et $\Phi \circ \Phi^{-1} = (Id_{E'})_{|\Omega'}$ par définition de Φ^{-1}. Comme Id_E et $Id_{E'}$ sont linéaires continues, on en déduit (en différentiant et en utilisant le théorème des fonctions composées) qu'on a $D\Phi^{-1}(\Phi(u))D\Phi(u) = Id_E$ pour tout $u \in \Omega$ et $D\Phi(\Phi^{-1}(u'))D\Phi^{-1}(u') = Id_{E'}$ pour tout $u' \in \Omega'$. En prenant $u' = \Phi(u)$ la deuxième identité s'écrit $D\Phi(u)D\Phi^{-1}(\Phi(u)) = Id_{E'}$, et on voit donc que $D\Phi(u)$ est inversible d'inverse $D\Phi^{-1}(\Phi(u))$. $\square$

COROLLAIRE 2.3. *S'il existe un difféomorphisme d'un ouvert de E sur un ouvert de E', alors E et E' sont isomorphes. En particulier, si $m, n \in \mathbb{N}^*$, alors il ne peut exister de difféomorphisme de $\mathbb{R}^n$ sur $\mathbb{R}^m$ que si $n = m$.*

Remarque. En fait, un résultat beaucoup plus fort (et beaucoup plus difficile à démontrer) est vrai : s'il existe un **homéomorphisme** (i.e. une bijection continue, de réciproque continue) d'un ouvert de $\mathbb{R}^n$ sur un ouvert de $\mathbb{R}^m$, alors $n = m$. C'est un théorème célèbre dû à Brouwer.

PROPOSITION 2.4. *Soient E et E' deux evn, et soit $\Phi : \Omega \to \Omega'$ une bijection d'un ouvert $\Omega \subset E$ sur un ouvert $\Omega' \subset E'$. Soit aussi $p \in \Omega$, et soit $p' = \Phi(p)$. On suppose que Φ est différentiable en p, que $D\Phi(p) : E \to E'$ est inversible, et que Φ^{-1} est* continue *au point p'. Alors Φ^{-1} est différentiable en p' et $D\Phi^{-1}(p') = D\Phi(p)^{-1} = D\Phi(\Phi^{-1}(p'))^{-1}$.*

Démonstration. Posons $L = D\Phi(p)$. Il s'agit de montrer qu'on peut écrire, pour h' tendant vers 0 dans E' :

$$\Phi^{-1}(p'+h') = \Phi^{-1}(p') + L^{-1}h' + o(\|h'\|)\,.$$

Comme Φ^{-1} est continue au point p', on peut certainement écrire

$$\Phi^{-1}(p'+h') = \Phi^{-1}(p') + \varepsilon(h') = p + \varepsilon(h')\,, \tag{2.1}$$

où $\varepsilon(h')$ tend vers 0 quand $h' \to 0$. Ce qu'il reste à faire est de montrer qu'on a en fait $\varepsilon(h') = L^{-1}h' + o(\|h'\|)$.

En appliquant Φ à (2.1) et en utilisant la différentiabilité de Φ au point p, on obtient

$$\begin{array}{rcl} p'+h' & = & \Phi(p+\varepsilon(h')) \\ & = & \Phi(p) + L\varepsilon(h') + o(\|\varepsilon(h')\|) \\ & = & p' + L\varepsilon(h') + o(\|\varepsilon(h')\|)\,; \end{array}$$

autrement dit $h' = L\varepsilon(h') + o(\|\varepsilon(h')\|$, ou encore $\varepsilon(h') = L^{-1}h' + L^{-1}\big(o(\|\varepsilon(h')\|)\big)$. Mais comme l'application linéaire L^{-1} est *continue*, on a aussi $L^{-1}(u) = O(\|u\|)$ et donc $L^{-1}\big(o(\|\varepsilon(h')\|)\big) = o(\|\varepsilon(h')\|)$; d'où

$$\varepsilon(h') = L^{-1}h' + o(\|\varepsilon(h')\|)\,. \tag{2.2}$$

Par définition d'un "o", on donc peut trouver $\delta > 0$ tel que

$$\|h'\| \leq \delta \implies \|\varepsilon(h')\| \leq \|L^{-1}h'\| + (1/2)\|\varepsilon(h'))\|,,$$

autrement dit $\|\varepsilon(h')\| \leq 2\|L^{-1}h'\|$ pour $\|h'\| \leq \delta$. Ainsi, on a $\|\varepsilon(h')\| = O(\|L^{-1}h'\|) = O(\|h'\|)$ au voisinage de 0, d'où la conclusion souhaitée en revenant à (2.2) :

$$\varepsilon(h') = L^{-1}h' + o(\|h'\|)\,.$$

□

Corollaire 2.5. *Soit $\Phi : \Omega \to \Omega'$ une bijection $\mathcal{C}^1$ d'un ouvert $\Omega \subset E$ sur un ouvert $\Omega' \subset E'$. Si $D\Phi(u)$ est inversible pour tout $u \in \Omega$* et si Φ^{-1} est continue *sur Ω', alors Φ est un difféomorphisme (i.e. Φ^{-1} est $\mathcal{C}^1$).*

Démonstration. D'après la proposition, Φ est différentiable en tout point avec $D\Phi^{-1}(x) = D\Phi(\Phi^{-1}(x))^{-1}$ pour tout $x \in \Omega'$. Comme Φ^{-1} est continue et comme l'application $X \mapsto X^{-1}$ est continue sur $GL(E, E')$ d'après le lemme 1.2, on voit que $D\Phi^{-1}$ est continue, i.e. que Φ^{-1} est $\mathcal{C}^1$. □

Remarque. On verra plus loin que l'hypothèse de continuité faite sur Φ^{-1} est en fait superflue.

3. Le théorème d'inversion locale

3.1. Énoncé et conséquences immédiates. Dans ce qui suit, E et E' sont des evn, Ω est un ouvert de E, et Ω' est un ouvert de E'.

Définition 3.1. *Soit $\Phi : \Omega \to \Omega'$, et soit $p_0 \in \Omega$. On dit que l'application Φ est un **difféomorphisme local** au point p_0 s'il existe un voisinage ouvert U de p_0 (contenu dans Ω) tel que $U' = \Phi(U)$ est un ouvert de E' et $\Phi_{|U} : U \to U'$ est un difféomorphisme de U sur U'.*

Remarque. Cela signifie qu'on peut "localement inverser" l'application Φ au voisinage de p_0. Autrement dit : pour tout $\lambda \in E'$ suffisamment proche de $\lambda_0 = \Phi(p_0)$ (i.e. $\lambda \in U'$) il existe un unique $u = u(\lambda) \in E$ proche de p_0 (i.e. $u(\lambda) \in U$) solution de l'équation $\Phi(u) = \lambda$; et de plus l'application $\lambda \mapsto u(\lambda)$ (i.e. $(\Phi_{|U})^{-1} : U' \to U$) est de classe $\mathcal{C}^1$.

Exemple. On prend $E = \mathbb{R} = E'$. Si $\varphi : \Omega \to \mathbb{R}$ est une fonction de classe $\mathcal{C}^1$ sur un ouvert $\Omega \subset \mathbb{R}$ et si $t_0 \in \Omega$ est tel que $\varphi'(t_0) \neq 0$, alors φ est un difféomorphisme local au point t_0.

Démonstration. Par continuité de φ', on peut trouver $\delta > 0$ tel que $\varphi'(t) \neq 0$ pour tout $t \in I :=]-\delta, \delta[$. On sait qu'alors $J = \varphi(I)$ est un intervalle ouvert et que $\varphi_{|I}$ est un difféomorphisme de I sur J (cf le 2è exemple de difféomorphisme donné plus haut). □

Si $\Phi : \Omega \to E'$ est un difféomorphisme local au point p_0, alors $D\Phi(p_0) : E \to E'$ est nécessairement inversible, d'après la proposition 2.2. Le théorème suivant (beaucoup plus difficile) montre que la réciproque est vraie.

