24h de Calcul Différentiel

L3 Mathématiques fondamentales, Univ. Rennes

San Vũ Ngọc

BROUILLON — TRAVAIL EN COURS

Voici 24h d'introduction au calcul différentiel, dans son jus (pur sucre), tel qu'enseigné par l'auteur à l'université de Rennes entre 2023 et 2024.

UNIVERSITÉ DE RENNES

Version du 27 novembre 2024

Table des matières

Avant-propos

Il faut être honnête, ces notes de cours ne prétendent à aucune originalité dans leur contenu. On trouve sur internet de très bons cours de calcul différentiel (y compris en français) qui sont souvent bien plus complets. Seuls le choix des sujets abordés (avec la contrainte des 24 heures), et l'enchaînement des arguments, pourront témoigner d'une réflexion personnelle de l'auteur.

2024 — En travaux

Ce texte, destiné à aider les étudiantes et étudiants qui suivent le cours de Calcul différentiel (L3) à l'université de Rennes, n'est pas (encore ?) terminé ! Certaines sections sont très incomplètes. Dans ce cas, j'essaie au moins d'énoncer les résultats principaux. J'ai bon espoir que dans une avenir pas trop lointain, ce texte s'améliore (ou, au moins, se complète) et puisse servir de base à des étudiant·e·s qui n'auraient pas la possibilité d'être présent·es physiquement au cours. Pour ce faire, j'accepte bien entendu avec gratitude toutes les demandes de corrections et remarques (si possible constructives).

1. Introduction

1.1 Calcul différentiel

Historiquement, le *problème de la tangente* et le *problème de la quadrature* servent d'impulsion à développer le *calcul infinitésimal* au $17^{\text{ème}}$ siècle, et deviendront, respectivement, le *calcul différentiel* et le *calcul intégral*. Même si on cite souvent Isaac NEWTON (1643 — 1727) et Gottfried Wilhelm LEIBNIZ (1646 — 1716) comme inventeurs (ou découvreurs, selon votre philosophie de la science) du calcul infinitésimal, tout en mentionnant la controverse qui les a opposés sur la paternité de la théorie, il s'agit en fait d'un courant d'idées auquel de nombreux·ses mathématicien·nes ont contribué. En particulier, les idées de Leibniz et Newton s'appuient sur une découverte importante de Pierre de FERMAT (né entre 1601 et 1609, et mort en 1665). D'autre part, il a fallu attendre le $19^{\text{ème}}$ siècle, avec l'invention des nombres réels et de la notion de limite, pour que le calcul infinitésimal devienne une théorie rigoureuse acceptée par tous. Avant cela, les mathématicien·nes manipulaient les "infiniment petits" de façon efficace mais en étant conscients du (et gênés par le) fait qu'ils n'avaient pas de moyen de définir proprement les objets utilisés.

De nos jours, le calcul infinitésimal, qu'on associe souvent à la naissance de l'analyse mathématique, est en fait présent dans l'ensemble des branches des mathématiques, et est devenu un outil incontournable dans toutes les sciences et technologies.

Bien sûr, vous connaissez déjà le calcul des dérivées et des intégrales, depuis le lycée et les deux premières années d'université (ou équivalentes), pour des fonctions d'une variable réelle, à valeurs réelles (voire, peut-être, complexes). Le but de ce cours est de placer ces connaissances dans un cadre plus général afin d'ouvrir grandement le champ des applications. On étudiera la notion de dérivée pour des fonctions définies entre des espaces bien plus "gros" que $\mathbb{R}$, potentiellement de dimension infinie. Une suite logique de ce cours serait la *géométrie différentielle*, qu'on abordera discrètement au dernier chapitre.

À lire ces lignes, on pourrait penser que nous allons simplement apprendre à *calculer des dérivées*, ce qui semble (peut-être) peu enthousiasmant. J'espère vous rassurer et vous motiver en affirmant que, pour ce faire, nous aurons besoin d'un savant mélange de topologie, d'analyse, de géométrie, et aussi d'algèbre n !

1.2 Présentation enseignants

— San Vũ Ngọc
— Karim Bekka

1.3 Modalités de contrôle

— 2 QCM de 30m en fin de cours (QCM_1, QCM_2) Oui, je pense qu'apprendre par cœur les énonces principaux est utile, ça permet de mettre à profit notre «système 1» [5].
— 1 Contrôle 1h en cours (C_1)
— 1 Contrôle 2h en amphi à prévoir (T)
— 14 Semaines : (soit 21h de cours, 21h de TD)

Soit CC la moyenne des 3 premiers contrôles :

$$\mathrm{CC} = (\mathrm{QCM}_1 + \mathrm{QCM}_2 + 2 \times \mathrm{C}_1)/4$$

Note finale :

$$N = \max(T; (\mathrm{CC} + T)/2; \min(10, \mathrm{CC}))$$

Autrement dit, si on a 10 au CC, on est assuré de valider le module, le T est du bonus. C'est bien entendu une mauvaise idée de raisonner ainsi, car les notions qui sont abordées en fin de cours et évaluées par l'examen T sont les plus importantes pour la suite de votre carrière mathématique !

Les règles suivantes seront appliquées en cas d'absences :

— en cas d'absence injustifiée à un contrôle la note 0 est attribuée à ce contrôle ;
— les absences justifiées donnent lieu à une « neutralisation » de la note, càd :
    — en cas d'absence justifiée aux 3 premiers contrôles on attribue la note $\mathrm{CC} = 0$;
    — en cas d'absence justifiée à un ou deux des 3 premiers contrôles on attribue aux contrôles non faits la seule note obtenue ou la moyenne des 2 notes obtenues ;
    — en cas d'absence justifiée au troisième contrôle on attribue à ce contrôle la note CC.

1.4 Plan du cours

1. Fonctions d'une variable à valeur dans un Espace Affine normé
2. Interlude : applications linéaires continues
3. Différentielle d'une application $f : E \to F$
4. Chemins dans un Banach
5. Difféomorphismes et inversion locale
6. Différentielles d'ordre supérieur
7. Extrema et sous-variétés

1.5 Références

Ce texte ne suit pas une référence précise, mais a été inspiré par plusieurs ouvrages. On pourra consulter par exemple, en cas de doute ou de désir de détails,

— *Calcul différentiel*, Christol, Gilles & Cot, Anne & Marle, Charles-Michel [4]

— *Calcul Différentiel et Calcul Intégral*, Marc Chaperon [3]
— (Ouvrage de référence !) *Cours de calcul différentiel*, Henri Cartan [2]

Prérequis Réviser les fonctions $\mathbb{R} \to \mathbb{R}$!

2. Fonctions d'une variable à valeurs dans un EVN (ou EAN)

Qu'est-ce qu'une fonction ?

FIGURE 2.1 – Paul Gauguin, *D'où venons-nous Que sommes-nous Où allons-nous* — Museum of Fine Arts Boston, Domaine public, https://commons.wikimedia.org/w/index.php?curid=36264337

À quoi sert une fonction ?
Fonction ou application ?
Observer $X \to \mathbb{R}$,
Transformer, $X \to X$
Transporter $X \to Y$
Dessiner (ou cheminer) $\mathbb{R} \to X$

> « Comment s'étaient-ils rencontrés ? Par hasard, comme tout le monde. Comment s'appelaient-ils ? Que vous importe ? D'où venaient-ils ? Du lieu le plus prochain. Où allaient-ils ? Est-ce que l'on sait où l'on va ? » (*Jacques le Fataliste*, Diderot)

Dans ce chapitre on va commencer doucement avec $f : \mathbb{R} \to X$ (c'est plus facile que $f : X \to \mathbb{R}$). Par exemple, si $X = \mathbb{R}^2$, c'est juste deux fonctions $f = (f_1, f_2)$ Si $X = \mathbb{R}^n$, on a n fonctions. Si X est un espace vectoriel de dimension infinie, on a quoi ?

But du chapitre : $f : \mathbb{R} \to E$: (fonction *vectorielle*) qu'est-ce qui change par rapport à $\mathbb{R} \to \mathbb{R}$ (fonction *scalaire*). Qu'est-ce qui marche, ne marche pas ?

■ **Exemple 2.1** Une suite réelle $(u_n)(t)$ dépendant d'un paramètre réel t est une fonction de $\mathbb{R}$ dans $\mathbb{R}^{\mathbb{N}}$. ■

2.1 Rappels sur les espaces vectoriels normés et les espace affines

Définition 2.2 Un **espace vectoriel** est un ensemble muni d'une loi de groupe commutatif (addition) (donc automatiquement non vide) et d'une multiplication externe par un corps $\mathbb{K}$ (également commutatif, ici on prendra $\mathbb{K} = \mathbb{R}$), de façon *compatible*. Précisément, notons E cet ensemble. L'addition de deux éléments $v, w \in E$ sera notée $v + w$; la multiplication d'un élément $v \in E$ par un scalaire $\lambda \in \mathbb{K}$ sera notée *à gauche* : $\lambda \cdot v$; les conditions de compatibilité sont :

— $0 \cdot u = 0$,
— $1 \cdot v = v$.
— $\lambda \cdot (\mu \cdot u) = (\lambda \mu).u$,
— $(\lambda + \mu) \cdot u = \lambda \cdot u + \mu \cdot u$,
— $\lambda \cdot (u + v) = \lambda \cdot u + \lambda \cdot v$,

De façon informelle, un ensemble est une collection d'objets ; ces objets sont appelés les *éléments* de cet ensemble. Dans le cas d'un espace vectoriel E, ses éléments sont en principe appelés des **vecteurs**. Néanmoins, on pourra aussi les appeler des **points**. Pourquoi donc ? Cela vient du fait qu'un espace vectoriel peut aussi être considéré comme un *espace affine*, voir ci-dessous. Pour bien comprendre le calcul différentiel, il sera important (et très pratique !) de savoir distinguer à quel moment les objets qu'on manipule sont des vecteurs, et à quel moment ce sont des points.

Le corps $\mathbb{K}$ s'appelle le *corps des scalaires*. Un espace vectoriel sur le corps $\mathbb{R}$ sera appelé un espace vectoriel *réel*.

Un espace vectoriel seul n'est pas suffisant pour faire du calcul différentiel. Nous aurons besoin également d'une *norme*.

Définition 2.3 Soit E un espace vectoriel sur un corps $\mathbb{K}$ muni d'une valeur absolue (si $\mathbb{K} = \mathbb{R}$, on prendra la valeur absolue usuelle, et si $\mathbb{K} = \mathbb{C}$, on prendra le module). Une **norme** sur E est une application $\|\cdot\| : E \to \mathbb{R}$ telle que :

(N1) $\|v\| = 0$ si et seulement si $v = 0$ *[séparation]*;
(N2) $\|\lambda \cdot v\| = |\lambda| \, \|v\|, \forall v \in E, \forall \lambda \in \mathbb{K}$, *[homogénéité]*;
(N3) $\|v + w\| \leqslant \|v\| + \|w\|, \forall v, w \in E$. *[inégalité triangulaire]*

Alors, $(E, \|.\|)$ est un **espace vectoriel normé** (evn).

Exercice 2.4 Montrer qu'une norme est toujours à valeurs dans $[0, +\infty[$. ■

Définition 2.5 Un **espace affine**, est, de façon très abstraite, un ensemble X (dont les éléments sont appelés des *points*) muni d'une "action libre et transitive d'un espace vectoriel E (vu comme groupe additif)".

R

Remarque 2.6 On pourra écrire « Soit X un espace affine » ou bien « Soit (X,E) un espace affine », selon qu'on veuille ou non introduire la notation E pour l'espace vectoriel associé.

En fait, on va pouvoir très rapidement oublier cette définition, car en fin de compte, c'est très simple. Commençons par expliciter les termes. Une *action* de E sur X, ça veut dire que pour tout point $x \in X$ et tout vecteur $v \in E$, on *peut construire un autre point* qu'on *note* [1] $x+v$. Et si on part du point $x+v$ et qu'on "agit" avec le vecteur $-v$, on retombe sur le point initial x. L'associativité habituelle est valable : $(x+v_1)+v_2 = x+(v_1 +_E v_2)$. (On a noté ici "$+_E$" la loi de groupe de E (addition de deux vecteurs), afin de la distinguer de l'action de E sur X, notée "$+$" ; dans toute la suite, on utilisera malheureusement la même notation $+$ pour les deux lois, comme c'est la coutume).

Le fait que l'action soit *libre* dit que la seule façon de laisser un point x inchangé en ajoutant un vecteur v c'est de choisir le vecteur nul $v = 0$.

Enfin, le fait que l'action soit *transitive* dit que pour tout couple de points $(x,y) \in X \times X$, il existe un vecteur $v \in E$ qui permet de passer de l'un à l'autre : $x+v = y$.

Exercice 2.7 Montrer que, étant donnés $x,y \in X$, il existe un *unique* $v \in E$ tel que $x+v = y$. ▪

Exercice 2.8 Tout espace vectoriel est aussi un espace affine (il "agit" sur lui-même !). ▪

Ceci explique que lorsqu'on travaille avec un espace vectoriel, on peut voir ses éléments comme des *points* ou des *vecteurs*, selon le point de vue adopté.

La proposition suivante montre que, finalement, un espace affine c'est juste "un espace vectoriel et un point origine". En pratique, c'est cette proposition qu'on va utiliser, et on peut oublier la définition 2.5.

Proposition 2.9 Soit X un espace affine, et E l'espace vectoriel associé. Pour tout point $x_0 \in X$ on a l'égalité

$$X = \{x_0\} + E$$

(On note $\{x_0\}+E$ (ou parfois simplement x_0+E) l'ensemble des points de X de la forme x_0+v, avec $v \in E$.)

Démonstration. Procédons par double inclusion.

1. Si $x \in X$, alors par transitivité (pour la paire (x_0,x)), il existe $v \in E$ tel que $x = x_0+v$. Donc $x \in \{x_0\}+E$.
2. Réciproquement, si $x \in \{x_0\}+E$, ça veut dire $x = x_0+v$ pour un certain $v \in E$, et par définition $x \in X$.

■

Définition 2.10 Soit X un espace affine, et E l'espace vectoriel associé. On dira que E est **l'espace tangent à** X.

Si E est un espace vectoriel, qu'on voit comme un espace affine, alors son espace tangent est (bien sûr) égal à E lui-même.

1. Notez l'ordre : *point* + *vecteur*. On ne doit **pas** écrire *vecteur* + *point*.

■ **Notation 2.11 — différence de deux points.** On a vu qu'étant donné un espace affine X, d'espace tangent E, si on se donne deux points x et y dans X, il existe un unique vecteur $v \in E$ tel que $y = x + v$. Il est donc naturel de noter ce vecteur

$$v = y - x.$$

On peut maintenant dire que, dans un espace affine, les vecteurs de l'espace vectoriel associé sont simplement les *différences de deux points*.

D'ailleurs, vous connaissez certainement le cas où X est un sous-ensemble d'un espace vectoriel F : par exemple une droite affine dans $\mathbb{R}^2$. Dans ce cas, la différence $y - x$ est bien égale à la différence de vecteurs provenant du grand espace F. Ainsi, sur une droite affine, la différence de deux point donne un vecteur directeur de la droite (Figure 2.2).

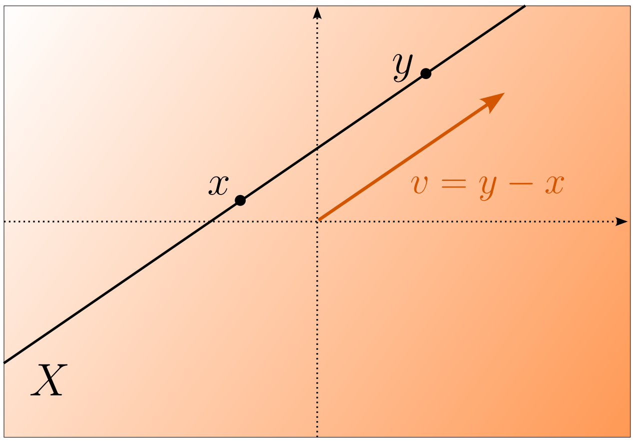

FIGURE 2.2 – Droite affine. La différence des deux points x et y est le vecteur v.

Définition 2.12 Un **espace affine normé** est un espace affine dont l'espace tangent est un espace vectoriel normé.