Théorème 3.2. (théorème d'inversion locale)

On suppose que les evn E et E' sont complets. *Soit $\Phi : \Omega \to E'$ de classe $\mathcal{C}^1$, et soit $p_0 \in \Omega$. Si $D\Phi(p_0) : E \to E'$ est inversible, alors Φ est un difféomorphisme local au point p_0.*

Remarque. Si $E = \mathbb{R}^n = E'$, la condition "$D\Phi(p_0)$ inversible" signifie que le déterminant de la matrice jacobienne de Φ au point p_0 est non-nul. Ce déterminant s'appelle le **déterminant jacobien** de Φ au point p_0, et se note $J_\Phi(p_0)$.

Corollaire 3.3. *Soit $\Phi : \Omega \to \Omega'$ une bijection $\mathcal{C}^1$ d'un ouvert $\Omega \subset E$ sur un ouvert $\Omega' \subset E'$. Si $D\Phi(u)$ est inversible pour tout $u \in \Omega$, alors Φ est un difféomorphisme.*

Démonstration. Soit $p_0' \in \Omega'$ quelconque, et soit $p_0 = \Phi^{-1}(p_0')$. En appliquant le théorème d'inversion locale au point p_0, on voit que Φ^{-1} est de classe $\mathcal{C}^1$ au voisinage de p_0' ; et comme p_0' est arbitraire, cela montre que Φ^{-1} est $\mathcal{C}^1$ sur Ω'. □

Corollaire 3.4. *Si $\Phi : \Omega \to \Omega'$ est une bijection $\mathcal{C}^1$ entre deux ouverts de $\mathbb{R}^n$ et si $J_\Phi(u) \neq 0$ pour tout $u \in \Omega$, alors Φ est un difféomorphisme.*

Remarque. L'intérêt "pratique" de ce dernier résultat est évident : pour montrer que Φ^{-1} est de classe $\mathcal{C}^1$, il est inutile de déterminer explicitement Φ^{-1} ! Il suffit de vérifier que le déterminant jacobien *de* Φ ne s'annule jamais.

Exemple 3.5. (coordonnées polaires)

Soit $\Phi :]0, \infty[\times]0, 2\pi[\to \mathbb{R}^2$ l'application définie par $\Phi(r, \theta) = (r\cos\theta, r\sin\theta)$. Il est clair que Φ est de classe $\mathcal{C}^1$, et que Φ est une bijection de $\Omega :=]0, \infty[\times]0, 2\pi[$ sur $\Omega' := \mathbb{R}^2 \setminus \{(x, 0);\ x \geq 0\}$. (Les seuls points non atteints par Φ sont les points p pour lesquels la demi-droite $[0, p)$ fait un angle égal à 0 avec l'axe $[0, x)$, i.e. les points de la forme $(x, 0)$ avec $x \geq 0$; et tout point $(x, y) \in \Omega'$ s'écrit de manière unique sous la forme $(r\cos\theta, r\sin\theta)$ si on impose $0 < \theta < 2\pi$). De plus, le déterminant jacobien de Φ est facile à calculer : on a

$$J_\Phi(r, \theta) = \det\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} = r\,,$$

donc en particulier $J_\Phi(r, \theta) \neq 0$ pour tout $(r, \theta) \in \Omega'$. Ainsi, Φ est un difféomorphisme de Ω sur Ω'.

Exercice. Donner une formule explicite pour Φ^{-1}.

Mentionnons encore deux autres conséquences très simples du théorème d'inversion locale.

COROLLAIRE 3.6. (théorème de l'application ouverte)
*Soit $\Phi : \Omega \to E'$ de classe $\mathcal{C}^1$. Si $D\phi(u)$ est inversible pour tout $u \in \Omega$, alors Φ est une application **ouverte** : l'image par Φ de tout ouvert $O \subset \Omega$ est un ouvert de E'.*

DÉMONSTRATION. Soit $O \subset \Omega$ ouvert, et soit $p_0' \in \Phi(O)$ quelconque. Choisissons un point $p_0 \in O$ tel que $\Phi(p_0) = p_0'$. Commme $D\Phi(p_0)$ est inversible, le théorème d'inversion locale appliqué dans O fournit un voisinage ouvert U de p_0 contenu dans O tel que $U' = \Phi(U)$ est un ouvert de E', et donc un voisinage ouvert de p_0'. Comme $U' \subset \Phi(O)$, cela montre que tout point $p_0' \in \Phi(O)$ est intérieur à $\Phi(O)$, autrement dit que $\Phi(O)$ est un ouvert de E'. $\square$

COROLLAIRE 3.7. *Soit $\Phi : \Omega \to E'$ de classe $\mathcal{C}^1$. Si Φ est* injective *et si $D\phi(u)$ est inversible pour tout $u \in \Omega$, alors $\Omega' := \Phi(\Omega)$ est un ouvert de E' et Φ est un difféomorphisme de Ω sur Ω'.*

DÉMONSTRATION. C'est immédiat d'après les corollaires 3.6 et 3.3. $\square$

3.2. Preuve. La preuve du théorème d'inversion locale va utiliser de manière essentielle le **théorème du point fixe**, sous la forme suivante.

LEMME 3.8. (théorème du point fixe)
Soit M un fermé d'un espace de Banach E, et soit $f : M \to E$. On suppose qu'on a $f(M) \subset M$, et que f est k-lipschitzienne pour un certain $k < 1$. Alors f possède un unique point fixe dans M.

DÉMONSTRATION. Soit $x_0 \in M$. Comme M est stable par f, on peut définir une suite $(x_n)_{n\geq 0} \subset M$ par la récurrence $x_{n+1} = f(x_n)$. Comme f est k-lipschitzienne, on vérifie facilement par récurrence qu'on a

$$\|x_n - x_{n-1}\| \leq k^{n-1}\,\|x_1 - x_0\|$$

pour tout $n \geq 1$. Comme $k < 1$, cela montre que la série $\sum_{n\geq 1} \|x_n - x_{n-1}\|$ est convergente ; et comme l'espace E est supposé complet, on en déduit que la série $\sum_{n\geq 1}(x_n - x_{n-1})$ converge dans E. Mais $x_n = x_0 + \sum_{i=1}^n (x_i - x_{i-1})$ pour tout $n \geq 1$, donc cela signifie que la suite (x_n) est convergente, $x_n \to \alpha \in E$. Le point α appartient à M car $(x_n) \subset M$ et M est fermé dans E ; et par continuité de f, on a $f(\alpha) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} x_{n+1} = \alpha$. On vient ainsi de montrer que f possède un point fixe. L'unicité est laissée en exercice. $\square$

PREUVE DU THÉORÈME D'INVERSION LOCALE. Soit $\Phi : \Omega \to E'$ de classe $\mathcal{C}^1$ (où Ω est un ouvert de E) et soit $p_0 \in \Omega$ tel que $D\Phi(p_0) : E \to E'$ est inversible. On cherche un voisinage ouvert U de p_0 tel que $U' = \Phi(U)$ est un ouvert de E' et $\Phi_{|U}$ est un difféomorphisme de U sur U'. La démonstration va se faire en plusieurs étapes.

FAIT 0. On peut supposer $E' = E$ et $D\Phi(p_0) = Id_E$.

DÉMONSTRATION. Posons $L = D\Phi(p_0)$, et soit $\widetilde{\Phi} : \Omega \to E$ l'application définie par $\widetilde{\Phi}(u) = L^{-1}(\Phi(u))$. Comme L^{-1} est linéaire continue, $\widetilde{\Phi}$ est de classe $\mathcal{C}^1$ (par composition) et on a $D\widetilde{\Phi}(p_0) = DL^{-1}(\Phi(p_0)) \circ D\Phi(p_0) = L^{-1} \circ D\Phi(p_0) = Id$. De plus, si on sait montrer que $\widetilde{\Phi}$ est un difféomorphisme local en p_0, alors on saura le faire pour Φ car $\Phi = L \circ \widetilde{\Phi}$ et L est un difféomorphisme de E sur E'. $\square$

Fait 1. On peut trouver un voisinage ouvert U de p_0 (contenu dans Ω) tel que $D\Phi(u)$ est inversible pour tout $u \in U$ et l'application $u \mapsto u - \Phi(u)$ est $\frac{1}{2}$ - lipschitzienne sur U.

Démonstration. Comme $D\Phi$ est continue au point p_0 et $D\Phi(p_0) = Id$, on peut trouver $\delta > 0$ tel que $U := B(p_0, \delta) \subset \Omega$ et $\|D\Phi(u) - Id\| \leq 1/2$ pour tout $u \in U$. Alors $D\Phi(u)$ est inversible pour $u \in U$, d'après le lemme 1.3. Si on définit $h : U \to E$ par $h(u) = \Phi(u) - u$, i.e. $h = (\Phi - Id)_{|U}$, alors $\|Dh(u)\| = \|D\Phi(u) - Id\| \leq 1/2$ pour tout $u \in U$; donc h est $\frac{1}{2}$-lipschitzienne sur U, d'après l'inégalité des accroissements finis (car la boule U est convexe). $\square$

Fait 2. L'application Φ est injective sur U, et $(\Phi_{|U})^{-1}$ est continue sur $U' = \Phi(U)$.

Démonstration. Si $u, v \in U$, alors

$$\begin{aligned} \|\Phi(v) - \Phi(u)\| &= \|(\Phi(v) - v) - (\Phi(u) - u) + (v - u)\| \\ &\geq \|v - u\| - \|(\Phi(v) - v) - (\Phi(u) - u)\| \\ &\geq \frac{1}{2}\|v - u\|, \end{aligned}$$

d'après le fait 1.

On en déduit immédiatement que Φ est injective sur U : si $\Phi(u) = \Phi(v)$, alors $\frac{1}{2}\|v - u\| \leq \|\Phi(v) - \Phi(u)\| = 0$ et donc $u = v$. D'autre part, si $u' = \Phi(u)$ et $v' = \Phi(v)$ sont quelconques dans $U' = \Phi(U)$, alors $\|(\Phi_{|U})^{-1}(v') - (\Phi_{|U})^{-1}(u')\| = \|v - u\| \leq 2\|\Phi(v) - \Phi(u)\| = 2\|v' - u'\|$. Ainsi, l'application $(\Phi_{|U})^{-1}$ est 2-lipschitzienne, et donc continue sur U'.

$\square$

Fait 3. L'ensemble $U' = \Phi(U)$ est un ouvert de E.

Démonstration. Soit $u'_0 \in U'$ quelconque, et écrivons $u'_0 = \Phi(u_0)$, où $u_0 \in U$. On cherche $\delta > 0$ tel que $B(u'_0, \delta) \subset U'$, autrement dit tel que pour tout $u' \in B(u'_0, \delta)$, l'équation $\Phi(u) = u'$ possède une solution $u \in U$.

Soit $\alpha > 0$ tel que $\overline{B}(u_0, \alpha) \subset U$. On va montrer que $\delta := \alpha/2$ convient. Plus précisément, on va montrer que pour tout $u' \in E$ vérifiant $\|u' - u'_0\| \leq \delta$, l'équation $\Phi(u) = u'$ possède une solution $u \in \overline{B}(u_0, \alpha)$.

Fixons u' vérifiant $\|u' - u'_0\| \leq \delta$, et soit $\Psi : \overline{B}(u_0, \alpha) \to E$ l'application définie par

$$\Psi(u) = u - \Phi(u) + u'.$$

Par définition, l'équation $\Phi(u) = u'$ est équivalente à $\Psi(u) = u$. Il s'agit donc montrer que Ψ possède un *point fixe* dans $\overline{B}(u_0, \alpha)$. Comme l'espace E est supposé *complet*, il suffit pour cela de vérifier que les hypothèses du théorème du point fixe (lemme 3.8) sont vérifiées avec $M := \overline{B}(u_0, \alpha) \subset E$ et $f = \Psi$.

Par définition, Ψ diffère de l'application $u \mapsto u - \Phi(u)$ par une constante ; donc Ψ est $\frac{1}{2}$-lipschitzienne sur $\overline{B}(u_0, \alpha)$ d'après le fait 1. Par conséquent, il reste simplement à montrer que la boule $\overline{B}(u_0, \alpha)$ est stable par Ψ.

Si $u \in \overline{B}(u_0, \alpha)$, alors

$$\begin{aligned} \|\Psi(u) - u_0\| &\leq \|\Psi(u) - \Psi(u_0)\| + \|\Psi(u_0) - u_0\| \\ &= \|\Psi(u) - \Psi(u_0)\| + \|u' - \Phi(u_0)\| \\ &\leq \frac{1}{2}\|u - u_0\| + \|u' - u_0'\| \\ &\leq \frac{\alpha}{2} + \delta = \alpha\,. \end{aligned}$$

On a donc bien $\Psi(u) \in \overline{B}(u_0, \alpha)$ pour tout $u \in \overline{B}(u_0, \alpha)$, ce qui termine la démonstration du fait 3. □

On peut maintenant conclure. D'après les faits 1, 2, 3, on sait que $\Phi_{|U} : U \to U'$ est une bijection $\mathcal{C}^1$ de U sur U', que U' est un ouvert de E, que $D\Phi(u)$ est inversible pour tout $u \in U$, et que $(\Phi_{|U})^{-1}$ est continue sur U'. D'après le corollaire 2.5, on en déduit que $\Phi_{|U}$ est un difféomorphisme de U sur U' ; donc Φ est un difféomorphisme local au point p_0. □

3.3. Un exemple. Pour illustrer le théorème d'inversion locale, on va maintenant expliquer comment il peut permettre de démontrer l'existence de solutions pour certaines équations différentielles. De façon précise, on va donner les grandes lignes de la preuve du résultat suivant. Les détails débordent un peu le cadre de ce cours, car on va avoir besoin d'utiliser le **théorème d'Ascoli**, qui est un résultat non trivial d'analyse fonctionnelle.

Proposition 3.9. *Soit I un intervalle ouvert de $\mathbb{R}$ et soit $p : I \to \mathbb{R}$ une application de classe $\mathcal{C}^1$ strictement croissante et surjective. Pour toute fonction $g : \mathbb{R} \to \mathbb{R}$ continue et 2π-périodique, il existe une fonction $y : \mathbb{R} \to \mathbb{R}$ de classe $\mathcal{C}^1$, 2π-périodique, à valeurs dans I, telle que $u' + p \circ u = g$.*

Preuve abrégée. Dans ce qui suit, on notera $\mathcal{C}^0_{2\pi}$ l'espace des fonctions continues 2π-périodiques de $\mathbb{R}$ dans $\mathbb{R}$, muni de la norme $||\ .\ ||_\infty$, et $\mathcal{C}^1_{2\pi}$ l'espace des fonctions 2π-périodiques de classe $\mathcal{C}^1$, muni de la norme définie par $||u|| = ||u||_\infty + ||u'||_\infty$. On vérifie sans difficulté que les espaces $\mathcal{C}^0_{2\pi}$ et $\mathcal{C}^1_{2\pi}$ sont *complets* pour leurs normes respectives.

Posons $\mathcal{U} = \{u \in \mathcal{C}^1_{2\pi};\ u(\mathbb{R}) \subset I\}$, et introduisons l'application $\Phi : \mathcal{U} \to \mathcal{C}^0_{2\pi}$ définie par

$$\Phi(u) = u' + p \circ u\,.$$

Il s'agit très exactement de montrer que l'application Φ est *surjective.*

Fait 1. Si $\alpha \in \mathcal{C}^0_{2\pi}$ vérifie $\int_0^{2\pi} \alpha(t)\,dt \neq 0$ alors, pour toute fonction $v \in \mathcal{C}^0_{2\pi}$, il existe une unique fonction $h \in \mathcal{C}^1_{2\pi}$ telle que $h' + \alpha h = v$.

Preuve du fait 1. Une fonction $v \in \mathcal{C}^0_{2\pi}$ étant fixée, il s'agit de montrer que l'équation différentielle linéaire $h' + \alpha h = v$ possède une et une seule solution 2π-périodique.

On sait que les solutions maximales de cette équation sont définies sur $\mathbb{R}$ et de la forme $h_C(t) = h_0(t) + Ce^{-\int_0^t \alpha(s)\,ds}$, où u_0 est une solution particulière (obtenue par la méthode de variation de la constante) et C est une constante. Comme $\rho = e^{-\int_0^{2\pi} \alpha(s)\,ds} \neq 1$ par hypothèse, il existe une unique constante C telle que $h_C(2\pi) = h_C(0)$ (il s'agit de résoudre l'équation $h_0(2\pi) + C\rho = h_0(0) + C$). Pour cette valeur de C, la fonction $\widetilde{h}_C$ définie par $\widetilde{h}_C(t) = u_C(t + 2\pi)$ est alors solution de la même équation différentielle que h_C et vérifie $\widetilde{h}_C(0) = h_C(0)$. Donc $\widetilde{h}_C = h_C$ d'après le théorème de Cauchy-Lipschitz, autrement dit h_C est 2π-périodique.