Exercice 2.13 Tout espace vectoriel normé est naturellement un espace affine normé. ■

■ **Exemple 2.14** $\mathbb{R}$ muni de la norme "valeur absolue" est un EVN (espace vectoriel normé) de dimension 1, sur le corps $\mathbb{K} = \mathbb{R}$. Dans tout ce cours, $\mathbb{R}$ sera toujours muni de cette norme dite "standard". Dans le cours de Topologie vous pourrez vous amuser à munir $\mathbb{R}$ d'autres topologies (et donc, d'autres notions de continuité), mais pour le cours de calcul différentiel on ne s'autorisera pas ces fantaisies ! ■

■ **Exemple 2.15** Le produit cartésien $\mathbb{R}^n$ (qui est un espace vectoriel réel de dimension n, avec les opérations habituelles) peut aussi être muni d'une norme, mais pour $n \geqslant 2$ il n'y a pas de norme standard ; on prendra en général la norme $\|(x_1,\dots,x_n)\|_\infty := \max\{|x_1|,\dots,|x_n|\}$, ou encore la norme euclidienne $\|(x_1,\dots,x_n)\|_2 := \sqrt{\sum_{i=1}^n |x_i|^2}$. ■

■ **Exemple 2.16** Il existe de nombreux espaces vectoriels de dimension *infinie*. Par exemple l'espace des suites réelles $\mathbb{R}^{\mathbb{N}}$ (ce sont les fonctions de $\mathbb{N}$ dans $\mathbb{R}$). On peut essayer de le munir de l'analogue des normes de l'exemple 2.15 :

$$\left\|(x_n)_{n\in\mathbb{N}}\right\|_\infty = \sup_n |x_n| \quad \text{et} \quad \left\|(x_n)_{n\in\mathbb{N}}\right\|_2 := \sqrt{\sum_n |x_n|^2},$$

cependant ces “normes” ne sont pas bien définies (elles peuvent prendre la valeur $+\infty$). On va donc introduire le *sous-espace* $\ell_\infty(\mathbb{R})$ des suites $x \in \mathbb{R}^{\mathbb{N}}$ pour lesquelles $\|x\|_\infty < +\infty$, ainsi que le sous-espace $\ell_2(\mathbb{R})$ des suites $x \in \mathbb{R}^{\mathbb{N}}$ pour lesquelles $\|x\|_2 < +\infty$. (Il faut montrer que ce sont bien des sous-espaces vectoriels, c'est un exercice classique.) Ceci fait, les espaces $\ell_2(\mathbb{R})$ et $\ell_\infty(\mathbb{R})$ deviennent bien entendu des EVN. ■

■ **Exemple 2.17** L'ensemble X des fonctions continues $f : [0,1] \to \mathbb{R}$ d'intégrale égale à 1 :

$$\int_0^1 f(t)\,\mathrm{d}t = 1$$

n'est *pas* un espace vectoriel ; mais c'est un espace *affine*. Puisqu'il est contenu dans l'espace vectoriel de toutes les fonctions continues sur $[0,1]$, pour trouver son espace tangent E, il suffit de calculer les différences $f-g$ pour $f,g \in X$: ce sont des fonctions continues d'intégrale nulle. Réciproquement, si on fixe un fonction $f_0 \in X$, il est facile de voir que $X = \{f_0\} + E$, ou E est l'espace (vectoriel !) des fonctions continues d'intégrale nulle, ce qui montre que X est bien affine. ■

Définition 2.18 Soient X,Y des espace affines et $f : X \to Y$ une application. On dit que f est **affine** si elle transporte les vecteurs (les différences de deux points) en des vecteurs par une application linéaire. Précisément : soient E,F les espaces tangents respectifs de X,Y. L'application f est affine s'il existe une application linéaire $A : E \to F$ telle que

$$\forall x_1, x_2 \in X, \quad f(x_1) - f(x_2) = A(x_1 - x_2)\,.$$

On en déduit immédiatement la forme habituelle des applications affines : pour tout $x_0 \in X$ fixé, f est définie par

$$\forall x \in X, \quad f(x) = f(x_0) + A(x - x_0)\,.$$

Dans toute la suite de ce cours, si la notion d'espace affine n'est pas suffisamment claire pour vous, vous pouvez sans danger remplacer tous les espaces affines par des espaces vectoriels. Ceci dit, si vous arriver à digérer cette notion d'espace affine, vous constaterez que toutes les définitions du cours deviendront plus naturelles, et les risques de se tromper diminueront fortement.

Pour aller plus loin avec les espaces affines (en lien avec le calcul différentiel), on pourra consulter [6].

2.2 Chemins

La notion de *chemin* dans un espace X apparaît dans de nombreuses branches des mathématiques, parfois avec des définitions légèrement différentes ; mais le point fondamental c'est qu'il s'agit d'une application *continue* f d'un *intervalle* I de $\mathbb{R}$ à valeurs dans X :

$$f : I \to X\,.$$

Évidemment, du coup, la question principale est : qu'est-ce que la *continuité* dans X ? Munir un ensemble X d'une notion de continuité revient à faire de X un *espace topologique*. Il convient de suivre un cours dédié à la Topologie. Dans notre cours de Calcul différentiel, nous utiliserons la caractérisation suivante d'une application continue dans un espace affine (ou vectoriel) *normé* :

Définition 2.19 Soit X un espace affine normé. Un **chemin** dans X est une application $f : I \to X$, où I est un intervalle de $\mathbb{R}$, qui est continue au sens où :

$$\forall t \in I, \quad \forall \varepsilon > 0, \quad \exists \alpha > 0 \quad \text{tel que :} \tag{2.1}$$

$$\text{si } s \in I \text{ est tel que } |t - s| < \alpha, \text{ alors } \|f(t) - f(s)\| < \varepsilon \tag{2.2}$$

Remarquons que la différence $f(t) - f(s)$ est bien un vecteur (voir la notation 2.11), donc sa norme est bien définie. Par contre, la "norme de $f(t)$" n'a pas de sens ! (à moins de fixer une origine x_0 et d'utiliser la proposition 2.9).

Si $X = \mathbb{R}$, on retrouve bien sûr la notion de fonction continue que vous connaissez depuis la L2.

Définition 2.20 On appelle (parfois) un **arc** un chemin défini sur un intervalle fermé borné $[a, b]$. Voir Figure 2.3.

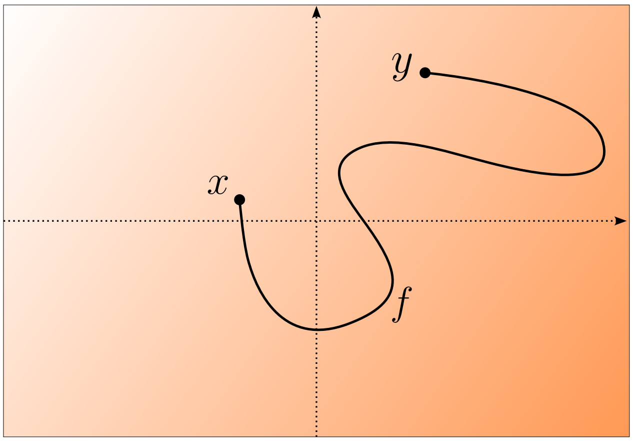

FIGURE 2.3 – Image (ou tracé) d'un chemin (ou arc) f dans le plan $\mathbb{R}^2$. Les points x et y sont les *extrémités* du chemin : si f est défini sur l'intervalle $[a, b]$, alors $x = f(a)$ et $y = f(b)$. Dans ce cours, on ne demandera pas nécessairement $a = 0$ et $b = 1$, mais bien sûr c'est facile de s'y ramener par un changement de variables affines dans $\mathbb{R}$, en conservant le même tracé.

On rappelle que si $f : I \to \mathbb{R}$ est une fonction continue, alors $f([a, b])$ est un intervalle compact, en particulier f atteint ses bornes. En cours de Topologie on montre que pour un espace topologique X séparé général, si $f : I \to X$ est continue, alors $f([a, b])$ est compact, mais on n'a pas la notion de "bornes". Si X est un espace vectoriel ou affine normé, alors l'ensemble $f([a, b])$ est "borné" au sens où il est contenu dans une boule, mais on n'a toujours pas de notion de "bornes" (sup et inf) car X n'est en général pas *ordonné*.

2.3 Tangente à un chemin différentiable

Nous voila prêts à introduire les premières notions de calcul infinitésimal. Commençons par les très utiles (mais subtiles) *notations de Landau*[2] : o (petit "o") et $\mathcal{O}$ (grand "o").

2. On dit aussi "Bachmann-Landau". Paul Bachmann (1837 — 1920) et Edmund Landau (1877 — 1938) sont deux mathématiciens allemands.

Définition 2.21 — petit o. Soient E, F des espaces vectoriels normés (en général, on prendra $F = \mathbb{R}$). Soit I un intervalle de $\mathbb{R}$ et $t_0 \in I$. Soient $f : I \setminus \{t_0\} \to E$, $g : I \setminus \{t_0\} \to F$ et $t_0 \in I$. Alors

$$f = o(g) \text{ en } t_0 \qquad (\text{ou bien } f(t) = o(g(t)) \text{ quand } t \to t_0)$$

(on dit "f est petit o de g en t_0" ou bien "$f(t)$ est petit o de $g(t)$ lorsque t tend vers t_0") s'il existe une fonction $\varepsilon(t) : I \setminus \{t_0\} \to \mathbb{R}$ *qui tend vers* 0 *en* t_0 telle que

$$\|f(t)\|_E = \varepsilon(t)\, \|g(t)\|_F \text{ près de } t_0\,.$$

Si on écrit tout ça avec des quantificateurs, on obtient, lorsque $t_0 \in \mathbb{R}$:

Définition 2.22 $f = o(g)$ en t_0 si

$$\forall \varepsilon > 0, \exists \alpha > 0, \quad \forall t \in I \setminus \{t_0\}, |t - t_0| < \alpha \Rightarrow \|f(t)\| \leqslant \varepsilon \|g(t)\|\,.$$

Exercice 2.23 On suppose que la fonction g ne s'annule pas au voisinage de t_0. Montrer qu'on a la caractérisation (utile, en pratique) :

$$(f = o(g) \text{ en } t_0) \Longleftrightarrow \left(\frac{\|f(t)\|_E}{\|g(t)\|_F} \to 0 \text{ lorsque } t \to t_0, t \neq t_0 \right).$$

■

■ Exemple 2.24

— $I = [0, +\infty[$, $t_0 = 0$: on a $\ln t = o(\frac{1}{t})$ en $t = 0$ car $t \ln t \to 0$ quand $t \to 0^+$.
— $I = \mathbb{R}$, $t_0 = 0$: on a $\sin(t) - t = o(t^2)$ en $t = 0$ car $\frac{\sin(t) - t}{t^2} \to 0$ lorsque $t \to 0$, $t \neq 0$.

■

Remarque 2.25 Montrer que $f = o(g)$ si et seulement si $f = o(\|g\|)$, qui est lui-même équivalent à $\|f\| = o(g)$, ou encore à $\|g\| = o(\|g\|)$. Ne vous étonnez donc pas si l'on passe, au grès de l'humeur, de l'une à l'autre de ces notations !

La plupart des ouvrages ont une définition plus restrictive et demandent que g soit à valeurs réelles. Autrement dit, ils utilisent donc toujours la forme $f = o(\|g\|)$. Je pense cependant qu'il est pratique (et sans risque) d'utiliser la forme générale de la définition 2.22.

Remarque 2.26 Le symbole $=$ est utilisé de façon *abusive* dans la notation $f = o(g)$. En effet, $t = o(1)$ en $t = 0$, et aussi $t^2 = o(1)$, mais $t \neq t^2$! Il serait plus raisonnable de considérer $o(g)$ comme une classe de fonctions, et de noter $f \in o(g)$; je le ferai parfois dans ce cours. Cependant, l'utilisation abusive du symbole $=$ est déjà largement passée dans le pratique courante.

Remarque 2.27 Dans la lignée de la remarque 2.26 ci-dessus, on s'autorisera (car c'est très pratique !) à écrire

$$f_1 = f_2 + o(g)$$

pour signifier

$$f_1 - f_2 = o(g)\,.$$

R

Remarque 2.28 On peut, dans la définition 2.21, remplacer I par n'importe quel sous-ensemble de $\mathbb{R} \cup \{-\infty, +\infty\}$, et t_0 un point d'accumulation de I. Dans ce cas, dans la formule de la définition 2.22, lorsque $t_0 = +\infty$, il faut remplacer $|t - t_0| < \alpha$ par $t > \alpha$; et lorsque $t_0 = -\infty$, par $t < -\alpha$.

L'intérêt est d'utiliser la même notion pour les suites : on prend $I = \mathbb{N} \cap \{+\infty\}$, $t_0 = +\infty$, et une suite u_n est une fonction $n \mapsto u_n = f(n)$ de I dans $\mathbb{R}$ (ou dans un espace affine normé). On peut alors écrire par exemple $n^4 = o(e^n)$ lorsque $n \to \infty$.

Exercice 2.29 Si $f_1 = o(g)$ et $f_2 = o(g)$, alors $f_1 + f_2 = o(g)$. ■

Définition 2.30 — grand o. Soient E, F des espaces vectoriels normés (en général, on prendra $F = \mathbb{R}$). Soit I un intervalle de $\mathbb{R}$ et $t_0 \in I$. Soient $f : I \setminus \{t_0\} \to E$, $g \setminus \{t_0\} : I \to F$ et $t_0 \in I$. Alors

$$f = \mathcal{O}(g) \text{ en } t_0$$

(on dit "f est grand o de g en t_0") s'il existe une fonction $b(t) : I \setminus \{t_0\} \to \mathbb{R}$ *bornée près de* t_0 telle que

$$\|f(t)\|_E = b(t)\, \|g(t)\|_F \text{ près de } t_0\,.$$

Avec des quantificateurs :

Définition 2.31 $f = \mathcal{O}(g)$ en t_0 si :

$$\exists C > 0, \alpha > 0, \quad \forall t \in I \setminus \{0\}, |t - t_0| \leqslant \alpha \Rightarrow \|f(t)\|_E \leqslant C\, \|g(t)\|_F\,.$$

Exercice 2.32 On suppose que la fonction g ne s'annule pas au voisinage de t_0, sauf peut-être en t_0 lui-même. Montrer qu'on a la caractérisation (utile, en pratique) : $f = \mathcal{O}(g)$ en t_0 si la fonction $t \mapsto \frac{\|f(t)\|_E}{\|g(t)\|_F}$ est *bornée* au voisinage (épointé) de t_0. ■

■ **Exemple 2.33** Soit X un espace affine normé et $f : \mathbb{R} \to X$ une application affine. Alors $f(t) - f(t_0) = \mathcal{O}(t - t_0)$ en $t = t_0$. En effet, on a (voir la définition 2.18) $f(t) = f(t_0) + A(t - t_0)$ où $A : \mathbb{R} \to E$ est linéaire, et E est l'espace vectoriel associé à X. Notons $c = A(1)$; par linéarité, $A(s) = s \cdot A(1) = s \cdot c$, pour tout $s \in \mathbb{R}$. Donc

$$\|f(t) - f(t_0)\| = \|(t - t_0) \cdot c\| = |t - t_0|\, \|c\|\,,$$

ce qui montre que $\frac{\|f(t) - f(t_0)\|}{|t - t_0|}$ est borné au voisinage de t_0 (et même partout). ■

Les remarques 2.25 et 2.26, ainsi que 2.28 s'appliquent aussi à la notation "grand o".

> **Attention,** *il faut parfaitement maîtriser ces notations de Landau pour profiter pleinement de la suite du cours.*

Définition 2.34 Soit E un EVN. On dit que $f : I \to E$ est **tangente à zéro** en t_0 si $f(t) = o(t - t_0)$ en $t = t_0$.

En particulier, $\lim_{\substack{t \to t_0 \\ t \neq t_0}} f(t) = 0$. Donc, d'une part, si $f(t_0) = 0$, alors f est continue en t_0, et d'autre part si f est continue en t_0, alors $f(t_0) = 0$. Intuitivement, pour des fonctions f continues (les chemins), les f tangentes à 0 en t_0 sont les chemins qui passent par l'origine en $t = t_0$ avec une "tangente horizontale".

Exercice 2.35 Montrer que les applications linéaires (ou affines) tangente à zéro en t_0 sont les applications nulles. ■

Exercice 2.36 Donner un exemple de $f : \mathbb{R} \to \mathbb{R}$ tangente à 0 en 0, non nulle. ■

Cette notion permet de définir une relation d'équivalence entre chemins :

Définition 2.37 Deux chemins f et $g : I \to X$, où X est affine, sont dits **tangents entre eux** en $t_0 \in I$ si $f - g$ est tangent à zéro en t_0.

Comment vérifier que deux chemins sont tangents ? Comme X est affine on dispose de chemins particuliers : les chemins affines (dont l'image est appelée une **droite affine**) :

$$a(t) = x_0 + t \cdot v,$$

où $x_0 \in X$ et $v \in E \setminus \{0\}$. Le vecteur $v \in E$ est un vecteur directeur de la droite. Ainsi, si f et g sont deux chemins tangents entre eux, et si chacun de son côté est tangent à un chemin affine, alors ces deux chemins affines sont tangents entre eux (transitivité). L'exercice 2.35 montre que deux chemins affines tangents entre eux doivent être égaux : il s'agit donc forcément de la même droite pour les deux, avec la même paramétrisation. En particulier, elles ont le même vecteur v. Donnons-lui un nom :

Définition 2.38 **Le vecteur vitesse** du chemin f en t_0 (s'il existe) est un vecteur $v_0 \in E$ tel que f est tangente (en t_0) au chemin affine $t \mapsto f(t_0) + (t - t_0) \cdot v_0$

Tout vecteur colinéaire au vecteur vitesse s'appelle un vecteur tangent au chemin f ; autrement dit :

Définition 2.39 $v \in E$ est **un vecteur tangent** à f en $f(t_0)$ si f est tangente (en t_0) à une droite affine de vecteur directeur colinéaire à v.

Évidemment un vecteur directeur d'une droite est lui-même un vecteur tangent à cette droite, c'est essentiellement la même notion, excepté pour le fait qu'un vecteur tangent peut être nul, ce qui n'est pas le cas d'un vecteur directeur. Voir Figure 2.4.

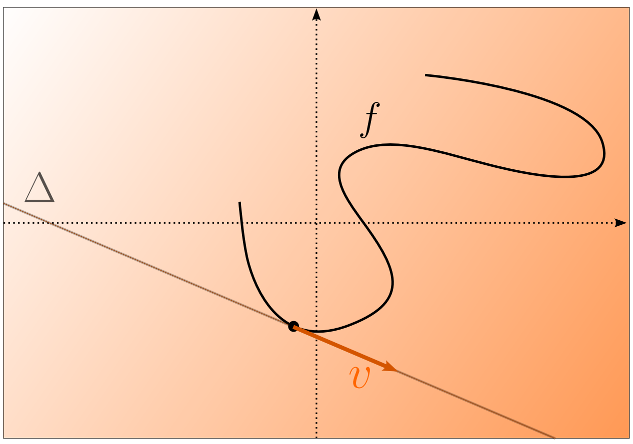

FIGURE 2.4 – La droite affine Δ grise est tangente au chemin f. Le vecteur v est un vecteur tangent à f au point noir ; c'est peut-être le vecteur vitesse, mais on ne peut pas savoir tant qu'on ne connaît pas la paramétrisation précise du tracé de ce chemin.

■ **Exemple 2.40** Soit e un vecteur de E. Alors $f(t) = t \cdot e$ et $g(t) = 2t \cdot e$ ont tous les deux le vecteur e comme vecteur tangent. Pourtant f et g ne sont pas tangents entre eux ! ils n'ont pas le même vecteur vitesse. ■

Il reste la question délicate : *Tous les chemins ont-ils un vecteur vitesse ?* Bien entendu, vous devez savoir que la réponse est négative. C'est pour cela qu'on introduit la notion de *differentiabilité*.