□

Fait 2. $\mathcal{U}$ est un ouvert de $\mathcal{C}^1_{2\pi}$, l'application Φ est de classe $\mathcal{C}^1$ sur $\mathcal{U}$, et la différentielle de Φ est donnée par la formule

$$D\Phi(u)h = h' + (p' \circ u) \times h\,.$$

Preuve du fait 2. Soit $u_0 \in \mathcal{U}$ quelconque. Comme u_0 est continue et périodique, $u_0(\mathbb{R})$ est un intervalle *compact* de $\mathbb{R}$, contenu dans I par définition de $\mathcal{U}$. Écrivons $u_0(\mathbb{R}) = [\alpha, \beta]$. Comme I est un intervalle ouvert, on peut trouver $\varepsilon > 0$ tel que $[\alpha - \varepsilon, \beta + \varepsilon] \subset I$. Si $u \in \mathcal{C}^1_{2\pi}$ vérifie $\|u - u_0\| < \varepsilon$, alors a fortiori $\|u - u_0\|_\infty < \varepsilon$ et donc $u(\mathbb{R}) \subset [\alpha - \varepsilon, \beta + \varepsilon] \subset I$. Ainsi, la boule $B(u_0, \varepsilon)$ est contenue dans $\mathcal{U}$, ce qui prouve que $\mathcal{U}$ est un ouvert de $\mathcal{C}^1_{2\pi}$.

La deuxième partie est un exercice tout à fait instructif. □

Fait 3. $\Phi(\mathcal{U})$ est un ouvert de $\mathcal{C}^0_{2\pi}$.

Preuve du fait 3. D'après le théorème de l'application ouverte (corollaire 3.6), il suffit de montrer que pour tout $u \in \mathcal{U}$, la différentielle $D\Phi(u)$ est un isomorphisme de $C^1_{2\pi}$ sur $C^0_{2\pi}$. Si on pose $\alpha = p' \circ u$, alors $D\Phi(u)h = h' + \alpha h$ pour tout $h \in \mathcal{C}^1_{2\pi}$, d'après le fait 2. De plus, on a $\int_0^{2\pi} \alpha(t)dt \neq 0$ car p' est strictement positive sur I. D'après le fait 1, on en déduit que $D\Phi(u)$ est bijective de $C^1_{2\pi}$ sur $C^0_{2\pi}$, et donc un isomorphisme d'après le théorème d'isomorphisme de Banach.

□

Fait 4. $\Phi(\mathcal{U})$ est également *fermé* dans $\mathcal{C}^0_{2\pi}$.

Preuve du fait 4. Commençons par observer que si $u \in \mathcal{C}^1_{2\pi}$, alors

$$\|p \circ u\|_\infty \leq \|\Phi(u)\|_\infty. \tag{3.1}$$

En effet, comme p est strictement croissante, $\|p \circ u\|_\infty$ est atteinte en un point a où u atteint sa borne supérieure ou sa borne inférieure. On a donc $u'(a) = 0$, d'où $\|p \circ u\|_\infty = p \circ u(a) = \Phi(u)(a) \leq \|\Phi(u)\|_\infty$.

Soit maintenant (g_n) une suite d'éléments de $\Phi(\mathcal{U})$ convergeant dans $\mathcal{C}^0_{2\pi}$ vers une fonction g. Il s'agit de montrer que la fonction g est encore dans $\in \Phi(\mathcal{U})$.

Écrivons $g_n = \Phi(u_n)$, où $u_n \in \mathcal{U}$. D'après (3.1) et comme la suite convergente $(\Phi(u_n)) = (g_n)$ est *bornée* dans $\mathcal{C}^0_{2\pi}$, on voit que la suite $(p \circ u_n)$ est uniformément bornée ; et comme $u'_n = \Phi(u_n) - p \circ u_n$ on en déduit que la suite (u'_n) est elle aussi uniformément bornée. Ainsi, les u_n sont C-lipschitziennes, pour une certaine constante C indépendante de n. D'après le théorème d'Ascoli, on peut donc supposer, quitte à extraire une sous-suite, que la suite (u_n) converge uniformément sur tout compact vers une fonctions continue $u : \mathbb{R} \to \mathbb{R}$. Comme de plus les u_n sont 2π-périodiques, la convergence est en fait uniforme sur $\mathbb{R}$, et la fonction u est 2π-périodique. En revenant à l'identité $u'_n = \Phi(u_n) - p \circ u_n = g_n - p \circ u_n$, on voit maintenant que la suite (u'_n) converge uniformément sur $\mathbb{R}$. Donc u est de classe $\mathcal{C}^1$, i.e. $u \in \mathcal{C}^1_{2\pi}$, et u' est la limite (uniforme) des u'_n.

En utilisant à nouveau le fait que la suite $(p \circ u_n)$ est uniformément bornée et en se souvenant que $p(x)$ tend vers $\pm\infty$ quand x tend vers les bornes de l'intervalle I (puisque $p(I) = \mathbb{R}$ par hypothèse), on voit que les fonctions u_n prennent leurs valeurs dans un compact $K \subset I$ indépendant de n. Comme $u_n(t) \to u(t)$, on en déduit que $u(t) \in K \subset I$ pour tout $t \in \mathbb{R}$, et donc que $u \in \mathcal{U}$. Enfin, en passant à la limite dans l'identité $g_n = u_n' + p \circ u_n$, on obtient $g = u' + p \circ u = \Phi(u)$. Ainsi $g \in \Phi(\mathcal{U})$, ce qui prouve que $\Phi(\mathcal{U})$ est fermé dans $\mathcal{C}^0_{2\pi}$. □

La preuve de la proposition est maintenant terminée : par *connexité* de $\mathcal{C}^0_{2\pi}$, on déduit des faits 3 et 4 que $\Phi(\mathcal{U}) = \mathcal{C}^0_{2\pi}$; autrement dit l'application Φ est surjective, ce qui est la conclusion souhaitée. □

4. Le théorème des fonctions implicites

4.1. Introduction. Le théorème d'inversion locale permet de résoudre localement des équations du type $\Phi(u) = \lambda$, ou si on préfère $\Phi(u) - \lambda = 0$. On va maintenant s'intéresser à des équations plus générales, de la forme $F(\lambda, u) = 0$. Dans ce qui suit, Λ, E et E' sont des evn.

DÉFINITION 4.1. *Soit $F : \Lambda \times E \to E'$, et soit $(\lambda_0, u_0) \in \Lambda \times E$ tel que $F(\lambda_0, u_0) = 0$. On dira que* ***l'équation*** $F(\lambda, u) = 0$ ***définit implicitement*** u ***comme fonction*** $\mathcal{C}^1$ ***de*** λ ***au voisinage de*** (λ_0, u_0) *s'il existe un voisinage ouvert U de u_0 dans E et un voisinage ouvert V de λ_0 dans Λ tels que*

(i) *pour tout $\lambda \in V$, il existe un unique $u = u(\lambda) \in U$ tel que $F(\lambda, u(\lambda)) = 0$;*

(ii) *l'application $\lambda \mapsto u(\lambda)$ est de classe $\mathcal{C}^1$ sur V.*

Remarque. D'après la propriété d'unicité dans (i), on a alors $u(\lambda_0) = u_0$.