Définition 2.41 Soit $t_0 \in I$, et (X, E) un espace affine normé. Un chemin $f : I \to X$ est dit **différentiable** en t_0 s'il existe un vecteur $v_0 \in E$ tel que f soit tangent à $t \mapsto f(t_0) + (t - t_0) \cdot v_0$ en t_0 ; autrement dit, on demande la formule familière :

$$f(t) = f(t_0) + (t - t_0) \cdot v_0 + o(t - t_0) \quad \text{lorsque } t \to t_0 \tag{2.3}$$

ou encore, en posant $h = t - t_0$:

$$\lim_{\substack{h \to 0 \\ h \neq 0 \\ t_0 + h \in I}} \frac{f(t_0 + h) - f(t_0)}{h} = v_0 \in E\,. \tag{2.4}$$

Le membre de gauche dans la formule (2.4) s'appelle **la limite du taux d'accroissement**.

Ainsi, une fonction $f : \mathbb{R} \to \mathbb{R}$ dérivable est (par définition) un chemin différentiable.

■ **Exemple 2.42** Soit $f :]-1, 1[\to \mathbb{R}^2$ le chemin donné par

$$f(t) = \begin{cases} (t, -\sin t) & \text{pour } t \in]-1, 0[\\ (-t, -\sin t) & \text{pour } t \in [0, 1[\end{cases}$$

C'est bien une fonction continue (à vérifier) mais elle n'est pas différentiable en $t = 0$: on voit facilement sur le dessin (Figure 2.5) que la « tangente à gauche » (la limite en $t \to 0^-$ du taux d'accroissement $v_- = (1, -1)$) est différente de la « tangente à droite » (la limite en 0^+, $v_+ = (-1, -1)$). La limite (2.4) n'existe donc pas. ■

■ **Exemple 2.43** Soit $f :]-1, 1[\to \mathbb{R}^2$ le chemin donné par

$$f(t) = \begin{cases} (t, t^2) & \text{pour } t \in]-1, 0[\\ (-t, -t^2) & \text{pour } t \in [0, 1[\end{cases}$$

C'est bien une fonction continue (à vérifier) mais elle n'est pas différentiable en $t = 0$: les limites à gauche et à droite du taux d'accroissement ne sont pas égales : $v_- = (1, 0)$ et $v^+ = (-1, 0)$. La limite (2.4) n'existe donc pas. Par contre, ces limites étant colinéaires, il existe une unique direction tangente au point $(0, 0)$, et donc en tout point ! Voir Figure 2.6. ■

Remarque 2.44 On dit aussi chemin **dérivable**, et le vecteur vitesse v_0 est la **dérivée** de f en t_0. On le note $v_0 = f'(t_0)$ (notation de Lagrange, d'après Euler), ou $v_0 = \frac{\mathrm{d}f}{\mathrm{d}t}(t_0)$ (notation de Leibniz), ou encore $v_0 = \dot{f}(t_0)$ (notation de Newton). Voir aussi la notation 4.10.

Il semblerait que cette idée de différentiabilité (sans la définition précise de limite) soit due à Fermat, puis reprise par Leibniz et Newton. Lorsque $E = \mathbb{R}$ on retrouve bien sûr la notion habituelle de fonction dérivable.

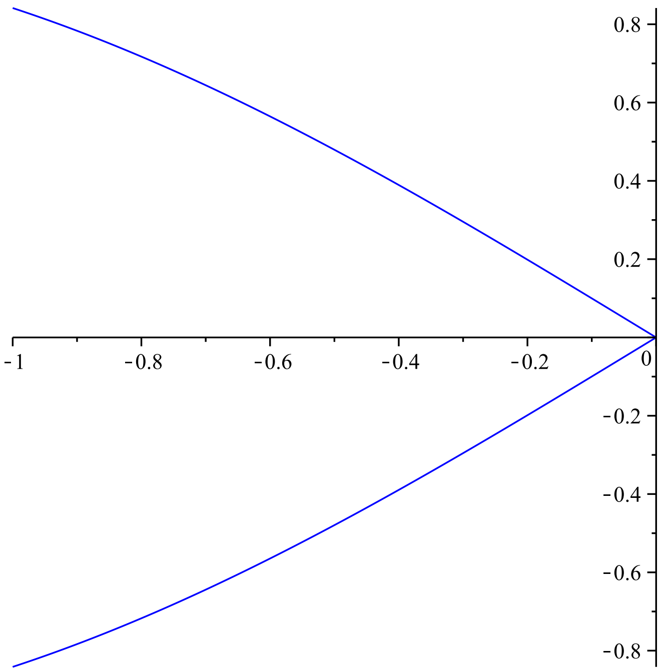

FIGURE 2.5 – Le tracé du chemin f de l'exemple 2.42

Remarque 2.45 Lorsque I est un intervalle fermé à gauche, et t_0 est sa borne de gauche, on obtient la notion de *dérivée à droite* (dans la limite, on doit prendre $h > 0$). De même, on obtient la dérivée à gauche si I est fermé à droite ($h < 0$).

Remarque 2.46 On peut également définir la différentiabilité pour une fonction $f : I \to X$ qui n'est pas nécessairement continue. Cependant, on déduit directement de (2.3) que $f(t)$ tend vers $f(t_0)$ lorsque $t \to t_0$: autrement dit, si f est différentiable en t_0, elle est nécessairement continue en ce point.

Exercice 2.47 (Voir la remarque 2.46 ci-dessus). Montrer qu'une fonction $f : I \to X$ différentiable en $t_0 \in I$ est continue en t_0, en utilisant précisément la définition 2.19. ■

Remarque 2.48 D'un point de vue plus abstrait, le vecteur vitesse répond à la question : quelle est la classe d'équivalence des chemins dérivables qui sont tangents entre eux en t_0 ? (C'est ceux qui ont le même vecteur vitesse.)

Remarque 2.49 Comme petit aperçu de *géométrie différentielle*, définissons :

Définition 2.50 L'**espace tangent** à X en x_0, noté $T_{x_0}X$, est l'ensemble de tous les vecteurs vitesses possibles pour des chemins qui passent par x_0.

Lorsque X est un espace affine, ce n'est pas très palpitant, en effet :

Proposition 2.51 Soit X un espace affine normé, associé à l'espace vectoriel E. Pour tout $x_0 \in X$, $T_{x_0}X = E$.

Démonstration. Prendre les droites ! ■

Cette proposition justifie la terminologie de la définition 2.10.

Mais, examinez un instant la définition 2.50 en supposant que X, au lieu d'être un espace affine, est le *cercle unité* dans $\mathbb{R}^2$... En chaque point du cercle, on obtient que

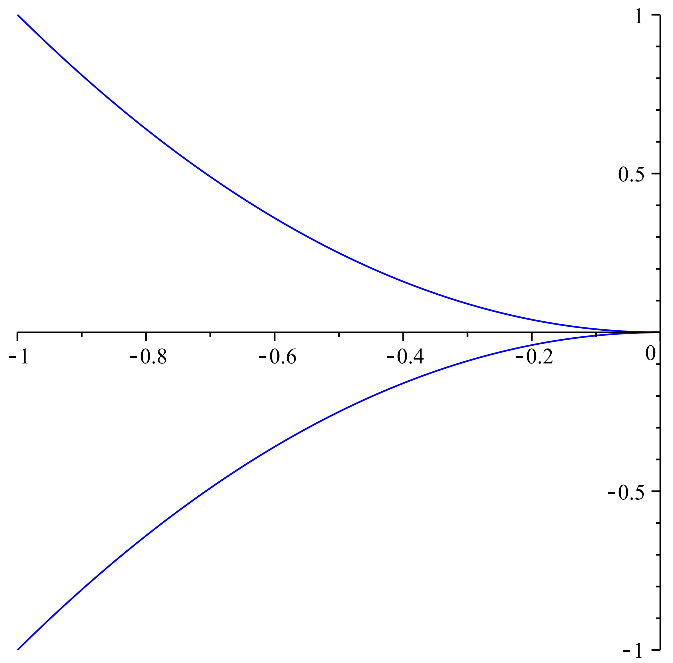

FIGURE 2.6 – Le tracé du chemin f de l'exemple 2.43

l'espace tangent est la droite tangente au cercle en ce point, au sens habituel. Vous pouvez deviner alors qu'on va pouvoir généraliser cette notion à tout *espace courbe* (= variété différentielle) et obtenir en chaque point un espace vectoriel intrinsèque, "tangent" à cet espace courbe.

Proposition 2.52 Considérons l'espace $\mathbb{R}^n$, muni de la norme du sup. Soit $f : I \to \mathbb{R}^n$; notons $f(t) = (f_1(t), \dots, f_n(t))$. Alors f est différentiable si et seulement si ses composantes f_i le sont, et la dérivée est le vecteur des dérivées des composantes.

Démonstration. Voir TD02. ■

Corollaire 2.53 C'est encore vrai pour n'importe quelle norme sur $\mathbb{R}^n$, car elles sont équivalentes, et donc définissent les mêmes limites. (TOPG)

Ainsi, si on revient aux exemples 2.42 et 2.43, on voit en appliquant la proposition 2.52 que dans les deux cas, f est différentiable partout *sauf* en $t = 0$.

Définition 2.54 Si $f : I \to X$ est différentiable en tout t dans I, on dit que f est **différentiable (ou dérivable) sur** I. L'application $t \mapsto v(t)$ est la **dérivée** ou **différentielle** de f, notée f' ou Df ou $\frac{\mathrm{d}f}{\mathrm{d}t}$.

Attention : la multiplicité du vocabulaire (dérivée, différentielle) et surtout des notations (Df, f', etc.) peut être perturbante ! On en verra d'autres... C'est à la fois une reflet de l'histoire du calcul infinitésimal, et un signe que cette notion est utilisée dans de nombreux domaines différents, chaque domaine ayant ses notations préférées.

Dans ce cours on *essaiera* (sans garantie) d'utiliser la notation de Lagrange $f'(t)$ pour les fonctions d'une variable, et la notation Df dans le cas de plusieurs variables (chapitre 4).

Définition 2.55 — classe C^1. On dit qu'une fonction $f : I \to X$ est (de classe) C^1 sur l'intervalle I si f y est dérivable (donc continue) et f' y est continue. On note $f \in C^1(I;X)$.

Après toutes ces définitions, voici les premiers résultats importants de calcul différentiel.

Proposition 2.56 — dérivée d'une somme. Soit X un espace affine normé et E son espace vectoriel associé. Si $f : I \to X$ et $g : I \to E$ on peut définir $f + g : t \to f(t) + g(t) \in X$. Si f est g sont dérivables en un point t_0, alors $f + g$ aussi et on a

$$(f+g)'(t) = f'(t) + g'(t)\,.$$

Exercice 2.57 Démontrer cette proposition. ■

La dérivée d'un « produit bilinéaire » (ici le produit d'un scalaire par un vecteur, voir aussi l'exemple 4.30) est donnée par la fameuse **formule de Leibniz**.

Proposition 2.58 — dérivée d'un produit — formule de Leibniz. Soit E un espace vectoriel normé. Si $f : I \to E$ et $g : I \to \mathbb{R}$ on peut définir $g \cdot f : t \to g(t) \cdot f(t) \in E$. Si f est g sont dérivables en un point t_0, alors $g \cdot f$ aussi et on a

$$(g \cdot f)'(t) = g'(t) \cdot f(t) + g(t) \cdot f'(t)$$

Démonstration. c'est la même preuve que celle que vous connaissez pour les fonctions de $\mathbb{R} \to \mathbb{R}$; il suffit de considérer

$$(g \cdot f)(t_0 + h) - (g \cdot f)(t_0) = (g(t_0 + h) - g(t_0)) \cdot f(t_0 + h) + g(t_0) \cdot (f(t_0 + h) - f(t_0))\,.$$

On divise par h et on prend la limite (les lois d'addition et de multiplication par un scalaire sont continues ! voir le cours de Topologie) ■

Proposition 2.59 — dérivée d'une composition. Soient I, J deux intervalles de $\mathbb{R}$, et $t_0 \in I$. Si $g : I \to J$ et $f : J \to X$ sont dérivables en t_0 et $g(t_0)$ resp., alors $f \circ g$ aussi et

$$(f \circ g)'(t_0) = g'(t_0) \cdot f'(g(t_0))\,.$$

Démonstration. $g(t_0 + h) = g(t_0) + hg'(t_0) + o(h)$ par définition de dérivabilité de g. Donc par définition de la dérivabilité de f,

$$\begin{aligned} f(g(t_0 + h)) &= f\left(g(t_0) + hg'(t_0) + o(h)\right) \\ &= f(g(t_0)) + [hg'(t_0) + o(h)] \cdot f'(g(t_0)) + o(hg'(t_0) + o(h)) \\ &= f(g(t_0)) + hg'(t_0) \cdot f'(g(t_0)) + o(h)\,, \end{aligned}$$

ce qui démontre la proposition. On a utilisé le fait que « $o(\mathcal{O}(h)) \subset o(h)$ », c'est-à-dire : si une fonction g_1 est $\mathcal{O}(h)$, alors toute fonction $g_2 \in o(g_1)$ est aussi $o(h)$. En effet, $\frac{\|g_2\|}{h} = \frac{\|g_2\|}{\|g_1\|}\frac{\|g_1\|}{h}$ tend vers 0 quand $h \to 0$ car c'est un produit d'une fonction qui tend vers 0 par une fonction bornée. ■

2.4 Accroissements finis

Théorème 2.60 — égalité des accroissements finis (TAF) pour les fonctions à valeurs réelles — rappel. Soient des réels $a < b$, soit $f : [a,b] \to \mathbb{R}$ continue, différentiable (=dérivable) sur l'ouvert $]a,b[$. Soit g l'unique fonction affine qui coïncide avec f en a et b, alors il existe c dans $]a,b[$ tel que $f'(c) = g'(c)$.

Exercice 2.61 Quelle est cette fameuse fonction affine g ? Que vaut $g'(c)$? ▪

Démonstration. On considère $h = f - g$ et on utilise le lemme fondamental : (uniquement valable pour les fonctions à valeurs réelles !)

Lemme 2.1 Soit $h :]a,b[\to \mathbb{R}$ une fonction différentiable. Si elle admet un extremum en un point $c \in]a,b[$, alors sa dérivée y est nulle : $f'(c) = 0$.

Preuve du lemme. En considérant les taux d'accroissements à gauche et à droite, on voit que $f'(c)$ est à la fois $\geqslant 0$ et $\leqslant 0$. ■

Fin de la preuve du théorème : h est continue et vaut 0 en a et b. Soit elle est nulle (ok) soit le max est atteint en c dans $]a,b[$ (elle atteint ses bornes par continuité et compacité). Alors $h'(c) = 0$, CQFD. ■

Corollaire 2.62 Si f à valeurs réelles est dérivable, alors f est croissante sur I ssi $f' \geqslant 0$.

Démonstration. Par l'absurde : si f n'est pas croissante, $\exists a < b, f(b) < f(a)$, par le TAF, $\exists c$ tq $f'(c) < 0$. ■

Remarque 2.63 **Le TAF 2.60 est faux pour f vectorielle**, càd à valeurs dans un espace vectoriel de dim $\geqslant 2$. Considérons l'application $f : [0, 2\pi] \to \mathbb{R}^2$ définie par $f(t) = (\cos t, \sin t)$. On a $f(0) = f(2\pi)$ mais pourtant $f'(t)$ ne peut pas s'annuler.

En dimension $\geqslant 2$, le théorème 2.60 ne s'applique pas, mais on a une *inégalité* des accroissements finis. C'est le premier résultat fondamental de ce cours.

Théorème 2.64 — *Inégalité* des AF pour f vectorielle. Soit X un espace affine normé, et I un intervalle ouvert de $\mathbb{R}$. Soit $f : I \to X$ continue, dérivable et telle que

$$\|f'(t)\| \leqslant g'(t)$$

pour une fonction $g : I \to \mathbb{R}$ dérivable. Alors pour tous $a < b$ dans I, on a

$$|f(b) - f(a)| \leqslant g(b) - g(a)$$

Corollaire 2.65 f est constante si et seulement si f est dérivable et f' est identiquement nulle.

Corollaire 2.66 La conclusion du théorème reste vraie si f et g sont continues sur $[a,b]$ et dérivables à l'intérieur. (Par continuité.)

Démonstration. On trouve dans la littérature de nombreuses preuves de ce théorème. Nous allons faire ici une preuve par contradiction et dichotomie, comme Avez [1].

Supposons le résultat faux : il existe $M > 0$ tel que

$$F(a,b) := \|f(b) - f(a)\| - (g(b) - g(a)) = M.$$

On remarque que, pour tous a, b, c,

$$F(a,c) + F(c,b) = \|f(c) - f(a)\| + \|f(b) - f(c)\| - (g(b) - g(a)) \geqslant F(a,b) \quad (2.5)$$

par inégalité triangulaire. Donc l'un des deux $F(a,c)$ ou $F(c,b)$ doit être $\geqslant M/2$. On prend c la moitié $(a+b)/2$ et on recommence du côté où on a l'inégalité, qu'on note $[a_1, b_1]\dots$ On obtient par récurrence :

$$F(a_n, b_n) \geqslant \frac{M}{2^n} \quad \text{et} \quad b_n - a_n = \frac{b-a}{2^n},$$

et bien sûr les a_n et b_n (suites adjacentes) convergent vers le même point t dans $[a,b]$.

Or $F(a_n, b_n) \leqslant F(a_n, t) + F(t, b_n)$ (d'après (2.5)). On rappelle que

$$f(t+h) = f(t) + h \cdot f'(t) + o(h) \qquad \text{lorsque } h \to 0,$$

donc, quand $n \to \infty$,

$$F(a_n, t) = \left\|(t - a_n) \cdot f'(t)\right\| + o(t - a_n) - g'(t)(t - a_n) + o(t - a_n)$$

et

$$F(t, b_n) = \left\|(b_n - t) \cdot f'(t)\right\| + o(b_n - t) - g'(t)(b_n - t) + o(b_n - t).$$

Ainsi :

$$\frac{M}{2^n} \leqslant F(a_n, b_n) \leqslant (|t - a_n| + |b_n - t|) \left\|f'(t)\right\| - g'(t)(b_n - a_n) + o(t - a_n) + o(b_n - t).$$

On divise par $b_n - a_n$ qui vaut $(b-a)/2^n$ (sachant que $|t - a_n| + |b_n - t| = |b_n - a_n|$ car $t \in [a_n, b_n]$) :

$$\frac{M}{b-a} \leqslant \left\|f'(t)\right\| - g'(t) + o(1) \qquad \text{lorsque } n \to \infty$$

et donc $M/(b-a) \leqslant \|f'(t)\| - g'(t)$ ce qui contredit l'hypothèse. ∎

Corollaire 2.67 Si f est dérivable au voisinage de 0, $k \geqslant 0$ et $\|f'(h)\| = o(h^k)$ lorsque $h \to 0$, alors $f(h) - f(0) = o(h^{k+1})$

Démonstration. Attention, une fonction $o(h^k)$ n'est pas forcément dérivable, pensez par exemple à $h \mapsto |h|$ qui n'est pas dérivable en 0.

Par hypothèse, pour tout $\varepsilon > 0$, on a un $\alpha > 0$ tel que pour tous $|h| < \alpha$,

$$\left\|f'(h)\right\| \leqslant \varepsilon\,|h|^k\,.$$

Fixons un tel h. Pour $t \in [0,1]$, on a donc $\|f'(th)\| \leqslant \varepsilon t^k\,|h|^k$. On peut alors poser $g(t) = \varepsilon\,|h|^{k+1}t^{k+1}/(k+1)$, qui est bien dérivable, de sorte que $\|h\cdot f'(ht)\| \leqslant g'(t)$ pour tout $t \in [0,1]$; donc en posant $F(t) = f(ht)$ on a $\|F'(t)\| \leqslant g'(t)$, donc par les AF pour $t \in [0,1]$, $\|F(1)-F(0)\| \leqslant g(1)-g(0) \leqslant \varepsilon\,|h|^{k+1}$.

Donc le résultat. ■

Exercice 2.68 Dans le preuve du corollaire 2.67, on a utilisé le fait que, si $F(t) = f(ht)$, alors $F'(t) = h\cdot f'(ht)$. Montrez-le. ■

Un autre corollaire des accroissements finis est la *Formule de la moyenne*, mais elle nécessite une hypothèse supplémentaire pour E, qui doit être un espace de Banach. On reporte cette discussion au chapitre 5.

2.5 Dérivées d'ordre supérieures dans un EVN/EAN

Si $f : I \to X$ est à valeurs dans un espace affine normé, alors sa dérivée $f' : I \to E$ est à valeurs dans un espace vectoriel normé. On peut donc se demander si f' est différentiable... et ainsi de suite. Pour appliquer directement à f' la discussion des sections précédentes, il sera agréable de supposer que f' est elle-même un chemin (donc continue). Autrement dit, que f est C^1 (définition 2.55).

Définition 2.69 *Définition récursive* : Soit $k \geqslant 1$: un chemin $f : I \to X$ est dit C^k s'il est dérivable et sa dérivée [a] est C^{k-1}.

a. On rappelle qu'on utilise les mots « différentiable » et « dérivables » comme synonymes. De même pour « dérivée » et « différentielle ».

On pourra noter $C^k(I)$ l'ensemble des fonctions qui sont C^k sur I. On définit aussi par récurrence les dérivées successives $f^{(k)} : I \to E$

Définition 2.70 $f^{(0)} = f$ et $\forall k \geqslant 1$, $f^{(k)}$ est la dérivée de $f^{(k-1)}$.

Exercice 2.71 Si f est C^k, alors les $f^{(j)}$, $j = 0,\dots,k$ sont continues. ■

Définition 2.72 On dit que f est C^∞ (ou « lisse ») si elle est C^k pour tous k.

■ **Exemple 2.73** Toute **fonction polynomiale**

$$P(t) = x_0 + t\cdot a_1 + t^2\cdot a_2 + \cdots + t^n\cdot a_n$$

avec $x_0 \in X$ et $a_j \in E$ est C^∞. ■

On a défini les classes C^k ; on peut aussi introduire une notion plus restrictive : on dira que f est k fois **différentiable au point** t_0 si f est C^{k-1} sur un intervalle ouvert contenant t_0, et que la dérivée $f^{(k-1)}$ est différentiable en t_0.

Dans ce cas, on ne sait pas si $f^{(k)}$ existe sur tout I (seulement en t_0). Même si $f^{(k)}$ existe sur tout I, on ne sait pas si elle est continue.

Définition 2.74 Soit $f : I \to E$. On dit que f est **tangente à** 0 **à l'ordre** k **en** t_0 si $f(x) = o((t-t_0)^k)$.

Ainsi : tangent à 0 (au sens de la définition 2.34) est la même chose que tangent à 0 à l'ordre 1.

Intuitivement, plus k est grand, plus le graphe de f est « plat » en l'origine.

Définition 2.75 On dit que f est **plate** en 0 si elle est C^∞ et tangente à 0 à tout ordre.

Définition 2.76 Si $f : I \to X$ est k fois dérivable en $t_0 \in \mathring{I}$, la fonction polynomiale

$$P(t) = \sum_{j=0}^{k} \frac{(t-t_0)^j}{j!} f^{(j)}(t_0) = f(t_0) + \cdots \in X$$

s'appelle le **polynôme de Taylor** de f en t_0 à l'ordre k.

Le polynôme de Taylor est une *bonne approximation de f au voisinage de* t_0.

Théorème 2.77 — développement de Taylor. Si $f : I \to X$ est k fois dérivable en $t_0 \in \mathring{I}$ ($k \geqslant 1$) alors f est tangente à l'ordre k en t_0 à son polynôme de Taylor de degré k. En formule :

$$f(t_0+h) = f(t_0) + \sum_{j=1}^{k} \frac{h^j}{j!} f^{(j)}(t_0) + o(h^k)\,. \tag{2.6}$$

Démonstration. Récurrence sur k. Pour $k=1$ c'est la définition de différentiable. Supposons $k \geqslant 2$ et la formule vraie pour $k-1$. Posons

$$F(h) := f(t_0+h) - f(t_0) - \sum_{j=1}^{k} \frac{h^j}{j!} f^{(j)}(t_0).$$

On remarque que $F(0) = 0$. F est continue, dérivable au voisinage de 0 et on a

$$F'(h) = f'(t_0+h) - \sum_{j=1}^{k} \frac{h^{j-1}}{(j-1)!} f^{(j)}(t_0) \tag{2.7}$$

$$= f'(t_0+h) - f'(t_0) - \sum_{j=1}^{k-1} \frac{h^j}{j!} f^{(j+1)}(t_0)\,. \tag{2.8}$$

Par hypothèse de récurrence, puisque f' est $(k-1)$ fois dérivable, on sait que

$$f'(t_0+h) - f'(t_0) - \sum_{j=1}^{k-1} \frac{h^j}{j!} (f')^{(j)}(t_0) = o(h^{k-1})\,.$$

Autrement dit, $F'(h) = o(h^{k-1})$. Par le corollaire 2.67 de l'inégalité des AF :

$$\|F(h)\| = \|F(h) - F(0)\| = o(h^k)\,,$$

ce qui est le résultat voulu. ■

R

Remarque 2.78 Dans l'équation (2.6), le premier terme $f(t_0)$ est un *point* et le reste $\sum_{j=1}^{k} \frac{h^j}{j!} f^{(j)}(t_0) + o(h^k)$ est un *vecteur*.

■ **Exemple 2.79** Le polynôme de Taylor d'une fonction plate est toujours le polynôme nul, quelque soit l'ordre k. Ça n'implique pas qu'une fonction plate soit forcément nulle ! Par exemple $f(t) = \exp(-1/t^2)$ (prolongée par 0 en $t = 0$) est plate en 0. (Exercice) ■

Attention, on a obtenu les implications suivantes :
— $f \in C^k$ implique f différentiable en tout point de I ;
— f différentiable en t_0 implique qu'elle est tangente à l'ordre k en t_0 à un polynôme (son polynôme de Taylor).

Mais aucune des deux réciproques n'est valable en général.

■ **Exemple 2.80** La fonction $f(t) = t^2 \sin(\frac{1}{t})$ est différentiable sur $\mathbb{R}$ mais n'est pas C^1. En effet elle est dérivable en 0 (écrire le taux d'accroissement, ou simplement $f(t) = o(t)$), et elle est dérivable sur $\mathbb{R} \setminus \{0\}$ par les théorèmes d'opérations habituels. Cependant la dérivée f' n'est pas continue car pour $t \neq 0$, $f'(t) = t \sin \frac{1}{t} - \cos(\frac{1}{t})$, qui n'admet pas de limite quand $t \to 0$. ■

■ **Exemple 2.81** Lorsque $k = 1$, le fait que f soit tangent à l'ordre 1 à une fonction affine (polynôme de degré 1) est équivalent à la différentiabilité de f (c'est la définition !). Mais ce n'est plus le cas pour $k \geqslant 2$. Par exemple soit $f(t) = t^3 \sin(\frac{1}{t})$, et $f(0) = 0$. Alors f est continue et on a clairement $f(t) = o(t^2)$ en $t = 0$. En particulier $f(t) = o(t)$, ce qui dit que f est différentiable en 0 avec $f'(0) = 0$. Cependant le taux d'accroissement

$$\frac{f'(t) - f'(0)}{t} = 3t \sin(\frac{1}{t}) - \cos(\frac{1}{t})$$

n'a pas de limite quand $t \to 0$. Donc f n'est pas deux fois différentiable en 0. ■

En renforçant l'hypothèse sur f on peut préciser le $o(h^k)$ dans le développement de Taylor.

Théorème 2.82 — reste de Lagrange. Si $f : I \to X$ est $\underline{k+1}$ fois différentiable sur I ($k \geqslant 1$) alors pour a, b dans I :

$$\left\| f(b) - f(a) - \sum_{j=1}^{k} \frac{(b-a)^j}{j!} f^{(j)}(a) \right\| \leqslant \frac{(b-a)^{k+1}}{(k+1)!} \sup_{[a,b]} \left\| f^{(k+1)} \right\|$$

Démonstration. Voir TD. ■

Ainsi, le $o(h^k)$ dans (2.6) est devenu un $\mathcal{O}(h^{k+1})$. On verra qu'on pourra donner une formule exacte pour ce reste (le *reste intégral*) lorsque E sera un Banach, au chapitre 5, qui impliquera facilement le reste de Lagrange.

3. Morphismes d’EVN : applications linéaires continues

Le contenu de ce chapitre sera repris, avec davantage de détails, dans le cours de Topologie. Je me permettrai donc d’aller assez vite, sans faire les preuves.

On s’intéresse ici à changer d’EVN en « préservant les structures » : En général, dans toutes les branches des mathématiques, une application qui préserve une structure (une « forme ») s’appelle un *morphisme* (morphè (en grec) = forme). Quels sont donc les **morphismes d’EVN** ? Que doit satisfaire une application $A : E \to F$ où E et F sont des EVN (sur $\mathbb{R}$ ou $\mathbb{C}$) pour recevoir ce titre prestigieux de « morphisme » ?

Plus précisément, en ce qui nous concerne, si $f : I \to E$ est un chemin différentiable, on voudrait que $A \circ f : I \to F$ soit encore un chemin différentiable.

Un evn est donné par un espace vectoriel et une norme. On va donc déjà demander que A préserve la structure d’ev : on dit que A est **linéaire** :

$$A(u + \lambda \cdot v) = Au + \lambda \cdot Av$$

(On notera indifféremment Au ou $A(u)$ la valeur de A sur le vecteur u — on dit aussi l’action de A sur u, et A est aussi appelé un *opérateur*.)

Quid de la norme ? Que demander de $\|Au\|$? On ne cherche pas des automorphismes (= bijectifs), seulement des morphismes : par exemple 0 doit être un bon morphisme. Donc on ne peut pas demander que la norme soit préservée (le fait d’être une *isométrie* : $\|Au\| = \|u\|$, impliquerait A injectif). Par contre on voudrait préserver la notion de chemin ainsi que la différentiabilité des chemins : autrement dit, préserver les limites. On va donc demander

$$\|Au\| \leqslant C\,\|u\|$$

pour une certaine constante C (dépendant de A).

Cette inégalité permet d’assurer que $\|u(t)\| \to 0 \Rightarrow \|Au(t)\| \to 0$.

Définition 3.1 Un **morphisme** entre deux espaces vectoriels normés E et F est une application linéaire $A : E \to F$ telle qu’il existe $C > 0$ tel que, pour tout u dans E, $\|A(u)\| \leqslant C\,\|u\|$. On note $\mathscr{L}(E,F)$ ou $L_c(E,F)$ ou $B(E,F)$ l’ensemble des morphismes de E dans F.

FIGURE 3.1 – *Morphée*, sculpture de Jean Antoine Houdon, 1777. Musée du Louvre, Paris. Du grec « morphè », la forme, le dieu de la mythologie grecque Morphée, représenté comme une créature (homme ou femme) ailée, est connu comme le dieu des rêves, dans lesquels il apparaît sous de multiples formes. Ces « formes » peuvent représenter des personnes, mais aussi des concepts abstraits.

Exercice 3.2 $\mathscr{L}(E,F)$ est aussi un espace vectoriel. ■

Remarque 3.3 Une application linéaire $A : E \to F$ est un morphisme ssi A est un « opérateur borné » : l'image de la boule unité est bornée. D'où la notation $B(E,F)$. La constante $C = C(A)$ *optimale* dans la définition 3.1 est appelée la *borne* de A, ou la *norme* de A. En fait $A \mapsto C(A)$ est bien une norme sur l'EV $\mathscr{L}(E,F)$. On l'appelle parfois la **norme subordonnée** (car elle dépend de la norme sur E). On la note $\|A\|$ si c'est clair, $|||A|||$ (norme triple !) ou $\|A\|_{\mathscr{L}(E,F)}$ s'il faut être précis.

Question : Que veut dire : « la constante C optimale » ? ?

Par définition, la constante C est un majorant de l'ensemble des $\frac{\|A(u)\|}{\|u\|}$, $u \in E \setminus \{0\}$. La constante optimale est donc le plus petit majorant possible, c'est ce qu'on appelle la *borne supérieure* de l'ensemble des $\frac{\|A(u)\|}{\|u\|}$.

Définition 3.4 La borne de A (ou **norme subordonnée**) est

$$\|A\|_{\mathscr{L}(E,F)} := \sup\left\{\frac{\|A(u)\|}{\|u\|}, \quad u \neq 0\right\}.$$

c'est aussi

$$\sup\{\|Au\|,\quad \|u\|=1\}$$

ou encore

$$\sup\{\|Au\|,\quad \|u\|\leqslant 1\}.$$

Proposition 3.5 la composée de morphismes est un morphisme, càd : si $A\in\mathscr{L}(E,F)$ et $B\in\mathscr{L}(G,E)$, alors $A\circ B\in\mathscr{L}(G,F)$; et la norme subordonnée vérifie :

$$\|AB\|\leqslant\|A\|\,\|B\|\,.$$

Ainsi, si $f:I\to E$ est un chemin continu, alors $A\circ f:I\to F$ est aussi un chemin continu [1].

Démonstration. Soit $\varepsilon>0$ donné, si α est tel que $|t-t_0|<\alpha\Rightarrow\|f(t)-f(t_0)\|<\varepsilon/C$ alors on voit que pour les mêmes t, on a $\|Af(t)-Af(t_0)\|<\varepsilon$. ■

Et même mieux :

Théorème 3.6 Si f est différentiable $I\to E$, et $A\in\mathscr{L}(E,F)$, alors $A\circ f$ est différentiable de $I\to F$, de dérivée $(A\circ f)'=A\circ f'$.

Démonstration. Si f est tangente à 0 alors Af aussi, car $A(o(h))\in o(h)$ (Exercice !). Donc si f est tangente à une fonction affine g, alors Af est tangente à Ag qui est aussi affine. ■

Ne poussons pas le suspense trop loin : on peut en fait montrer :

Théorème 3.7 (TOPG) : $\mathscr{L}(E,F)$ est l'ensemble des **applications linéaires continues** de E dans F.

(D'où la notation $L_c(E,F)$, l'indice c étant pour « continu ».)

On en déduit que si le morphisme A est inversible et A^{-1} est aussi un morphisme, alors A est un *homéomorphisme*, et l'application $x\to\|A(u)\|$ est une norme équivalente à $\|u\|$.

Exercice 3.8 (TOPG) une application linéaire est continue ssi elle est continue en 0. ■

Exercice 3.9 Une application linéaire n'est pas toujours continue. ■

Théorème 3.10 Si E est de dimension finie, toute application linéaire de $E\to F$ est continue.

(TOPG). Prendre une base de E : $(e_1,dots,e_n)$; alors on peut prendre $C=\|Ae_1\|+\cdots+\|Ae_n\|$, si on met la norme sup sur E. Or les normes sur E sont équivalentes. ■

Théorème 3.11 — Theorème d'Hahn-Banach (admis). Soit E un evn et $u_0\in E$, non nul. Alors il existe une *forme linéaire continue* $L:E\to\mathbb{R}$ (donc $L\in L_c(E,R)$) telle que

1. $\|L\|=1$

1. On note indifféremment $A\circ B$ ou AB, $A\circ f$ ou Af, etc.

2. $L(u_0) = \|u_0\|$.

Application : preuve des AF avec Hahn-Banach. Si on utilise un théorème fort d'analyse fonctionnelle (Hahn-Banach, sur tout EVN, nécessite l'axiome du choix), on peut se ramener aux fonctions de $\mathbb{R}$ dans $\mathbb{R}$ et appliquer le résultat connu dans ce cas. Voici quelques détails.

D'après le corollaire du TAF scalaire, g est croissante. Si $f(b) = f(a)$, il n'y a rien à prouver ; sinon, le théorème de Hahn-Banach affirme l'existence d'une forme linéaire continue L, telle que

1. $|L(u)| \leqslant \|u\|$ pour tous u, et
2. $L(f(b)-f(a)) = \|f(b)-f(a)\|$. (Donc $\|L\| = 1$.)