Cette définition est purement "analytique", mais on peut la formuler de manière plus géométrique en considérant l'"hypersurface" d'équation $F(\lambda, x) = 0$:

LEMME 4.2. *Soit $\Gamma = \{(\lambda, u) \in \Lambda \times E;\ F(\lambda, u) = 0\}$. Dire que l'équation $F(\lambda, u) = 0$ définit implicitement u comme fonction $\mathcal{C}^1$ de λ au voisinage d'un point $(\lambda_0, u_0) \in \Gamma$ signifie qu'on peut trouver un voisinage $\mathcal{W}$ de (λ_0, u_0) tel que $\Gamma \cap \mathcal{W}$ est le graphe d'une fonction de classe $\mathcal{C}^1$.*

DÉMONSTRATION. Si $U \ni u_0$ et $V \ni \lambda_0$ sont comme dans la définition, on peut prendre $\mathcal{W} = V \times U$. Inversement, si $\mathcal{W}$ est un voisinage ouvert de (λ_0, u_0) tel que $\Gamma \cap \mathcal{W}$ est le graphe d'une fonction $\lambda \mapsto u(\lambda)$ de classe $\mathcal{C}^1$, il suffit de choisir des voisinages ouverts U et V de u_0 et λ_0 tels $V \times U \subset \mathcal{W}$.

□

Exemple 1. Soit $F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ la fonction définie par $F(x, y) = x^2 + y^2 - 1$.

Dans ce cas, $\Gamma = \{(x, y);\ F(x, y) = 0\}$ est le cercle de centre $(0, 0)$ et de rayon 1.

Géométriquement, on voit tout de suite (en dessinant le cercle) que l'équation $F(x, y) = 0$ définit implicitement y comme fonction $\mathcal{C}^1$ de x au voisinage de tout point $(x_0, y_0) \in \Gamma$ tel que $y_0 \neq 0$, i.e. $x_0 \neq \pm 1$: si on choisit un voisinage ouvert $\mathcal{W}$ de (x_0, y_0) ne contenant pas les points $(1, 0)$ et $(-1, 0)$, il est "clair" que $\mathcal{W} \cap \Gamma$ est le graphe d'une fonction $x \mapsto y(x)$ de classe $\mathcal{C}^1$. Analytiquement, on le justifie comme suit. Pour $y_0 > 0$, on peut prendre $V =]-1, 1[$ et $U = \{y \in \mathbb{R};\ y > 0\}$. Si $x \in]-1, 1[$, l'équation $F(x, y) = 0$ possède une unique solution $y(x) > 0$, à savoir $y(x) = \sqrt{1 - x^2}$, et la fonction $x \mapsto y(x)$ est de classe $\mathcal{C}^1$ sur $V =]-1, 1[$. Pour $y_0 < 0$, on peut prendre $V =]-1, 1[$ et $U = \{y \in \mathbb{R};\ y < 0\}$, et dans ce cas la solution est $y(x) = -\sqrt{1 - x^2}$.

Au voisinage des points $(1, 0)$ et $(-1, 0)$, il est tout aussi "clair" géométriquement que l'équation $F(x, y) = 0$ ne définit pas implicitement y comme fonction $\mathcal{C}^1$ de x : si

$\mathcal{W}$ est un voisinage ouvert de $(1,0)$ ou de $(-1,0)$, l'ensemble $\mathcal{W}\cap\Gamma$ n'est visiblement pas le graphe d'une fonction $x\mapsto y(x)$. Analytiquement, on peut le voir de 2 façons, par exemple pour le point $(1,0)$. Si V est un voisinage ouvert de $x_0=1$ et U est un voisinage ouvert de $y_0=0$, alors V contient à la fois des points x pour lesquels l'équation $F(x,y)=0$ ne possède aucune solution (n'importe quel $x>1$), et des points x pour lesquels l'équation possède 2 solutions (n'importe quel $x<1$ suffisamment proche de 1 pour que $\sqrt{1-x^2}$ et $-\sqrt{1-x^2}$ appartienne à U). En fait, il n'est même pas possible de définir une fonction $x\mapsto y(x)$ de classe $\mathcal{C}^1$ sur un intervalle $I=]\alpha,1]$ telle que $F(x,y(x))=0$ sur I. En effet, on aurait alors (en dérivant par rapport à x) $\frac{\partial F}{\partial x}(x,y(x))+y'(x)\frac{\partial F}{\partial y}(x,y(x))=0$ pour tout $x\in I$, autrement dit $2x+2y(x)\,y'(x)=0$; et comme $y(1)=0$ (puisque $1^2+y(1)^2-1=0$) on en déduirait une contradiction en prenant $x=1$.

On montrerait de même que l'équation $F(x,y)=0$ définit implicitement x comme fonction $\mathcal{C}^1$ de y au voisinage de tout point $(x_0,y_0)\in\Gamma$ tel que $x_0\neq 0$, mais pas au voisinage des points $(0,1)$ et $(0,-1)$.

Sur cet exemple, on *constate* que les deux seuls points de Γ au voisinage desquels l'équation $F(x,y)=0$ ne définit pas implicitement y comme fonction $\mathcal{C}^1$ de x (à savoir $(1,0)$ et $(-1,0)$) sont également les deux seuls points de Γ où la dérivée partielle $\frac{\partial F}{\partial y}=2y$ s'annule. Et de même, les deux seuls points de Γ au voisinage desquels l'équation ne définit pas implicitement x comme fonction $\mathcal{C}^1$ de y sont ceux pour lesquels $\frac{\partial F}{\partial x}=0$.

Exemple 2. Soit $F:\mathbb{R}^2\to\mathbb{R}$ définie par $F(x,y)=x^2-y^2$.

Dans ce cas, $\Gamma=\{(x,y);\ F(x,y)=0\}$ est la réunion des deux droites d'équations $y=x$ et $y=-x$.

On vérifie facilement que l'équation $F(x,y)=0$ définit implicitement y comme fonction de x et x comme fonction de y au voisinage de tout point $(x_0,y_0)\in\Gamma$ différent de $(0,0)$. (*Exercice*).

Au voisinage de $(0,0)$, l'équation $F(x,y)=0$ ne définit pas y comme fonction de x, et pas non plus x comme fonction de y. C'est évident géométriquement : pour tout voisinage $\mathcal{W}$ de $(0,0)$, l'ensemble $\mathcal{W}\cap\Gamma$ n'est certainement pas un graphe, car toute droite verticale ou horizontale passant par un point de $\mathcal{W}\cap\Gamma$ différent de $(0,0)$ et suffisamment proche de $(0,0)$ rencontre deux fois $\mathcal{W}\cap\Gamma$.

Sur cet exemple, on constate à nouveau que l'équation $F(x,y)=0$ définit implictement y comme fonction $\mathcal{C}^1$ de x au voisinage de tout point $(x_0,y_0)\in\Gamma$ tel que $\frac{\partial F}{\partial y}(x_0,y_0)=2y_0\neq 0$, et qu'elle définit implicitement x comme fonction $\mathcal{C}^1$ de y au voisinage de tout point $(x_0,y_0)\in\Gamma$ tel que $\frac{\partial F}{\partial x}(x_0,y_0)=2x_0\neq 0$.

Ce qu'on vient d'observer sur deux exemples est un phénomène très général : c'est exactement le contenu de ce qu'on appelle le **théorème des fonctions implicites**.

4.2. Une version "pour enfants". Sous sa forme générale, le théorème des fonctions implicites sera énoncé et démontré plus bas. La proposition suivante est un cas particulier, où on ne considère que des fonctions F de deux variables réelles, et à valeurs réelles (i.e. $\Lambda=E=E'=\mathbb{R}$). La preuve est élémentaire, mais cependant non-triviale et intéressante.

PROPOSITION 4.3. (théorème des fonctions implicites dans le plan)

Soit $F : \mathbb{R}^2 \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$. Si (x_0, y_0) est tel que $F(x_0, y_0) = 0$ et $\frac{\partial F}{\partial y}(x_0, y_0) \neq 0$, alors l'équation $F(x, y) = 0$ définit implicitement y comme fonction de x au voisinage de (x_0, y_0).

Démonstration. Fixons (x_0, y_0) vérifiant les hypothèss ci-dessus. On cherche des intervalles ouverts $I \ni x_0$ et $J \ni y_0$ tels que

(i) pour tout $x \in I$, il existe un unique $y = y(x) \in J$ tel que $F(x, y(x)) = 0$;
(ii) l'application $x \mapsto y(x)$ est de classe $\mathcal{C}^1$ sur I.