Puisque L est continue, le théorème 3.6 nous assure que $L \circ f$ est dérivable $I \to \mathbb{R}$, de dérivée $L(f'(t))$, donc majorée en norme par $g'(t)$. Donc $g - L\circ f$ est croissante, et ça donne le résultat en testant sur a et b.

Exercice 3.12 Définir la bonne notion de morphisme d'espaces *affines* normés ■

Applications *multilinéaires* continues. Au fait, qu'est-ce qu'une application *multilinéaire* ? Vous connaissez certainement le cas *bilinéaire*.

Proposition 3.13 Soient $(E_1, \|\cdot\|_1), \dots, (E_n, \|\cdot\|_n)$, $(F, \|\cdot\|_F)$ des evn et $f : E_1 \times \cdots \times E_n \to F$ une application n-linéaire. On munit $E_1 \times \cdots \times E_n$ de la norme $\|(u_1, \dots, u_n)\| = \sup\{\|u_1\|_1, \dots, \|u_n\|_n\}$ (ou de toute norme équivalente). On a les équivalences suivantes :

1. f est continue
2. f est continue en zéro
3. il existe $C > 0$ tel que $\|f(u_1, \dots, u_n)\|_F \leqslant C \|u_1\|_1 \cdots \|u_n\|_n$, $\forall (u_1, \dots, u_n) \in E_1 \times \cdots \times E_n$.

cas $n = 2$. TODO ? ou TOPG ■

Application :

Proposition 3.14 — Composition d'applications linéaires continues.
Soit $\mathscr{C} : \mathscr{L}(F,G) \times \mathscr{L}(E,F) \to \mathscr{L}(E,G)$ définie par

$$\mathscr{C} : (B,A) \mapsto B \circ A\,.$$

Alors $\mathscr{C}$ est linéaire continue.

Démonstration. On a vu en proposition 3.5 que $\|B\circ A\| \leqslant \|B\|\,\|A\|$. La continuité de $\mathscr{C}$ découle donc de la proposition 3.13. ■

Définition 3.15 Pour f multilinéaire continue, la **norme associée** $\|f\| = \dots$ exercice ... comme plus haut (on prend la meilleure constante C) !

On en déduit que la norme de l'application de composition est $\|\mathscr{C}\| = 1$.

4. Différentielle d'une application $f : E \to F$

On arrive au cœur de ce cours : les fonctions de plusieurs variables. Ces fonctions sont utilisées dans toutes les applications des mathématiques. Par exemple, une application $F : \mathbb{R}^2 \to \mathbb{R}^2$ peut servir à décrire l'écoulement d'un fluide ; dans ce cas on la voit comme un **champ de vecteurs** : en chaque *point* $p = (x,y)$, F associe un *vecteur* $v = F(x,y)$, qui représente la vitesse de la particule de fluide au point p. On peut facilement représenter une telle fonction en traçant le vecteur v comme une petite flèche partant du point p (Figure 4.1). (De façon plus générale, un champ de vecteur sera une application d'un espace affine — ou d'une variété différentiable — à valeurs dans son espace tangent.)

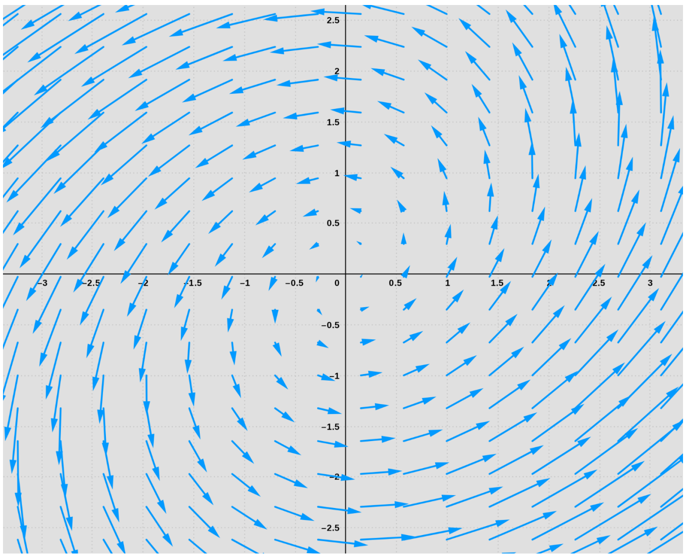

FIGURE 4.1 – Représentation du champ de vecteurs $F : \mathbb{R}^2 \to \mathbb{R}^2$ défini par $F(x,y) = (-y + x/2, y)$.

Dans ce cours, plutôt que de traiter le cas de deux variables, puis trois variables, etc., on va prendre directement la variable x dans un espace de dimension arbitraire (éventuellement infinie !). Traiter la dimension infinie n'est pas qu'un exercice académique : vous connaissez déjà de telles applications : par exemple l'application qui à une fonction associe sa dérivée est une « fonction à nombre infini de variables ». (En général, les espaces de fonctions sont de dimension infinie.) De façon plus générale, les opérateurs différentiels, qui sont

FIGURE 4.2 – *Nuit étoilée (cyprès et village)*, Vincent van Gogh, 1889. Remarquez la similitude avec la représentation d'un *champ de vecteur*, Figure 4.1.

présents dans presque toutes les équations de la physique, peuvent être vus comme des fonctions agissant sur des espaces de dimension infinie.

On se donne deux espaces affines normés X et Y (donc $X = x_0 + E$ et $Y = y_0 + F$, où E et F sont les evn associés). On rappelle (TOPG) que E et F sont munis d'une topologie (engendrée par les boules ouvertes.) Comme E est homéomorphe à X par l'application $u \mapsto x_0 + u$, la topologie de E se transporte sur X (également engendrée par les boules ouvertes).

Exercice 4.1 Définir une boule ouverte dans un espace affine normé. ■

On se donne un ouvert U de X. (Par exemple U = une boule ouverte.) On se donne $f : U \to Y$.

Comment définir la « dérivée » de f ?

On dispose de la notion d'accroissement de $f(x)$: mais dans quelle direction ? (en effet, x bouge dans U qui est multidimensionnel : pensez à un disque dans le plan $\mathbb{R}^2$.)

4.1 Applications tangentes à 0

Les notations de Landau s'étendent sans difficulté.

Définition 4.2 Soit X un espace affine normé, F, G des evn, U un ouvert de X. Soit $x_0 \in U$; on note $\dot{U} := U \setminus \{x_0\}$. Soient $f : \dot{U} \to F$, $g : \dot{U} \to G$. On dit que $f = o(g)$ en x_0 s'il existe une fonction $\varepsilon(x) : \dot{U} \to \mathbb{R}$ *qui tend vers* 0 *en* x_0 telle que

$$\|f(x)\|_F = \varepsilon(t)\,\|g(x)\|_G \text{ près de } x_0 .$$

On a alors :

Définition 4.3 Soit $g : U \to F$, $x_0 \in U$. On dit que g est tangente à 0 en x_0 (à l'ordre 1) si $\|g(x)\|_F = o(\|x - x_0\|_E)$.

Rappel : on peut aussi écrire $g(x) = o(x - x_0)$, c'est la même chose (par définition). Mais on risque moins de se tromper en indiquant directement les normes.

En particulier, g est continue en x_0 ssi $g(x_0) = 0$.

Définition 4.4 f_1 est tangente à f_2 si $(f_2 - f_1)$ est tangente à 0.

Proposition 4.5 Une application linéaire $A : E \to F$ est tangente à 0 en 0 si et seulement si elle est nulle.

Démonstration. Supposons A tangente à 0 en 0. Fixons $y \neq 0$. Soit $\varepsilon > 0$. Par définition, il existe $\alpha > 0$ tel que $\|x - 0\| < \alpha \Rightarrow \|A(x)\| < \varepsilon \|x\|$. Si on pose $x = (\alpha/2)y/\|y\|$, on a bien $\|x\| < \alpha$, donc par homogénéité,

$$\frac{\alpha}{2\|y\|}\|A(y)\| < \varepsilon \frac{\alpha}{2\|y\|}\|y\|$$

donc $\|A(y)\| \leqslant \varepsilon \|y\|$

Comme c'est vrai pour tout ε, pour ce y fixé, on en déduit que $A(y) = 0$. CQFD. ■

R **Remarque 4.6** Autre preuve : à partir de $\|A(y)\| \leqslant \varepsilon \|y\|$, on pourrait aussi en déduire que A est continue (càd dans $\mathscr{L}(E,F)$) de norme $< \varepsilon$, donc au final de norme 0…

4.2 Applications différentiables $U \to Y$, différentielle $E \to F$ (Fréchet)

Soient X, Y des EAN, U un ouvert de X.

Définition 4.7 Soit $f : U \to Y$, $a \in U$. On dit que f est **différentiable en a** s'il existe une *application linéaire continue* $A : E \to F$ telle que f soit tangente à l'application affine $x \to f(a) + A(x - a)$. En formule :

$$f(a+h) - f(a) = A(h) + o(\|h\|), \qquad h \to 0 \tag{4.1}$$

On appelle A **différentielle** de f en a. On dit aussi différentielle de Fréchet.

FIGURE 4.3 – Maurice FRÉCHET (1878 – 1973)

Proposition 4.8 Différentiable $\Rightarrow$ continue.

Démonstration. Puisque $A \in \mathscr{L}(E,F)$, on déduit de (4.1) que $f(a+h) \to f(a)$ quand $h \to 0$. ∎

Proposition 4.9 — (et définition). Si f est différentiable en a, la différentielle A est **unique**. On la note $Df(a)$ ou $f'(a)$ ou $T_a f$ (le « T » est là pour « application linéaire Tangente »).

Démonstration. Si B est une autre différentielle, alors $h \to Ah$ et $h \to Bh$ sont tangentes en $h = 0$, donc $(A - B)$ est tangente à 0 en 0, donc $A - B = 0$ par la proposition 4.5. ∎

▪ **Notation 4.10** Nous voici avec un problème de notation un peu embêtant : puisque $Df(a)$ est une application linéaire de $E \to F$. Comment noter l'application de $Df(a)$ au vecteur h ?

Dans la littérature, on peut trouver : $Df(a)(h)$, $Df(a) \cdot h$, $Df(a)h$, $T_a f(h)$, $d_a f(h)$, $f'(a)(h)$, $L_h f(a)$, etc.

Tout est admissible... ! il faudra bien faire attention de distinguer le *point a* du *vecteur h*. Dans ce cours, on essaiera de privilégier $Df(a) \cdot h$ ou $Df(a)h$, sauf dans le cas d'une variable. L'usage du point $(\cdot)$ pourrait prêter à confusion avec la multiplication par un scalaire dans un espace vectoriel : $\lambda \cdot u$; mais les risques de confusion sont très faibles, et l'avantage de cette notation par rapport à la parenthèse simple $Df(a)(h)$ est de mettre en avant le fait que l'action de $Df(a)$ sur h soit *linéaire*, tout comme l'action de λ sur u dans $\lambda \cdot u$!

Le cas d'une seule variable scalaire. Retour au cas $E = \mathbb{R}$. On avait défini $f'(a) \in F$. Alors que la définition 4.7 dit que $Df(a) \in \mathscr{L}(E,F)$. Comment réconcilier les définitions ?

En comparant $f'(a)$ défini à la définition 2.54 et $Df(a)$ défini à la définition 4.7, on voit que $A(h) = h \cdot f'(a)$. Autrement dit, si on s'autorise à écrire $f'(a)$ à la place de $Df(a)$, c'est qu'on *identifie* $f'(a)$ à l'application linéaire de $\mathbb{R}$ dans F donnée par : $h \mapsto h \cdot f'(a)$. En fait, l'identification de F et $\mathscr{L}(\mathbb{R},F)$ est naturelle. Tout élément $A \in \mathscr{L}(\mathbb{R},F)$ définit un vecteur $f'(a) = A(1)$, et réciproquement tout vecteur $v \in F$ définit l'application linéaire $h \mapsto h \cdot v \in \mathscr{L}(\mathbb{R},F)$. (En d'autres termes, « 1 » est la *base canonique* de $\mathbb{R}$.)

Revenons au cas général.

Définition 4.11 — Différentiabilité sur un ouvert. On dit que f est différentiable sur U si elle est différentiable en tout point de U. L'application $a \mapsto Df(a)$, $U \to \mathscr{L}(E,F)$, est **la différentielle de f**.

R **Remarque 4.12** Contrairement au cas d'une variable, il est délicat de définir la differentiabilité sur un ensemble non ouvert ; *dans ce cours, nous n'utiliserons que des ouverts*. En fait, pour les curieux, la façon la plus simple de définir la différentiabilité d'une fonction f sur un ensemble K quelconque est de demander que f soit égale à la restriction à K d'une fonction différentiable définie sur un ouvert contenant K.

Définition 4.13 — classe C^1. On dit que f est C^1 si sa différentielle est continue (donc de U dans $\mathscr{L}(E,F)$).

Attention à la confusion possible. L'application linéaire $Df(a)$ est *toujours continue* $E \to F$ par définition. Pour être C^1 c'est très différent : on demande que $a \mapsto Df(a)$ soit continue de $U \to \mathscr{L}(E,F)$ avec la norme subordonnée définie au chapitre précédent.
Rappel (exercice 3.9) une application linéaire n'est pas toujours continue. Ceci dit c'est automatique si l'espace de départ E est de dimension finie (théorème 3.10).

Exemples de différentielles :

■ **Exemple 4.14 — Différentielle d'une application constante.** La différentielle d'une application constante, en tout point, c'est $0 \in \mathscr{L}(E,F)$, évidemment. ■

■ **Exemple 4.15 — Différentielle d'une application linéaire continue.** Soit $A \in \mathscr{L}(E,F)$. Soit $f : E \to F$ donnée par $f(x) = Ax$. Quelle est sa différentielle ?

On calcule $f(a+h) - f(a) = Ah$, donc, par unicité (proposition 4.9), $Df(a) = A$ (indépendant de a !). Df est donc constant, donc en particulier C^1. ■

Exercice 4.16 — Différentielle d'une application affine continue. . Soit $f : X \to Y$ une application affine : définie par $f(x) = f(x_0) + A(x - x_0)$ sur un espace affine normé X d'espace vectoriel associé E, avec $x_0 \in X$ fixé, et $A \in \mathscr{L}(E,F)$, où F est l'espace vectoriel associé à Y. Montrer que f est différentiable et que sa différentielle est constante sur X. ■

4.3 Dérivées directionnelles (Gateaux)

On peut définir une notion plus faible que la différentielle habituelle (de Fréchet). C'est la différentielle de Gateaux (René GATEAUX, mathématicien français mort très jeune au début de la guerre de 14. Le 3/10/1914). C'est une dérivée directionnelle :

Pour un vecteur $h \in E$ fixé, on regarde la limite

$$\frac{f(a+t\cdot h) - f(a)}{t}, \quad t \to 0,$$

si elle existe, on la note $T_a f(h)$. On dit que f est Gateaux-différentiable en a si $h \to T_a f(h)$ est $\mathscr{L}(E,F)$. Bien sûr Fréchet $\Rightarrow$ Gateaux (et dans ce cas la différentielle est la même), mais pas l'inverse : pour Gateaux, la façon dont la limite converge peut dépendre de h, pour Fréchet c'est uniforme.

4.4 Différentielle d'une application composée

Proposition 4.17 Soient X,Y,Z trois espace affines normés. U ouvert de X, V ouvert de Y. On se donne $f : U \to Y$ et $g : V \to Z$. Soit $a \in U$ tel que $f(a) \in V$.

1. Si f est différentiable en a et g est différentiable en $f(a)$, alors $g \circ f$ est différentiable en a, et sa différentielle en ce point est
$$D(g\circ f)(a) = Dg(f(a)) \circ Df(a).$$
2. Si f est C^1 dans U et g est C^1 dans V, alors $g \circ f$ est C^1 dans son domaine de définition $U \cap f^{-1}(V)$

Démonstration. C'est fondamentalement la même idée que la preuve de la proposition 2.59. Cependant on voit ici que la continuité linéaire de la différentielle est essentielle.

1. On a, pour h dans E :

$$f(a+h) = f(a) + Df(a) \cdot h + o(h)$$

et pour u dans F :

$$g(f(a)+u) = g(f(a)) + Dg(f(a)) \cdot u + o(u)\,.$$

Donc :

$$g \circ f(a+h) = g(f(a+h)) = g(f(a) + Df(a) \cdot h + o(h))$$

Posons $u = Df(a) \cdot h + o(h)$ qui apparaît dans la formule ci-dessus de sorte que

$$g \circ f(a+h) = g(f(a)+u)$$

Donc $= g(f(a)) + Dg(f(a)) \cdot u + o(u)$.
Puisque $\|u\| \leqslant \|Df(a) \cdot h\| + o(h) \leqslant \|Df(a)\| \, \|h\| + o(h) = O(h)$ (on utilise ici la continuité de $Df(a)$), on voit que toute fonction $o(u)$ est aussi $o(h)$ (Exercice).
D'autre part

$$Dg(f(a)) \cdot (o(h)) \leqslant \|Dg(f(a))\| \, o(h) = o(h)\,.$$

(Là on a utilisé la continuité de $Dg(f(a))$.) Donc au final

$$g \circ f(a+h) = g \circ f(a) + Dg(f(a)) \cdot (Df(a)h) + o(h)\,.$$

Donc $g \circ f$ est différentiable et, par unicité (proposition 4.9), sa différentielle est l'application linéaire

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

(qui est bien continue comme composée de 2 applications (linéaires) continues).
2. La composée d'applications continues est continue : utiliser $(B,A) \to B \circ A$ qui est bilinéaire continue pour $A \in \mathscr{L}(E,F)$ et $B \in \mathscr{L}(F,G)$.

■

4.5 Linéarité de D, la « différentiation »

On peut appeler D l'opérateur de différentiation, qui à f associe Df.

Proposition 4.18 Soit X affine et F vectoriel (normés, évidemment).
1. Les fonctions $f : U \to F$ différentiables (en un point, ou sur un ouvert) forment un EV.
2. L'application $f \to Df$ est linéaire.