Supposons par exemple $\frac{\partial F}{\partial y}(x_0, y_0) > 0$. Par continuité de $\frac{\partial F}{\partial y}$, on peut alors trouver un voisinage ouvert $\mathcal{W}_0$ de (x_0, y_0) tel que $\frac{\partial F}{\partial y}(x, y) > 0$ sur $\mathcal{W}_0$; et on peut également supposer que $\mathcal{W}_0$ est de la forme $\mathcal{W}_0 = I_0 \times J_0$, où I_0 et J_0 sont des intervalles ouverts contenant respectivement x_0 et y_0. Alors la propriété suivante a lieu : pour tout $x \in I_0$ fixé, la fonction $y \mapsto F(x, y)$ est strictement croissante sur J_0.

Choisissons $\delta > 0$ de sorte que $[y_0 - \delta, y_0 + \delta] \subset J_0$. Comme $F(x_0, y_0) = 0$ et comme la fonction $y \mapsto F(x, y)$ est strictement croissante sur J_0, on a $F(x_0, y_0 - \delta) < 0$ et $F(x_0, y_0 + \delta) > 0$. Comme l'ensemble $\{x \in \mathbb{R};\ F(x, y_0 - \delta) < 0 \text{ et } F(x, y_0 + \delta) > 0\}$ est un ouvert de $\mathbb{R}$ par continuité de F, on en déduit qu'on peut trouver un intervalle ouvert $I \subset I_0$ contenant x_0 tel que $F(x, y_0 - \delta) < 0 < F(x, y_0 + \delta)$ pour tout $x \in I$. On va montrer que I et $J :=]y_0 - \delta, y_0 + \delta[$ conviennent.

Si $x \in I$, alors $F(x, y_0 - \delta) < 0 < F(x_0 + \delta)$. Comme la fonction $y \mapsto F(x, y)$ est continue et strictement croissante sur J_0 et donc sur $[y_0 - \delta, y_0 + \delta]$, on en en déduit (à l'aide du théorème des valeurs intermédiaires) que l'équation $F(x, y) = 0$ possède une unique solution $y(x) \in J$. Ainsi, (i) est vérifiée.

Admettons provisoirement que la fonction $x \mapsto y(x)$ est *continue* sur I, et montrons qu'elle est en fait de classe $\mathcal{C}^1$.

Fixons un point $x \in I$, et posons $a = \frac{\partial F}{\partial x}(x, y(x))$ et $b = \frac{\partial F}{\partial y}(x, y(x))$. Comme la fonction y est continue au point x, on peut écrire

$$y(x + h) = y(x) + \varepsilon(h)\,,$$

où $\varepsilon(h) \to 0$ quand $h \to 0$; et comme F est différentiable en $(x, y(x))$, on a donc

$$\begin{aligned} F(x + h, y(x + h)) &= F(x + h, y(x) + \varepsilon(h)) \\ &= F(x, y(x)) + a\,h + b\,\varepsilon(h) + R(h)\,, \end{aligned}$$

où $R(h) = o(|h| + |\varepsilon(h)|)$ quand $h \to 0$.

En se souvenant que $F(x, y(x)) = 0 = F(x+h, y(x+h))$ et comme $b = \frac{\partial F}{\partial y}(x, y(x)) \neq 0$ (car $y(x) \in I \subset J_0$) on en déduit

$$\varepsilon(h) = -\frac{a}{b}\,h + o(|h| + |\varepsilon(h)|)\,. \tag{4.1}$$

Pour h assez petit, on a donc en particulier $|\varepsilon(h)| \leq \frac{|a|}{|b|}\,|h| + \frac{1}{2}(|h| + |\varepsilon(h)|)$, autrement dit $|\varepsilon(h)| \leq 2\left(\frac{|a|}{|b|} + \frac{1}{2}\right)|h|$. Ainsi, $\varepsilon(h) = O(|h|)$ au voisinage de 0, d'où en revenant à (4.1) :

$$\varepsilon(h) = -\frac{a}{b}\,h + o(|h|)\,.$$

Par définition de $\varepsilon(h)$, cela signifie que la fonction y est dérivable au point x, avec $y'(x) = -\frac{b}{a}\cdot$ Ainsi, on vient de montrer que y est dérivable sur I avec

$$\forall x \in I \ : \ y'(x) = -\frac{\frac{\partial F}{\partial x}(x, y(x))}{\frac{\partial F}{\partial y}(x, y(x))}\,;$$

et comme la fonction y est continue, cette formule montre que y est en fait de classe $\mathcal{C}^1$.

Montrons maintenant que la fonction y est effectivement continue sur I. Soit (x_n) une suite de points de I convergeant vers $x_\infty \in I$, et supposons que $y(x_n)$ ne tende pas vers $y(x_\infty)$. Quitte à extraire une sous-suite, on peut supposer qu'on a $|y(x_n) - y(x_\infty)| \geq \varepsilon_0$ pour tout n et pour un certain $\varepsilon_0 > 0$. D'autre part, la suite $(y(x_n))$ est bornée car $y(x_n) \in J =]y_0 - \delta, y_0 + \delta[$. On peut donc trouver une sous-suite $y(x_{n_k})$ qui converge vers un certain $y_\infty \in [y_0 - \delta, y_0 + \delta]$. Par continuité de F, on a alors $F(x_\infty, y_\infty) = \lim F(x_{n_k}, y(x_{n_k})) = 0$; et comme la seule solution de l'équation $F(x_\infty, y) = 0$ dans l'intervalle $[y_0 - \delta, y_0 + \delta]$ est $y = y(x_\infty)$, on a donc $y_\infty = y(x_\infty)$. Ainsi, $y(x_{n_k})$ tend vers $y(x_\infty)$, d'où une contradiction puisque $|y(x_{n_k}) - y(x_\infty)| \geq \varepsilon_0 > 0$ pour tout k. On a donc bien montré que la fonction y est continue sur I. $\square$

Exercice. Vérifier que la démonstration précédente s'adapte au cas d'une fonction $F : \Lambda \times \mathbb{R} \to \mathbb{R}$, où Λ est un evn.

La proposition suivante donne une formule pour la dérivée de la fonction implicite. Il est inutile d'apprendre cette formule par coeur, mais *indispensable* de savoir la redémontrer rapidement.

Proposition 4.4. (dérivée de la fonction implicite)
Soit $F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ vérifiant les hypothèses du théorème des fonctions implicites au voisinage d'un point (λ_0, u_0), et soit $\lambda \mapsto u(\lambda)$ la fonction implicite associée. Alors $\frac{\partial F}{\partial u}(\lambda, u(\lambda))$ ne s'annule pas au voisinage de λ_0, et la dérivée de u est donnée par la formule

$$u'(\lambda) = -\frac{\dfrac{\partial F}{\partial \lambda}(\lambda, u(\lambda))}{\dfrac{\partial F}{\partial u}(\lambda, u(\lambda))}$$

Démonstration. Par continuité, $\frac{\partial F}{\partial u}(\lambda, u(\lambda))$ ne s'annule pas au voisinage de λ_0 puisque $\frac{\partial F}{\partial u}(\lambda_0, u(\lambda_0)) = \frac{\partial F}{\partial u}(\lambda_0, u_0) \neq 0$. En dérivant la relation $F(\lambda, u(\lambda)) = 0$ par rapport à λ, on obtient

$$\frac{\partial F}{\partial \lambda}(\lambda, u(\lambda)) + \frac{\partial F}{\partial u}(\lambda, u(\lambda)) \times u'(\lambda) = 0\,,$$

d'où la formule souhaitée. $\square$

Corollaire 4.5. *Sous les hypothèses précédentes, la fonction implicite u possède la même régularité que la fonction F : si F possède des dérivées partielles continues jusqu'à un certain ordre k, alors u est de classe $\mathcal{C}^k$.*

Démonstration. La formule donnant $u'(\lambda)$ montre que u' est de classe $\mathcal{C}^{k-1}$. $\square$

Exercice 1. Montrer que la relation $x^3y + 2x^4y^2 + x^2 + y = 0$ définit implicitement y comme fonction de classe $\mathcal{C}^\infty$ de x au voisinage de $(0, 0)$ et déterminer le développement limité à l'ordre 2 en 0 de la fonction $y(x)$.

Exercice 2. Soient $a, b \in \mathbb{R}$ avec $a < b$. Montrer que l'équation $(x-a)(x-b)+\varepsilon x^3 = 0$ définit implicitement x comme fonction $\mathcal{C}^\infty$ de ε au voisinage de $(0,a)$ et au voisinage de $(0,b)$. En déduire que pour $\varepsilon > 0$ assez petit, l'équation $(x-a)(x-b)+\varepsilon x^3 = 0 = 0$ possède trois solutions $x_1(\varepsilon) < x_2(\varepsilon) < x_3(\varepsilon)$, puis déterminer les développements limités à l'ordre 1 de $x_1(\varepsilon)$ et $x_2(\varepsilon)$ quand $\varepsilon \to 0^+$, et un développement asymptotique de $x_3(\varepsilon)$.

4.3. Le cas général. Passons maintenant à la version générale du théorème des fonctions implicites, où on revient au cas d'une fonction $F : \Lambda \times E \to E'$.

NOTATION. Pour toute fonction $F : \Lambda \times E \to E'$ pour $(\lambda_0, u_0) \in \Lambda \times E$ donné, on note $F_{\lambda_0} : E \to E'$ et $F^{u_0} : \Lambda \to E'$ les "fonctions partielles" définies par

$$F_{\lambda_0}(u) = F(\lambda_0, u) \quad \text{et} \quad F^{u_0}(\lambda) = F(\lambda, u_0)\,.$$

Si F est différentiable en (λ_0, u_0), les **différentielles partielles** de F au point (λ_0, u_0) sont les applications linéaires $D_\lambda F(\lambda_0, u_0) : \Lambda \to E'$ et $D_u F(\lambda_0, u_0) : E \to E'$ définies par

$$D_\lambda F(\lambda_0, u_0) = DF^{u_0}(\lambda_0) \quad \text{et} \quad D_u F(\lambda_0, u_0) = DF_{\lambda_0}(u_0)\,.$$

Le lemme suivant est l'analogue du théorème 2.5 du chapitre 3.

LEMME 4.6. *Si $F = \Lambda \times E \to E'$ est différentiable en (λ_0, u_0), alors $D_\lambda F(\lambda_0, u_0)$ et $D_u F(\lambda_0, u_0)$ s'identifient aux restrictions de $Df(\lambda_0, u_0)$ à $\Lambda \times \{0\} \simeq \Lambda$ et $\{0\} \times E \simeq E$. Pour $h = (h_\lambda, h_u) \in \Lambda \times E$, on a donc*

$$DF(\lambda_0, u_0)h = D_\lambda F(\lambda_0, u_0)h_\lambda + D_u F(\lambda_0, u_0)h_u\,.$$

DÉMONSTRATION. On a $F^{u_0} = F \circ I^{u_0}$ et $F_{\lambda_0} = F \circ I_{\lambda_0}$, où $I^{u_0} : \Lambda \to \Lambda \times E$ et $I_{\lambda_0} : E \times \Lambda \times E$ sont définies par $I^{u_0}(\lambda) = (\lambda, u_0) = (Id_\Lambda(\lambda), u_0)$ et $I_{\lambda_0}(u) = (\lambda_0, u) = (u_0, Id_E(u)$. Les applications Id_Λ et Id_E étant linéaires continues et les applications $\lambda \mapsto u_0$ et $u \mapsto \lambda_0$ constantes, on a $DI^{u_0}(\lambda_0)h_\lambda = (Id_\Lambda h_\lambda, 0) = (h_\lambda, 0)$ pour tout $h_\lambda \in \Lambda$, et $DI_{\lambda_0}(u_0)h_u = (0, h_u)$ pour tout $h_u \in E$. Par conséquent, $DF^{u_0}(\lambda_0)h_\lambda = DF(I^{u_0}(\lambda_0))(DI^{u_0}(\lambda_0)h_\lambda) = DF(\lambda_0, u_0)(h_\lambda, 0)$ et $DF_{\lambda_0}(u_0)h_u = DF(\lambda_0, u_0)(0, h_u)$. Cela prouve la première partie du lemme. La seconde partie en découle immédiatement : pour $h = (h_\lambda, h_u) \in \Lambda \times E$, on a

$$\begin{aligned} DF(\lambda_0, u_0)h &= DF(\lambda_0, u_0)(h_\lambda, 0) + DF(\lambda_0, u_0)(0, h_u) \\ &= D_\lambda F(\lambda_0, u_0)h_\lambda + D_u F(\lambda_0, u_0)h_u\,. \end{aligned}$$

□

THÉORÈME 4.7. (théorème des fonctions implicites)

Supposons les evn Λ, E et E' complets. *Soit $F : \Lambda \times E \to E'$ de classe $\mathcal{C}^1$ au voisinage d'un point (λ_0, u_0), avec $F(\lambda_0, u_0) = 0$. Si $DF_{\lambda_0}(u_0) : E \to E'$ est inversible, alors l'équation $F(\lambda, u) = 0$ définit implicitement u comme fonction $\mathcal{C}^1$ de λ au voisinage de (λ_0, u_0).*

Remarque 1. Si $\Lambda = E'$ et si F est de la forme $F(\lambda, u) = \Phi(u) - \lambda$ pour une certaine fonction $\Phi : E \to E'$, alors $DF_{\lambda_0} = D\Phi(u_0)$. On retrouve donc le théorème d'inversion locale.

Remarque 2. Bien entendu, il suffit de supposer que F est définie (et $\mathcal{C}^1$) seulement sur un voisinage de (λ_0, u_0).

PREUVE DU THÉORÈME. Soit $\Phi : \Lambda \times E \to \Lambda \times E'$ l'application définie par

$$\Phi(\lambda, u) = (\lambda, F(\lambda, u)).$$

Autrement dit, en écrivant $\Phi(\lambda, u)$ en colonne et en notant $\pi_\Lambda : \Lambda \times E \to \Lambda$ la "projection canonique" :

$$\Phi(\lambda, u) = \begin{pmatrix} \pi_\Lambda(\lambda,u) \\ F(\lambda,u) \end{pmatrix}.$$

L'application Φ est de classe $\mathcal{C}^1$, et pour $h \in \Lambda \times E$ écrit en colonne sous la forme $h = \begin{pmatrix} h_\lambda \\ h_u \end{pmatrix}$, on a

$$\begin{aligned} D\Phi(\lambda_0, u_0)h &= \begin{pmatrix} \pi_\Lambda h \\ DF(\lambda_0, u_0)h \end{pmatrix} \\ &= \begin{pmatrix} h_\lambda \\ D_\lambda F(\lambda_0, u_0)h_\lambda + D_u F(\lambda_0, u_0)h_u \end{pmatrix} \end{aligned}$$

Ainsi, $D\Phi(\lambda_0, u_0)$ s'écrit "matriciellement" sous la forme

$$D\Phi(\lambda_0, u_0) = \begin{pmatrix} Id_\Lambda & 0 \\ A & B \end{pmatrix}$$

où $A = D_\lambda F(\lambda_0, u_0) \in \mathcal{L}(\Lambda, E')$, $B = D_u F(\lambda_0, u_0) \in \mathcal{L}(E, E')$ et $0 \in \mathcal{L}(E, \Lambda)$.

Comme $B = D_u F(\lambda_0, u_0)$ est supposée inversible, on en déduit facilement que $D\Phi(\lambda_0, u_0)$ est inversible : ce serait évident dans le cas où $\Lambda = \mathbb{R}^p$ et $E = \mathbb{R}^q = E'$ en utilisant le déterminant jacobien puisque $J_\Phi(\lambda_0, u_0) = J_{F_{\lambda_0}}(u_0) \neq 0$; et dans le cas général on voit (*faire le calcul*) que $D\Phi(\lambda_0, u_0)^{-1}$ est donné matriciellement par

$$D\Phi(\lambda_0, u_0)^{-1} = \begin{pmatrix} Id_\Lambda & 0 \\ -B^{-1}A & B^{-1} \end{pmatrix}.$$

D'après le théorème d'inversion locale, Φ est un difféomorphisme local au point (λ_0, u_0). Comme $\Phi(\lambda_0, u_0) = (\lambda_0, 0)$, il existe donc un voisinage ouvert $\mathcal{U}$ de (λ_0, u_0) dans $\Lambda \times E$ tel que $\Phi_{|\mathcal{U}}$ est un difféomorphisme de $\mathcal{U}$ sur un ouvert $\mathcal{U}' \subset \Lambda \times E'$ contenant $(\lambda_0, 0)$.