Précisons un peu les espaces vectoriels.
1. $D : \mathrm{Diff}(x_0, F) \to \mathscr{L}(E,F)$
2. $D : \mathrm{Diff}(U,F) \to \mathrm{Applications}(U, \mathscr{L}(E,F))$

Exercice 4.19 Les fonctions différentiables de U dans un espace affine Y forment un espace affine ! ■

4.6 Applications à valeurs dans un produit

Soient E,F_1,F_2 les EV associés aux espaces affines X, Y_1, Y_2 respectivement. Le produit $Y_1 \times Y_2$ est un espace affine d'espace vectoriel $F_1 \times F_2$. Ce dernier est « naturellement » normé par $\|(v_1, v_2)\| := \max\{\|v_1\|_{F_1}, \|v_2\|_{F_2}\}$. La topologie associée est la topologie produit (voir cours de Topologie).

Proposition 4.20 Soit $f : U \to Y_1 \times Y_2$. On note ses composantes $f = (f_1, f_2)$. Alors f est différentiable si et seulement si ses composantes f_1 et f_2 le sont, et alors

$$Df(x) = (Df_1(x), Df_2(x)) \in \mathscr{L}(E, F_1 \times F_2)$$

Démonstration. Notons π_1 et π_2 les projections de $X_1 \times X_2$ sur les facteurs Y_1 et Y_2, respectivement. Ce sont des applications affines, donc différentiables. Donc, par composition (proposition 4.17), $f_1 (= \pi_1 \circ f)$ et $f_2 (= \pi_2 \circ f)$ sont différentiables dès que f l'est.

Réciproquement, pour $h \in E$, on a

$$\begin{aligned} f(a+h) &= (f_1(a+h), f_2(a+h)) \\ &= (f_1(a), f_2(a)) + (Df_1(a)h + o(h), Df_2(a)h + o(h)) \\ &= f(a) + Ah + o(h) \end{aligned}$$

avec A l'application linéaire $h \mapsto (Df_1(a)h, Df_2(a)h)$. Cette application est continue de norme $\leqslant \max(\|Df_1(a)\|, \|Df_2(a)\|)$. Donc f est différentiable en a et (par unicité de la différentielle), $Df(a) = A$. ■

4.7 Théorème des accroissements finis

Définition 4.21 Soient a et b deux points d'un espace affine normé X. On appelle **segment** d'extrémités a et b, et on note $[a,b]$, l'ensemble

$$[a,b] = \{a + t(b-a);\ 0 \leqslant t \leqslant 1\} \subset X$$

Théorème 4.22 Soient (X, E) et (Y, F) deux espaces affines normés, et $f : U \to Y$ une application différentiable d'un ouvert U de X. Soient a et b deux points de U tels que le segment $[a,b]$ soit contenu dans U. Alors l'application f vérifie

$$\|f(b) - f(a)\|_F \leqslant \sup_{z \in [a,b]} \|Df(z)\|_{\mathscr{L}(E,F)} \|b - a\|_E$$

Définition 4.23 On dit qu'un ensemble $U \subset X$ est **convexe** si $\forall x_0, x_1 \in U$, le segment $[x_0, x_1]$ est contenu dans U.

Corollaire 4.24 Si U est convexe et s'il existe un réel $M > 0$ tel que, pour tout $x \in U$, $\|Df(x)\| \leqslant M$, alors pour tout couple (a,b) de points de U, on a $\|f(b) - f(a)\| \leqslant$

$M\|b-a\|$.

Preuve du théorème. Considérons le chemin tracé dans Y :

$$[0,1] \ni t \to \phi(t) = f(a+t(b-a)).$$

C'est un chemin différentiable en tout point $t \in]0,1[$ et de différentielle

$$\phi'(t) = Df(a+t(b-a)) \cdot (b-a).$$

Nous avons donc, pour tout $t \in]0,1[$,

$$\|\phi'(t)\| \leqslant \|Df(a+t(b-a))\| \ \|b-a\| \leqslant C\|b-a\|$$

avec $C = \sup \|Df(z)\|_{z\in[a,b]}$. En appliquant l'inégalité des accroissements finis pour les chemins (théorème 2.64), avec $g(t) = tC\|b-a\|$, on obtient $\|\phi(1)-\phi(0)\| \leqslant g(1)-g(0)$, CQFD. ■

4.8 Applications définies sur un ouvert d'un produit

On se donne X_1, X_2, Y des espace affines normés et U un ouvert de $X_1 \times X_2$. Soient E_1, E_2, F les evn associés.

On se donne $f : U \to Y$.

On notera $x = (x_1,x_2)$ et $f(x) = f(x_1,x_2)$.

On rappelle (cours de Topologie) que l'ensemble U étant un ouvert, tout point $a = (a_1,a_2) \in U$ admet un voisinage dans U égal à un produit d'ouverts $U_1 \times U_2 \subset U$, avec $U_j \subset X_j$ ouvert. On peut donc définir les **applications partielles** $x_1 \mapsto f(x_1,a_2)$ et $x_2 \mapsto f(a_1,x_2)$, définies respectivement sur U_1, voisinage de a_1, et sur U_2, voisinage de a_2.

Définition 4.25 Soit $a = (a_1,a_2) \in U$.

1. Si l'application $x_1 \mapsto f(x_1,a_2)$ est différentiable en a_1, on dit que sa différentielle est la **différentielle partielle** de f par rapport à x_1 au point a. On la notera $D_1f(a) \in \mathscr{L}(E_1,F)$. Ou parfois $D_{x_1}f(a)$ ou $\partial_1 f(a)$ ou encore $\partial_{x_1} f(a)$.
2. Si l'application $x_2 \mapsto f(a_1,x_2)$ est différentiable en a_2, on dit que sa différentielle est la **différentielle partielle** de f par rapport à x_2 au point a. On la notera $D_2f(a) \in \mathscr{L}(E_2,F)$. Ou parfois $D_{x_2}f(a)$ ou $\partial_2 f(a)$ ou encore $\partial_{x_2} f(a)$.

Bien sûr on généralise à un nombre fini quelconque de facteurs X_i.

Exercice 4.26 Écrire la généralisation de la définition 4.25 à un nombre fini quelconque de facteurs X_i. ■

Proposition 4.27 Si f est différentiable en $a = (a_1,a_2)$ alors f admet des différentielles partielles par rapport à ses deux variables et on a, pour $h = (h_1,h_2) \in E_1 \times E_2$:

$$Df(a) \cdot h = D_1f(a) \cdot h_1 + D_2f(a) \cdot h_2\,.$$

Démonstration. Pour a fixé, l'application $x_1 \mapsto f(x_1,a_2)$ s'écrit aussi $f \circ j_1$ où $j_1(x_1) := (x_1,a_2)$. L'application j_1 est affine donc différentiable, et (exercice) $Dj_1(a_1)$ est l'application linéaire $h_1 \mapsto (h_1,0)$. Donc, par composition, $f \circ j_1$ est différentiable, et sa différentielle est

$$D_1 f(a) = D(f \circ j_1)(a_1) = Df(a) \circ Dj_1(a_1) = [h_1 \mapsto Df(a) \cdot (h_1, 0)]$$

On fait le même raisonnement pour $D_2 f(a)$, et en écrivant $Dj_1 \cdot h_1 + Dj_2 \cdot h_2 = (h_1, 0) + (0, h_2) = (h_1, h_2)$, on obtient la formule annoncée. ■

Exercice 4.28 Généraliser la proposition 4.27 à n facteurs $X_1 \times X_2 \times \cdots \times X_n$. ■

■ **Exemple 4.29 — Différentielle d'une forme bilinéaire continue.** Soit $f : E \times F \to \mathbb{R}$ bilinéaire continue. Montrer que f est différentiable et que, pour $a = (a_1, a_2) \in E \times F, h = (h_1, h_2) \in E \times F$, on a

$$Df(a) \cdot (h_1, h_2) = f(a_1, h_2) + f(h_1, a_2)$$

Démonstration. Développer les 4 termes de $f(a+h) = f(a_1 + h_1, a_2 + h_2)$. Montrer que le terme $f(h_1, h_2)$ est $o(h)$. ■

■

■ **Exemple 4.30 — Différentielle d'un « produit ».** Soit E un evn. Soit $E \ni x \mapsto A(x)$ dans $M_n(\mathbb{R})$ et $x \mapsto B(x)$ dans $M_n(\mathbb{R})$ des applications différentiables. Quelle est la différentielle de $F : x \mapsto A(x)B(x)$? (produit matriciel)

On l'écrit $\mathscr{C} \circ f(x)$ avec $f(x) = (A(x), B(x))$ et $\mathscr{C}(A,B) = AB$. L'application $\mathscr{C}$ est bilinéaire continue (voir la proposition 3.14) et donc, par l'exemple 4.29, différentiable, avec $D\mathscr{C}(A,B) = \mathscr{C}(A, \cdot) + \mathscr{C}(\cdot, B)$.

f est différentiable (par composantes) et $Df(x) = (DA(x), DB(x))$ où $DA(x) : E \to M_n(\mathbb{R})$.

Donc, puisque $F(x) = \mathscr{C} \circ f(x)$,

$$\begin{aligned} DF(x) \cdot u &= D\mathscr{C}(f(x)) \circ Df(x) \cdot u \\ &= \mathscr{C}(DA(x) \cdot u, B(x)) + \mathscr{C}(A(x), DB(x) \cdot u) \\ &= \underbrace{(DA(x) \cdot u)}_{\in M_n(\mathbb{R})}(B(x)) + (A(x))\underbrace{(DB(x) \cdot u)}_{\in M_n(\mathbb{R})}. \end{aligned}$$

Pour $n = 1$ on retrouve la formule de Leibniz (proposition 2.58) :

$$(A(x)B(x))' = A(x)'B(x) = A(x)B(x)',.$$

La même preuve fonctionne pour $A(x) \in \mathscr{L}(F,G)$ et $B(x) \in \mathscr{L}(E,F)$. ■

Attention, la réciproque de la proposition 4.27 est fausse. Par exemple $f(x,y) = xy/(x^2+y^2)$ et $f(0,0) = 0$: les applications partielles en $(0,0)$ sont nulles donc dérivables, mais f n'est pas continue en 0 donc pas différentiable. Cependant, pour la classe C^1 sur un *ouvert* ça marche :

Théorème 4.31 Soit f une application continue d'un ouvert U d'un produit $X = X_1 \times \cdots \times X_n$ de n espaces affines normés (X_i, E_i), dans un espace affine normé (Y, F). L'application f est de classe C^1 sur U *si et seulement si* les n différentielles partielles de f existent en tout point de U et sont des applications continues de U dans $\mathscr{L}(E_i, F)$, $1 < i < n$.

Démonstration.

Étape 1. Si f est C^1 en $a = (a_1, \dots, a_n)$, montrons que les différentielles partielles sont continues : on rappelle que la $i$ème différentielle partielle de f en a est la différentielle en a_i de $f \circ j_i$ avec

$$j_i(x_i) = (a_1, \dots, a_{i-1}, x_i, a_{i+1}, \dots, a_n) .$$

On rappelle aussi que j_i est affine donc différentiable et $Dj_i(x_i)$ est l'application (indépendante de x_i, on la notera $\tilde{j}_i$) d'injection de E_i dans le produit :

$$\tilde{j}_i : h_i \mapsto (0, \dots, 0, h_i, 0, \dots, 0) \in E_1 \times \cdots \times E_n .$$

Donc

$$\partial_j f(a) = Df(a) \circ \tilde{j}_i$$

qui est bien continue en a, car on a vu (Exemple 4.30) que l'application de composition bilinéaire $\mathscr{C} : (A, B) \mapsto A \circ B$ est continue.

Étape 2 : réciproque. On traite le cas $n = 2$. (Le cas général s'obtient alors par récurrence.)

Supposons que les différentielles partielles existent et sont continues : on a alors un candidat nécessaire pour la différentielle de f au point x, qui est

$$g(x) \cdot v = D_1 f(x) \cdot v_1 + D_2 f(x) \cdot v_2 .$$

Pour chaque x, c'est une application linéaire continue (par composition, car la projection $p_j : v \mapsto v_j$ est continue) et l'application $x \mapsto g(x) \in L(E, F)$ est continue, car le premier terme se factorise en :

$$x \mapsto (D_1 f(x), p_1) \mapsto \mathscr{C}(D_1 f(x), p_1) = D_1 f(x) \circ p_1$$

(et idem pour le 2ème terme).

On considère maintenant $f(x+h) - f(x) - g(x) \cdot h$: on voudrait montrer que c'est $o(h)$. On a

$$\begin{aligned} f(x+h) - f(x) - g(x) \cdot h = {} & f(x+h) - f(x_1, x_2 + h_2) - D_1 f(x) \cdot h_1 \\ & + f(x_1, x_2 + h_2) - f(x) - D_2 f(x) \cdot h_2 . \end{aligned}$$

Donc

$$\| f(x+h) - f(x) - g(x) \cdot h \| \leqslant N_1 + N_2$$

avec

$$N_1 := \| f(x+h) - f(x_1, x_2 + h_2) - D_1 f(x) \cdot h_1 \|$$

et

$$N_2 := \| f(x_1, x_2 + h_2) - f(x) - D_2 f(x) \cdot h_2 \| .$$

Par le fait que la dérivée partielle $D_2 f(x)$ existe, ce dernier terme N_2 est $o(h_2)$. Ainsi, pour tout $\varepsilon > 0$ fixé, on a un $\alpha_1 > 0$ tel que $N_2 < \varepsilon \|h_2\|$ dès que $\|h_2\| < \alpha_1$.

Mais pour le premier terme N_1, on n'est pas exactement au bon point… on a $D_1 f(x)$ et on voudrait $D_1 f(x_1, x_2 + h_2)$. Il faut utiliser la continuité de $D_1 f$. Notons $\tilde{x} = (x_1, x_2 + h_2)$. On avait fixé ci-dessus $\varepsilon > 0$. Par continuité de $D_1 f$, on a un $\alpha_2 > 0$ tel que

$$\|D_1 f(x) - D_1 f(\tilde{x})\| < \varepsilon$$

dès que $\|x - \tilde{x}\| < \alpha_2$. Par définition de la norme subordonnée, on a donc

$$\|D_1 f(x) \cdot h_1 - D_1 f(\tilde{x}) \cdot h_1\| < \varepsilon \|h_1\| .$$

Écrivons donc

$$\begin{aligned} N_1 &= \|f(x+h) - f(\tilde{x}) - D_1 f(x) \cdot h_1\| \\ &\leqslant \|f(x+h) - f(\tilde{x}) - D_1 f(\tilde{x}) \cdot h_1\| + \|D_1 f(x) \cdot h_1 - D_1 f(\tilde{x}) \cdot h_1\| . \end{aligned}$$

Le premier terme est $o(h_1)$ par existence de $D_1 f(\tilde{x})$, on obtient donc $N_1 < 2\varepsilon \|h_1\|$ si $\|h_1\|$ est assez petit.

On a donc $N_1 + N_2 \leqslant 2\varepsilon(\|h_1\| + \|h_2\|)$ pour h assez petit, ce qui dit bien $N_1 + N_2 = o(h)$, CQFD@. ■

Autre preuve de la dernière étape. On remarque que $N_1 = \|F(h_1) - F(0)\|$ avec h_2 fixé et

$$F : h_1 \to f(x_1 + h_1, x_2 + h_2) - D_1 f(x_1, x_2) \cdot (h_1) .$$

F elle est différentiable (pourquoi ?) et $DF(h_1) = D_1 f(x_1 + h_1, x_2 + h_2) - D_1 f(x_1, x_2)$ qui est majoré en norme par ε si $\|h\|$ est assez petit (continuité de Df_1). Donc par les accroissements finis :

$$\|F(h_1) - F(0)\| \leqslant \varepsilon \|h_1\| .$$

(si h est petit, le segment $[0, h_1]$ est dans le domaine de F)

Soit : $N_1 \leqslant \varepsilon \|h_1\|$. ■

4.9 Dimension finie

C'est bien beau de connaître la théorie générale, mais il faut en priorité savoir l'appliquer en dimension finie. Pour cela, des notations spécifiques sont utiles.

4.9.1 Dérivées partielles

On considère la différentielle d'une fonction *scalaire* $f : U \to \mathbb{R}$, U ouvert de $\mathbb{R}^n$. On écrit $f(x) = f(x_1, \ldots, x_n)$ et on note $\frac{\partial f}{\partial x_i}(x)$ la $i$ème différentielle partielle. Mais attention, on va utiliser ici la convention qu'on a prise pour des fonctions d'une variable (en effet, l'application partielle $x_i \mapsto f(x)$ est une fonction d'une variable) :

▪ **Notation 4.32** La **dérivée partielle** $\frac{\partial f}{\partial x_i}(x)$ désigne le **nombre réel** égal à la dérivée de l'application partielle $x_i \mapsto f(x)$.

Si on revient à la convention générale (définition 4.25), alors la différentielle partielle $D_i f(x)$ est un élément de $\mathscr{L}(\mathbb{R},\mathbb{R})$, autrement dit une matrice $1 \times 1\dots$ dont l'unique coefficient est $\frac{\partial f}{\partial x_i}(x)$.

On note souvent $\mathrm{d}f$ au lieu de Df pour la différentielle. (En fait, $\mathrm{d}f$ désigne la *différentielle extérieure*, mais ici — c'est-à-dire lorsque f est une fonction scalaire, il n'y a pas de différence avec Df : en chaque point x, $Df(x) \in \mathscr{L}(\mathbb{R}^n,\mathbb{R})$: c'est une forme linéaire sur $\mathbb{R}^n$).

En géométrie différentielle, on note aussi x_i l'application « $i$ème coordonnée » : $x_i(x) = x_i$ (eh oui, la notation peut surprendre...) C'est juste la projection sur le $i$ème facteur, qu'on a pu noter précédemment p_i. Elle est linéaire, donc sa différentielle $\mathrm{d}x_i$ en tout point est égale à elle-même, p_i. On obtient ainsi, à partir de la proposition 4.27, la célèbre formule :

Proposition 4.33

$$\mathrm{d}f(x) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x)\,\mathrm{d}x_i$$

R **Remarque 4.34** $\mathrm{d}f(x)$, étant une forme linéaire sur $\mathbb{R}^n$, se voit comme un « vecteur ligne », c'est utile pour calculer $\mathrm{d}f(x) \cdot h$ où h est un vecteur colonne, avec les formules habituelles de produit matriciel.