Choisissons un voisinage ouvert V de λ_0 dans Λ et un voisinage ouvert U de u_0 dans E tels que $V \times U \subset \mathcal{U}$ et $V \times \{0\} \subset \mathcal{U}'$. Alors, pour $(\lambda, u) \in V \times U$, on a les équivalences suivantes :

$$\begin{aligned} F(\lambda, u) = 0 &\iff \Phi(\lambda, u) = (\lambda, 0) \\ &\iff (\lambda, u) = (\Phi_{|\mathcal{U}})^{-1}(\lambda, 0) \\ &\iff u = \pi_E\Big[(\Phi_{|\mathcal{U}})^{-1}(\lambda, 0)\Big], \end{aligned}$$

où $\pi_E : \Lambda \times E \to E$ est la projection canonique.

On en déduit que pour tout $\lambda \in V$, il existe un unique $u = u(\lambda) \in U$ tel que $F(\lambda, u) = 0$. Ce point $u(\lambda)$ est donné par la formule $u(\lambda) = \pi_E\big[(\Phi_{|\mathcal{U}})^{-1}(\lambda, 0)\big]$ et dépend donc de façon $\mathcal{C}^1$ de λ puisque $\Phi_{|\mathcal{U}}^{-1}$ est $\mathcal{C}^1$. □

COROLLAIRE 4.8. *Soient $F_1, \dots, F_n : \mathbb{R}^{m+n} \to \mathbb{R}$ des fonctions de classe $\mathcal{C}^1$ au voisinage d'un point $p_0 \in \mathbb{R}^{m+n}$ tel que $F_1(p_0) = \cdots = F_n(p_0) = 0$, et soient $s_1, \dots, s_n \in \{1, \dots, m+n\}$ avec $s_1 < \cdots < s_n$. Si la matrice $\left(\frac{\partial F_i}{\partial x_{s_j}}(p_0)\right)_{i,j=1}^n$ est* *inversible, alors le système d'équations*

$$\begin{cases} F_1(x_1, \dots, x_{m+n}) = 0 \\ \vdots \\ F_n(x_1, \dots, x_{m+n}) = 0 \end{cases}$$

définit implicitement $x_{s_1}, \dots, x_{s_n}$ *comme fonctions* $\mathcal{C}^1$ *des autres variables (au voisinage de* p_0*).*

Démonstration. Quitte à permuter les variables, on peut supposer que $s_1, \dots, s_n$ sont les n derniers indices, i.e. $s_1 = m+1, \dots, s_n = m+n$. En identifiant $\mathbb{R}^{n+m}$ à $\mathbb{R}^m \times \mathbb{R}^n$, écrivons $p_0 = (\lambda_0, u_0)$ et définissons $F : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$ par

$$F(\lambda, u) = \begin{pmatrix} F_1(\lambda,u) \\ \vdots \\ F_n(\lambda,u) \end{pmatrix}.$$

Alors la matrice $\left(\frac{\partial F_i}{\partial x_{s_j}}(p_0)\right)_{i,j=1}^{n}$ est précisément la matrice jacobienne de l'application $F_{\lambda_0} : \mathbb{R}^n \to \mathbb{R}^n$. Par conséquent, le résultat découle directement du théorème des fonctions implicites. □

Corollaire 4.9. *Soit* $F : \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}$ *de classe* $\mathcal{C}^1$*, et soit* $(\lambda_0, u_0) \in \mathbb{R}^p \times \mathbb{R}$ *tel que* $F(\lambda_0, u_0) = 0$. *Si* $\frac{\partial F}{\partial u}(\lambda_0, u_0) \neq 0$*, alors l'équation* $F(\lambda, u) = 0$ *définit implicitement* u *comme fonction* $\mathcal{C}^1$ *de* λ *au voisinage de* (λ_0, u_0)*.*

Remarque. Lorsque $p = 1$, i.e. pour $F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, il est très fréquent que l'on ne se souvienne plus de l'hypothèse permettant d'appliquer le théorème des fonctions implicites : doit on supposer $\frac{\partial F}{\partial u}(\lambda_0, u_0) \neq 0$ pour pouvoir exprimer u en fonction de λ, ou bien $\frac{\partial F}{\partial \lambda}(\lambda_0, u_0) \neq 0$? Pour s'en souvenir, on peut raisonner comme suit : on veut qu'il n'y ait, pour un λ donné qu'une seule solution à l'équation $F(\lambda, u) = 0$; donc il faut que $F(\lambda, u)$ varie lorsque u varie et que λ reste fixe ; et ceci est vrai si $\frac{\partial F}{\partial u}$ ne s'annule pas car alors la fonction $u \mapsto F(\lambda, u)$ est strictement monotone. (C'est exactement le raisonnement utilisé dans la preuve de la proposition 4.3).

Exemple. Continuité des racines d'un polynôme.

Soit $n \in \mathbb{N}^*$, et notons $\mathbf{ZS}_n$ l'ensemble de tous les polynômes P à coefficients réels et de degré n possédant n racines réelles distinctes. Pour $P \in \mathbf{ZS}_n$, notons $\lambda_1(P), \dots, \lambda_n(P)$ les racines de P rangées par ordre croissant. On va montrer que $\mathbf{ZS}_n$ *est un ouvert de* $\mathbb{R}_n[X]$ (muni de n'importe quelle norme), et que *l'application* $P \mapsto (\lambda_1(P), \dots, \lambda_n(P))$ *est de classe* $\mathcal{C}^1$ *sur* $\mathbf{ZS}_n$*.*

Dans la suite, on posera $\bar{\lambda}(P) = (\lambda_1(P), \dots, \lambda_n(P))$ pour $P \in \mathbf{ZS}_n$.

Soit $\mathcal{U} = \{\bar{\lambda} = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n;\ \lambda_1 < \cdots < \lambda_n\}$, et considérons l'application $F : \mathbb{R}_n[X] \times \mathcal{U} \to \mathbb{R}^n$ définie par

$$F(P, \bar{\lambda}) = (P(\lambda_1), \dots, P(\lambda_n)).$$

L'application F est visiblement de classe $\mathcal{C}^1$. Pour $P \in \mathbb{R}_n[X]$ fixé, la matrice jacobienne de l'application "partielle" $F_P : \mathcal{U} \to \mathbb{R}^n$ est diagonale :

$$\mathrm{Jac}_{F_P}(\lambda_1, \dots, \lambda_n) = \begin{pmatrix} P'(\lambda_1) & & \\ & \ddots & \\ & & P'(\lambda_n) \end{pmatrix}.$$

Si $P \in \mathbf{ZS}_n$, alors $\bar{\lambda}(P) \in \mathcal{U}$ et $P'(\lambda_j(P)) \neq 0$ pour tout j car $\lambda_j(P)$ est une racine nécessairement *simple* de P ; donc la matrice $\mathrm{Jac}_{F_P}(\bar{\lambda}(P))$ est inversible. D'après le théorème des fonctions implicites, on en déduit que pour tout $P_0 \in \mathbf{ZF}_n$, il existe un voisinage V_0 de P_0 et une application $\bar{\mu} : V_0 \to \mathcal{U}$ de classe $\mathcal{C}^1$ tels que $F(P, \bar{\mu}(P)) = 0$ pour tout $P \in V_0$. En posant $\bar{\mu}(P) = (\mu_1(P), \dots, \mu_n(P))$, cela signifie qu'on a $\mu_1(P) < \cdots < \mu_n(P)$ et $P(\mu_j(P)) = 0$ pour tout $j \in \{1, \dots, n\}$. Ainsi, on voit que si $P \in V_0$, alors $P \in \mathbf{ZS}_n$ et les $\mu_j(P)$ sont les racines de P rangées par ordre croissant ; autrement dit $\bar{\mu}(P) = \bar{\lambda}(P)$.

On vient de montrer que pour tout $P_0 \in \mathbf{ZS}_n$, on peut trouver un voisinage ouvert V_0 de P_0 tel que $V_0 \subset \mathbf{ZS}_n$ et l'application $\bar{\lambda}$ est de classe $\mathcal{C}^1$ sur V_0. En d'autres termes, $\mathbf{ZS}_n$ est un ouvert de $\mathbb{R}_n[X]$ et $\bar{\lambda}$ est de classe $\mathcal{C}^1$ sur $\mathbf{ZS}_n$.