4.9.2 Matrice jacobienne

On se donne encore U ouvert de $\mathbb{R}^n$ et on considère maintenant

$$f : U \to \mathbb{R}^m, \quad f = (f_1, \dots, f_m)$$

La proposition 4.20 dit que $Df = (\mathrm{d}f_1, \dots, \mathrm{d}f_m)$: autrement dit, on a une collection de m vecteurs lignes, chacun étant dans $\mathbb{R}^n$, ce qui forme naturellement une **matrice** $m \times n$:

$$\begin{pmatrix} \mathrm{d}f_1 \\ \mathrm{d}f_2 \\ \vdots \\ \mathrm{d}f_m \end{pmatrix}$$

C'est ce qu'on appelle la matrice jacobienne de f.

Définition 4.35 La **matrice jacobienne**[a] de f est la matrice

$$\left(\frac{\partial f_i}{\partial x_j}(x)\right)_{\substack{i=1,\dots,m \\ j=1,\dots,n}} \in \mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$$

a. du mathématicien Charles JACOBI, 1804–1851

On rappelle que i est l'indice de ligne parmi m, tandis que j est l'indice de colonne parmi n.

On obtient immédiatement :

Proposition 4.36 Pour calculer $Df(x) \cdot h$ on fait le produit matriciel de la matrice jacobienne avec le vecteur colonne h.

On n'a donc en général pas de notation spécifique pour la matrice jacobienne, on utilise la même notation que la différentielle $Df(x)$… ce qui est cohérent avec notre notation $Df(x)(h) = Df(x) \cdot h$, qui se trouve ici être le produit d'une matrice $m \times n$ avec un vecteur colonne $h \in \mathbb{R}^n$. Autrement dit, la matrice jacobienne est la matrice qui représente l'application linéaire $Df(x)$ dans les bases canoniques de $\mathbb{R}^n$ et $\mathbb{R}^m$.

Lorsque $n = m$, $Df(x)$ est un *endomorphisme* de $\mathbb{R}^n$. Il admet donc un déterminant, qui est indépendant du choix d'une base de $\mathbb{R}^n$. C'est (donc) aussi le déterminant de la matrice jacobienne.

Définition 4.37 Déterminant jacobien

$$J(x) = \det(\text{matrice jacobienne}) = \det Df(x)\,.$$

4.9.3 Gradient

Si X est un espace Euclidien (typiquement, $X = \mathbb{R}^n$, muni de la norme euclidienne $\|\cdot\|_2$), alors on utilise fréquemment le *gradient* d'une fonction $f : X \to \mathbb{R}$. Quel est le lien avec la différentielle ?

Rappelez-vous que la différentielle $Df(x)$ est une forme linéaire sur l'espace vectoriel E. Ici, on a $E = \mathbb{R}^n$ et on a identifié cette forme linéaire à une vecteur *ligne*, de sorte que, si $v \in E$, l'évaluation $Df(x) \cdot v$ est exactement le produit matriciel de la ligne $Df(x)$ avec la colonne v.

Maintenant qu'on dispose d'un produit scalaire, on pourrait aussi vouloir écrire $Df(x) \cdot v$ comme le produit scalaire de v avec un vecteur. Ce vecteur s'appelle **le gradient de f (au point x)** on le note $\nabla f(x)$:

Définition 4.38 Soit $\langle \cdot, \cdot \rangle$ un produit scalaire euclidien sur E. Le **gradient** $\nabla f(x)$ est défini par

$$\forall v \in E, \quad Df(x) \cdot v = \langle \nabla f(x), v \rangle\,.$$

Exercice 4.39 Soit $f : \mathbb{R}^n \to \mathbb{R}$ différentiable en x. Avec le produit scalaire usuel de $\mathbb{R}^n$, le gradient de f est simplement le vecteur colonne $(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x))$, c'est-à-dire le *transposé* de la différentielle $Df(x)$, lorsque cette dernière est identifiée à la matrice jacobienne de f dans la base canonique. ▪

5. Chemins dans un espace de Banach

Le calcul différentiel est intimement relié au *calcul intégral*, en particulier par la formule emblématique dite « formule de la moyenne » (théorème 5.18) et ses généralisations que sont les formules de Taylor.

Cependant, pour disposer d'un calcul intégral, nous aurons besoin d'une *hypothèse supplémentaire* sur les espaces vectoriels normés utilisés. C'est la notion d'espaces de Banach.

5.1 Suites de Cauchy dans un EVN, espaces de Banach

Davantage de détails sur cette partie pourront être trouvés dans le cours de Topologie.

Définition 5.1 Soit (u_n) une suite dans un EVN. On dit que (u_n) est **de Cauchy** si : Pour tout $\varepsilon > 0$ il existe un rang N tel que $\|u_n - u_m\| < \varepsilon$ dès que n et m sont $\geqslant N$.

Exercice 5.2 Dans un EVN, toute suite convergente est de Cauchy. ■

La réciproque n'est pas vraie en général : il peut arriver (dans un EVN de dimension infinie) qu'il existe des suites de Cauchy qui ne soient pas convergentes. Voir le cours de Topologie.

Définition 5.3 Un **espace de Banach** est un EVN **complet**, c'est-à-dire un EVN où *toute* suite de Cauchy est convergente.

Remarque 5.4 La notion de suite de Cauchy s'adapte parfaitement au cas d'un espace *affine* normé (X,E). En effet $u_n - u_m$ est toujours un vecteur de E, l'espace vectoriel normé associé.

Remarque 5.5 — $\star$. Il y a *deux* généralisations possibles de la notion de suites de Cauchy et donc d'espaces complets :

1. suites de Cauchy dans un espace vectoriel topologique (hors programme) ;
2. suites de Cauchy dans un espace métrique (TOPG). Cela inclut le cas d'un espace affine mentionné ci-dessus.

5.2 Intégrale d'un chemin dans un Banach

Nous allons utiliser la technique des *sommes de Riemann*. Soit E un EVN. Soit $f : [a,b] \to E$ une fonction continue. On se donne une subdivision (t_i) de $[a,b]$: on prendra en général un pas constant $h = (b-a)/N$, et donc $t_i = a + ih$, avec $i = 0, \ldots, N$, de sorte que $t_0 = a$, $t_N = b$, et il y a N sous-intervalles. (On dira que (t_i) est une subdivision *régulière*. Plus généralement une subdivision de taille N est une famille finie croissante $(t_i)_{i=0,\ldots,N}$: $t_i \leqslant t_{i+1}$, avec $t_0 = a$ et $t_N = b$.)

Définition 5.6 Soit $f : [a,b] \to E$. La **somme de Riemann** de f pour la subdivision $t_i = a + i(b-a)/n$ est

$$S_N(f) = \sum_{i=1}^{N} (t_i - t_{i-1}) \cdot f(t_i) = \sum_{i=1}^{N} h f(t_i) = \frac{b-a}{N} \sum_{i=1}^{N} f(t_i)$$

On montre alors :

Proposition 5.7 Si $f : [a,b] \to E$ est continue, alors la somme de Riemann $(S_N(f))_{N\in\mathbb{N}}$ est une suite de Cauchy dans E.

Démonstration. On considère donc la différence $S_N(f) - S_M(f)$. Notons x_i la subdivision utilisée dans $S_N(f)$, soit : $x_i = a + (i/N)(b-a)$, et y_j la subdivision utilisée dans $S_M(f)$, c'est-à-dire $y_j = a + (j/M)(b-a)$

Notons (t_k) la subdivision (non régulière) obtenue en prenant l'union des (x_i) et des (y_j) (avec répétition le cas échéant : k varie de 0 à $M+N+1$, et on a $t_k \leqslant t_{k+1}$), et notons $S_{N,M}(f)$ la somme de Riemann correspondante :

$$S_{N,M}(f) = \sum_{k=1}^{N+M+1} (t_k - t_{k-1}) \cdot f(t_k) \,.$$

On regroupe les t_k qui tombent dans les intervalles $]x_{i-1}, x_i]$: on écrit donc

$$S_{N,M}(f) = \sum_{i=1}^{N} \sum_{t_k \in]x_{i-1},x_i]} (t_k - t_{k-1}) \cdot f(t_k) \,.$$

Comparons maintenant $S_N(f)$ avec $S_{N,M}(f)$:

$$S_N(f) - S_{N,M}(f) = \sum_{i=1}^{N} \left(\frac{b-a}{N} f(x_i) - \sum_{t_k \in]x_{i-1},x_i]} (t_k - t_{k-1}) \cdot f(t_k) \right) .$$

On remarque qu'en sommant la longueur des sous-intervalles on obtient la longueur de $]x_{i-1}, x_i]$:

$$\sum_{t_k \in]x_{i-1},x_i]} (t_k - t_{k-1}) = x_i - x_{i-1} = \frac{b-a}{N}$$

et du coup la somme précédente se réécrit en

$$S_N(f) - S_{N,M}(f) = \sum_{i=1}^{N} \left(\sum_{t_k \in]x_{i-1},x_i]} (t_k - t_{k-1}) \cdot (f(x_i) - f(t_k)) \right) .$$

Soit $\varepsilon > 0$. Par *continuité uniforme* de f sur le compact $[a,b]$ (théorème de Heine), si N est assez grand, $\|f(x) - f(y)\| < \frac{\varepsilon}{2(b-a)}$ lorsque $|x-y| < (b-a)/N$. C'est donc le cas pour tous les $\|f(x_i) - f(t_k)\|$ qui apparaissent dans la somme.

Donc

$$\|S_N(f) - S_{N,M}(f)\| \leqslant \sum_{i=1}^{N} \left(\sum_{t_k \in]x_{i-1},x_i]} \frac{\varepsilon}{2(b-a)}(t_k - t_{k-1}) \right) = \frac{\varepsilon}{2(b-a)} \sum_{i=1}^{N} \frac{b-a}{N} = \frac{\varepsilon}{2}.$$

Bien sûr, la même estimation est valable aussi pour $||S_M(f) - S_{N,M}(f)|| \leqslant \varepsilon/2$ si M est également assez grand. On obtient finalement, pour N et M assez grands :

$$\|S_M(f) - S_N(f)\| \leqslant \|S_M(f) - S_{N,M}(f)\| + ||S_{N,M}(f) - S_N(f)|| \leqslant \varepsilon.$$

La suite est donc bien de Cauchy. ■

R **Remarque 5.8** Si f dépend d'un autre paramètre s de sorte que $(t,s) \mapsto f(t,s)$ soit continue, alors la preuve est encore valable uniformément sur un voisinage de s : en effet dans la preuve du théorème de Heine, il suffit de recouvrir le compact $[a,b] \times \{s\}$ par un nombre fini de voisinages produits. [1] La suite $S_N(f(\cdot,s))$ est alors *uniformément de Cauchy* par rapport à s.

De la proposition 5.7, il découle qu'on peut dorénavant *définir* l'intégrale de f :

Théorème 5.9 — et définition. Soit I un intervalle de $\mathbb{R}$, et soit $f : I \to E$ un *chemin* (donc continu) dans un espace de Banach E. Soient a,b dans I. La somme de Riemann $S_N(f)$ associée est une suite de Cauchy dans E donc converge, et on définit :

$$\int_a^b f = \int_a^b f(t)\,\mathrm{d}t := \lim_{N\to\infty} S_N(f).$$

Quelques remarques :

R **Remarque 5.10** On peut montrer (de la même façon) que la limite de la somme de Riemann « générale » $S_N(f) = \sum_{i=1}^{N}(t_i - t_{i-1})f(t_i)$ ne dépend pas de la subdivision (t_i) choisie, non nécessairement régulière, tant que les pas tendent uniformément vers 0, càd $\max_i |t_i - t_{i-1}| \to 0$ quand $N \to \infty$.

R **Remarque 5.11** Si $b < a$, la même formule pour la somme de Riemann fonctionne (avec $h < 0$) et on obtient $\int_a^b f = -\int_b^a f$.

Proposition 5.12 Si f dépend continûment d'un autre paramètre s (voir remarque 5.8 ci-dessus), alors la convergence de $S_N(f)$ est uniforme dans un voisinage V de s :

$$\sup_{s\in V} \left\| S_N(f(\cdot,s)) - \int_a^b f(\cdot,s) \right\| \to 0 \qquad \text{quand } N \to \infty.$$

On en déduit en particulier que l'application $s \mapsto \int_a^b f(\cdot,s)$ est continue.

Démonstration. Lorsque E est un Banach, l'espace $C_b^0(V,E)$ (fonctions continues bornées) est un Banach pour la norme uniforme (voir Topologie). Autrement dit une suite uniformément de Cauchy est uniformément convergente. ■

1. Cette idée est utilisée également dans la preuve du théorème 5.31 plus loin.

Exercice 5.13 — relation de Chasles.

$$\int_a^b f + \int_b^c f = \int_a^c f.$$

(il faut utiliser l'indépendance de la subdivision, cf remarque 5.10 ci-dessus) ■

Définition 5.14 On étend l'intégrale aux fonctions **continues par morceaux** [a] de façon habituelle : on somme les intégrales sur chaque intervalle où f est continue de façon à satisfaire la relation de Chasles.

a. L'intervalle I est décomposé en un nombre fini de sous intervalles I_j, et pour chaque j, la restriction de f à l'intervalle ouvert $\mathring{I}_j$ est continue, et admet un prolongement par continuité au sous-intervalle fermé $\overline{I_j}$. En d'autres termes, il y a des limites à gauche et à droite partout.

On récupère facilement les propriétés habituelles de l'intégrale, par exemple :
— $\int_a^a f = 0$
— si $x_0 \in E$, l'intégrale de la fonction constante vaut : $\int_a^b x_0 = (b-a) \cdot x_0$
et aussi :

Proposition 5.15 Si $a \leqslant b$, on a

$$\left\| \int_a^b f(t)dt \right\| \leqslant \int_a^b \|f(t)\| \,\mathrm{d}t\,.$$

Pour tous $a, b \in I$, et pour la norme du sup, l'application

$$C^0([a,b],E) \ni f \mapsto \int_a^b f \in E$$

est linéaire continue de norme $|b-a|$. (Si $E \neq \{0\}$.)

Démonstration.

1. La linéarité de l'intégrale provient de la linéarité de $f \mapsto S_N(f)$, qui est préservée par passage à la limite.
2. On écrit

$$\left\| \int_a^b f \right\| = \|\lim S_N(f)\| = \lim \|S_N(f)\| \leqslant \lim |S_N(g)|\,,$$

où g est la fonction $g(t) = \|f(t)\|$. (On a utilisé la continuité de la norme pour la deuxième égalité.) Or $\lim S_N(g) = \int g$, on obtient donc bien $\|\int f\| \leqslant |\int g|$ (qui est $= \int g$ si $a \leqslant b$).
3. Du coup, pour calculer la norme subordonnée de l'application « $\int_a^b$ », on écrit

$$\left\| \int_a^b f \right\| \leqslant \left| \int_a^b \sup g \right| = |b-a|\,\|g\|_\infty = |b-a|\,\|f\|_\infty\,,$$

et cette norme est facilement atteinte avec une fonction constante.

■

Exercice 5.16 — TD. L'espace $C([a,b];E)$, muni de la norme L^1 : $\|f\|_1 = \int_a^b \|f(t)\|\,\mathrm{d}t$, est un EVN *non complet* (si $a<b$ et $E \neq \{0\}$).

Proposition 5.17 Si $A \in \mathscr{L}(E,F)$, alors $\int(Af) = A(\int f)$.

Démonstration. $\int(Af) = \lim S_N(Af) = \lim A(S_N(f))$ par linéarité de A, ce qui est égal à $A(\lim S_N(f))$ par continuité de A, c'est donc $= A(\int f)$. ■

Le résultat suivant ne vous surprendra sans doute pas, mais il reste « magique » malgré tout ! Il exprime le lien entre dérivée et intégrale, deux notions qui ont chacune été définies indépendamment, sans aucune relation l'une avec l'autre.

Théorème 5.18 — Formule de la moyenne. (Appelé aussi : *théorème fondamental du calcul infinitésimal.*)
Soit I un intervalle de $\mathbb{R}$. Soit E un Banach. Pour $f: I \to E$ continue sur I, et C^1 (par morceaux [a]), et a,b dans I, on a :

$$f(b) - f(a) = \int_a^b f'(t)\,\mathrm{d}t\,. \tag{5.1}$$

[a]. plus précisément, on demande que f soit C^1 sur les sous-intervalles ouverts, et que f' y admette des limites à gauche et à droite.

Démonstration. Montrons le lemme :

Lemme 5.1 Si g est continue par morceaux, la fonction $G: t \to \int_a^t g$ est continue. Si g est continue au point t, alors G est différentiable en t, de dérivée $G'(t) = g(t)$.

Preuve du lemme. La continuité provient du fait que $\|g(t)\|$ est bornée : car alors, par la proposition 5.15, G est Lipschitz.

Pour la différentiabilité, on considère :

$$G(t+h) - G(t) = \int_t^{t+h} g(s)\,\mathrm{d}s = hg(t) + \int_t^{t+h} (g(s) - g(t))\,\mathrm{d}s$$

(par Chasles). Par continuité de g en t, si $|h|$ est assez petit, $\|g(s)-g(t)\| < \varepsilon$, donc $\left|\int_t^{t+h}(g(s)-g(t))\,\mathrm{d}s\right| < h\varepsilon$. Autrement dit, ce terme est $o(h)$ lorsque $h \to 0$, ce qui prouve le lemme. ■

Preuve du théorème. La fonction $F(t) = f(t) - f(a) - \int_a^t f'$ a donc une dérivée nulle en tout point où f' est continue. Par le théorème des accroissements finis, cette fonction doit être constante sur tout intervalle où f' est continue. Elle est donc globalement constante, par continuité.

Donc $F(b) = F(a) = 0$, ce qui donne le résultat. ■

Remarque 5.19 La formule de la moyenne est encore valable si $f: I \to X$ où (X,E) est un espace affine. (On rappelle que $f'(t) \in E$ dans tous les cas, donc l'intégrale est bien définie.)

Voyons maintenant quelques conséquences intéressantes de cette formule.

Exercice 5.20 Si $f :]a,b[\to E$, où E est un Banach, est C^1 et f' est bornée, alors f admet un unique prolongement par continuité à $[a,b]$. (Et la formule de la moyenne (5.1) est alors valable pour la fonctions prolongée). ■

Proposition 5.21 — Intégration par parties. Si $g : I \to \mathbb{R}$ et $f : I \to E$ Banach sont C^1, alors

$$\int_a^b g(t)\cdot f'(t)\,\mathrm{d}t = [g(t)\cdot f(t)]_a^b - \int_a^b g'(t)\cdot f(t)\,\mathrm{d}t\,.$$

On a noté

$$[g(t)\cdot f(t)]_a^b := g(b)\cdot f(b) - g(a)\cdot f(a)\,.$$

Démonstration. Appliquer la formule de la moyenne à $t \to g(t)\cdot f(t)$. ■

Proposition 5.22 — Changement de variables dans l'intégrale. Soient $f : J \to E$ continue et $g : I \to J$ de classe C^1 par morceaux, où I et J sont des intervalles de $\mathbb{R}$. Alors, pour a,b dans I

$$\int_{g(a)}^{g(b)} f(t)\,\mathrm{d}t = \int_a^b g'(u)\cdot f(g(u))\,\mathrm{d}u\,. \tag{5.2}$$

Démonstration. Appliquer la formule de la moyenne à $\gamma\circ g$, avec $\gamma(s) := \displaystyle\int_{g(a)}^{s} f(t)\,\mathrm{d}t$. En effet, par le lemme 5.1, γ est bien C^1, de dérivée $\gamma' = f$. Donc $\gamma\circ g$ est C^1 par morceaux, et on peut appliquer a formule de la moyenne :

$$\gamma\circ g(b) - \gamma\circ g(a) = \int_a^b (\gamma\circ g)'(u)\,\mathrm{d}u = \int_a^b D\gamma(g(u))\cdot\gamma'(u)\,\mathrm{d}u = \int_a^b \gamma'(u)\cdot\gamma'(g(u))\,\mathrm{d}u\,,$$

ce qui donne bien la formule voulue. ■

R **Remarque 5.23** Dans la formule de changement de variables (5.2), remarquez que g n'est pas nécessairement bijectif.

R **Remarque 5.24** Le moyen mnémotechnique standard pour les changements de variables est de poser $t = g(u)$ et de calculer la « différentielle » : $\mathrm{d}t = g'(u)\,\mathrm{d}u$. Ceci suggère la notion d'intégrale d'une « forme différentielle », que nous n'aborderons pas ici.

La formule de la moyenne est un résultat pour les fonctions d'une variable ; mais on en déduit facilement des applications au cas d'un nombre quelconque de variables, grâce à la notion de *segment* que nous avons introduite à la définition 4.21 : le segment $[x_0,x_1]$ dans X affine c'est l'ensemble $\{x_0 + t(x_1 - x_0), t \in [0,1]\}$. On obtient alors

Corollaire 5.25 — formule de la moyenne, bis. Soient (X,E) et (Y,F) deux espaces affines normés, et $f : U \to Y$ une application C^1 sur un ouvert U de X. On suppose que F est un Banach. Soient x_0 et x_1 deux points de U tels que le segment $[x_0,x_1]$ soit contenu dans U. Alors

$$f(x_1) - f(x_0) = \int_0^1 Df(x_0 + t(x_1 - x_0)) \cdot (x_1 - x_0)\,\mathrm{d}t\,.$$

Démonstration. Poser $g(t) = f(x_0 + t(x_1 - x_0))$ pour $t \in [0,1]$ et appliquer la formule de la moyenne (théorème 5.18). ■

R **Remarque 5.26** Cette proposition permet de redémontrer directement le TAF (théorème 4.22). Cependant ici on a dû supposer que F était un Banach.

Corollaire 5.27 Soit U un ouvert *convexe* dans un espace affine normé X et $f : U \to Y$ une application C^1 à valeurs dans un espace affine normé (Y,F) où F est un espace de Banach. On suppose que $Df(x) = 0$ pour tout $x \in U$. Alors f est constante.

Démonstration. Appliquer le corollaire 5.25 à deux points quelconques $x_0, x_1 \in U$. ■

Exercice 5.28 Montrer que le corollaire 5.27 reste vrai si U est un ouvert *connexe*. (Voir cours de Topologie.) ■

5.3 Formules de Taylor avec reste intégral

Lorsqu'une fonction est plusieurs fois dérivable, on peut *itérer* la formule de la moyenne, ce qui donne les fameuses *formules de Taylor*.

On a vu, déjà, plusieurs variantes de ces formules, qu'on avait obtenues sans avoir recours à l'intégration (donc sans supposer E Banach) :

— Le développement limité de Taylor (théorème 2.77) : en supposant que f est k fois dérivable, on obtient un reste $o(h^k)$.

— Le reste de Lagrange (théorème 2.82) : si f est $k+1$ fois dérivable, alors le reste est majoré par

$$\frac{h^{k+1}}{(k+1)!} \sup_{[t,t+h]} \left\| f^{(k+1)} \right\|.$$

Maintenant, avec les hyopthèses plus fortes que f est de classe C^{k+1}, et est à valeurs dans un espace de Banach, on obtient *une formule explicite pour le reste*, qui est souvent la plus pratique à utiliser.

Théorème 5.29 — Taylor avec reste intégral. Soit $f : I \to E$ un chemin C^{k+1}, $k \in \mathbb{N}$, dans un espace de Banach E. Soit $t \in I$ et $h \in \mathbb{R}$ tel que $t + h \in I$. Alors

$$f(t+h) = f(t) + \sum_{j=1}^{k} \frac{h^j}{j!} \cdot f^{(j)}(t) + h^{k+1} \int_0^1 \frac{(1-s)^k}{k!} \cdot f^{(k+1)}(t+sh)\,\mathrm{d}s\,.$$

Démonstration. Pour $k = 0$, c'est la formule de la moyenne.

Supposons la formule vraie pour un rang $k \geqslant 0$. Par intégration par parties (proposition 5.21) : avec

$$g(s) = -\frac{(1-s)^{k+1}}{(k+1)!}, \quad g'(s) = \frac{(1-s)^k}{k!}$$

on obtient, en supposant $f \in C^{k+2}$,

$$\int_0^1 \frac{(1-s)^k}{k!} \cdot f^{(k+1)}(t+sh)\,\mathrm{d}s = \frac{1}{(k+1)!} \cdot f^{(k+1)}(t) + h \cdot \int_0^1 \frac{(1-s)^{k+1}}{(k+1)!} \cdot f^{(k+2)}(t+sh)\,\mathrm{d}s,$$

ce qui donne la formule au rang $k+1$. ∎

R **Remarque 5.30** Comme d'habitude, on aurait pu faire entrer le terme $f(t)$ dans la somme en débutant à $j=0$, mais on le présente comme ça pour mieux voir que la formule s'étend au cas affine : $f(t) \in X$, et le « reste » est un vecteur dans E.

5.4 Interversions limite/dérivée/intégrale

Le reste intégral

$$\int_0^1 \frac{(1-s)^k}{k!} \cdot f^{(k+1)}(t+sh)\,\mathrm{d}s$$

de la formule de Taylor (théorème 5.29) est une fonction **continue** de (t,h) ; ça découle de la proposition 5.12 et du fait que $f^{(k+1)}$ soit continue. En fait, un grand intérêt de cette formule est que ce reste est *aussi régulier que* $f^{(k+1)}$. Par exemple, si $f \in C^\infty$, alors le reste intégral sera une fonction C^∞ de (t,h). C'est une conséquence du théorème suivant.

Théorème 5.31 — Dérivation sous l'intégrale. Soient E un espace vectoriel (ou affine) normé et F un Banach. Soit I un intervalle de $\mathbb{R}$, et $U \subset E$ un ouvert. Soit

$$K : I \times U \to F \quad \text{continue, avec} \quad \partial_x K \text{ continue sur } I \times U\,.$$

Alors, pour $a, b \in I$, l'application « intégrale partielle »

$$f(x) := \int_a^b K(t,x)\,\mathrm{d}t$$

est de classe C^1 sur U, et

$$Df(x) \cdot h = \int_a^b \partial_x K(t,x) \cdot h\,\mathrm{d}t\,, \quad \forall h \in E\,.$$

Démonstration. (Voir aussi le TD.) Pour commencer, notons que la quantité $f(x)$ est bien définie car $K(x,\cdot)$ est continue (restriction d'une fonction continue) et F est un Banach.

Si on utilise simplement que $\partial_x K$ existe, on a, pour x et t fixés,

$$K(t,x+h) = K(t,x) + \partial_x K(t,x) \cdot h + o(h)\,;$$

on voudrait intégrer cette équation sur $[a,b]$ pour obtenir le résultat voulu. Le problème vient du terme de reste $o(h)$, qui dépend en fait de x et t : d'une part, est-il continu en t ? (afin de l'intégrer) : il faut pour cela que $t \mapsto \partial_x K(t,x) \cdot h$ soit continu, et d'autre part, la petitesse par rapport à h est-elle uniforme par rapport à t ?

C'est afin de contrôler ce terme $o(h)$ qu'on va supposer que $\partial_x K(t,x)$ est continue par rapport aux deux variables x et t. La continuité par rapport à x permet d'appliquer la formule de la moyenne (corollaire 5.25) à l'application partielle $x \mapsto K(t,x)$:

$$K(t,x+h) = K(t,x) + \int_0^1 \partial_x K(t,x(s)) \cdot h\,\mathrm{d}s\,, \tag{5.3}$$

avec $x(s) := x + s \cdot h$. Par continuité de $\partial_x K$ par rapport à (x,t), pour $\varepsilon > 0$, $t_0 \in [a,b]$ et $x \in U$ fixés, il existe $\alpha(t_0) > 0$ tel que, dès que $\|h\| < \alpha(t_0)$ et $|t - t_0| < \alpha(t_0)$,

$$\|\partial_x K(t, x(s)) - \partial_x K(t,x)\| < \varepsilon\,, \quad \forall s \in [0,1]\,. \tag{5.4}$$

L'ensemble des intervalles ouverts $I_{t_0} =]t_0 - \alpha(t_0), t_0 + \alpha(t_0)[$, lorsque t_0 varie dans $[a,b]$, forme un recouvrement du compact $[a,b]$; on peut donc le recouvrir avec un nombre fini de tels intervalles : $I_{t_0}, \ldots, I_{t_N}$. En choisissant $\alpha = \min \alpha(t_j)$ on obtient que (5.4) reste vraie pour tout $t \in [a,b]$.

Donc, en intégrant sur $[0,1]$, $\| \int_0^1 [\partial_x K(t,x(s)) - \partial_x K(t,x)] \cdot h \, \mathrm{d}s\| < \varepsilon \|h\|$. Par intégration de (5.3), en utilisant le fait que le terme $\partial_x K(t,x) \cdot h$ ne dépend pas de s et donc s'intègre sur $[0,1]$ en lui-même,

$$\int_a^b K(t, x+h)\,\mathrm{d}t = \int_a^b K(t,x)\,\mathrm{d}t + \int_a^b \partial_x K(t,x) \cdot h \,\mathrm{d}s + R$$

avec $\|R\| < \left| \int_a^b \varepsilon \|h\| \right| = |b-a|\, \varepsilon \|h\|$. Donc

$$f(x+h) = f(x) + \int_a^b \partial_x K(t,x) \cdot h\,\mathrm{d}s + o(h)\,,$$

ce qui donne le résultat. ■

R **Remarque 5.32** Le terme $\partial_x K(t,x)$ appartient à $\mathscr{L}(E,F)$ qui est lui-même un Banach (puisque F est un Banach). On peut donc définir $\int_a^b \partial_x K(t,x)\,\mathrm{d}t \in \mathscr{L}(E,F)$, et par continuité de l'application $A \mapsto A(h)$, $A \in \mathscr{L}(E,F)$, on obtient (utiliser la définition par les sommes de Riemann)

$$\int_a^b \partial_x K(t,x) \cdot h\,\mathrm{d}s = \left[\int_a^b \partial_x K(t,x)\,\mathrm{d}t \right] \cdot h\,.$$

Autrement dit, on peut écrire

$$Df(x) = \int_a^b \partial_x K(t,x)\,\mathrm{d}t\,.$$

Corollaire 5.33 Soit $f : I \to E$ un chemin C^∞, dans un espace de Banach E. Soit $t_0 \in I$ et $h \in \mathbb{R}$ tel que $t_0 + h \in I$. Alors il existe une fonction $g \in C^\infty$ telle que

$$f(t_0 + h) = f(t_0) + \sum_{j=1}^{k} \frac{h^j}{j!} \cdot f^{(j)}(t_0) + h^{k+1} \cdot g(h)\,.$$

On utilise souvent ce corollaire sous la forme suivante : si f est C^∞ au voisinage de 0 et f ainsi que ses k première dérivées sont nulles en 0, alors

$$f(h) = h^{k+1} \cdot g(h)$$

où g est une fonction C^∞. On appelle ça un « lemme de division », car il indique qu'on peut diviser f par h^{k+1} en restant dans la classe des fonctions C^∞.

Dérivée d'une suite de fonctions. — Il existe une autre preuve du théorème 5.31, un peu plus détournée, mais intéressante : l'intégrale étant obtenue comme une limite de sommes de Riemann, dériver sous un intégrale revient à intervertir limite et intégrale. Pour cela, on va utiliser ce résultat, d'intérêt indépendant (et dont la preuve, ironiquement, utilise aussi la formule de la moyenne [2])

Théorème 5.34 Soient U un ouvert *convexe* d'un espace vectoriel normé E, soit F un espace *de Banach*, et soit $(f_n,\ n \in \mathbb{N})$ une suite d'applications différentiables de U dans F, vérifiant les deux propriétés suivantes :

1. il existe un point $x_0 \in U$ tel que la suite $(f_n(x_0))$ converge (lorsque $n \to \infty$) ;
2. la suite (Df_n) converge uniformément sur U vers une application $g : U \to \mathscr{L}(E,F)$.

Alors, pour tout $x \in U$, la suite $(f_n(x))$ converge vers un élément, noté $f(x)$, de F, la convergence étant uniforme sur toute partie bornée de U. De plus, l'application $f : U \to F$ ainsi obtenue est différentiable sur U, et sa différentielle n'est autre que g.

Si en outre les f_n sont C^1, alors f est C^1.

Démonstration. Traitons ici le cas $f_n \in C^1$ (pour le cas où f_n est simplement différentiable, voir [4, théorème 5.8]).

On a

$$f_n(x) = f_n(x_0) + \int_0^1 Df_n(x_0 + t(x - x_0)) \cdot (x - x_0)\,\mathrm{d}t\,.$$

Le terme $f_n(x_0)$ converge par hypothèse. Notons $f(x_0)$ la limite.

La suite de fonctions $(t \to Df_n(x_0 + t(x - x_0)) \cdot (x - x_0))$ converge uniformément sur $[0,1]$ vers $g(x_0 + t(x - x_0)) \cdot (x - x_0)$. Donc, par continuité de $h \mapsto \int_0^1 h$ pour la norme uniforme (proposition 5.15), le terme intégral converge vers $\int_0^1 g(x_0 + t(x - x_0)) \cdot (x - x_0)\,\mathrm{d}t$.

Donc $f_n(x)$ converge vers $f(x) := f(x_0) + \int_0^1 g(x_0 + t(x - x_0)) \cdot (x - x_0)\,\mathrm{d}t$.

Puisque $f_n(x)$ converge, le même raisonnement en remplaçant x_0 par x donne, pour tous (x, y),

$$f(y) = f(x) + \int_0^1 g(x + t(y - x)) \cdot (y - x)\,\mathrm{d}t\,.$$

En notant $y = x + h$ on trouve

$$f(x+h) - f(x) = \int_0^1 g(x + th) \cdot h\,\mathrm{d}t\,.$$

On sait que g est continue en x car Df_n est continue. Donc pour $\varepsilon > 0$ donné, il existe $\alpha > 0$ tel que si $|h| < \alpha$ et $t \in [0,1]$, $\|g(x+th) - g(x)\| < \varepsilon$. Du coup

$$\|f(x+h) - f(x) - g(x) \cdot h\| < \varepsilon\,\|h\|\,,$$

autrement dit $f(x+h) - f(x) - g(x) \cdot h = o(h)$, donc f est bien différentiable de différentielle g. Ici, puisque g est continue (limite uniforme de fonctions continues), on voit que f est C^1. ∎

2. Non, il ne s'agit pas d'un argument circulaire !

Deuxième preuve du théorème 5.31. Écrivons la définition de l'intégrale par les sommes de Riemann : $f(x) = \lim_{n\to\infty} f_n(x)$ avec $f_n(x) := S_n(K(\cdot,x))$. On a $Df_n(x) = S_n(\partial_x K(\cdot,x))$. Puisque $\partial_x K$ est continue, cette somme de Riemann est uniformément convergente dans un voisinage de x (proposition 5.15). On applique le théorème 5.34, ce qui donne que f est différentiable (et même C^1) avec $Df(x) = \lim Df_n(x) = \int_a^b \partial_x K(t,x)\,\mathrm{d}t$. ■

6. Difféomorphismes et inversion locale

6.1 Difféomorphismes

L'isomorphisme naturel dans la catégorie des espaces topologiques est l'*homéomorphisme* ; on le particularise aux fonctions différentiables de façon assez évidente :

Définition 6.1 Soient E et F deux espaces vectoriels normés, U un ouvert de E et f une application de U dans F. On dit que f est un **difféomorphisme** de U sur un ouvert V de F, si f est différentiable sur U, est une bijection de U sur V, et si l'application réciproque $f^{-1} : V \to E$ est différentiable sur V. On dit que f est un **difféomorphisme de classe C^1** si f est un difféomorphisme, et si f et f^{-1} sont de classe C^1.

R **Remarque 6.2** On dit couramment « difféo » au lieu de difféomorphisme, c'est plus rapide ;)

R **Remarque 6.3** Un difféomorphisme de U sur V est évidemment un *homéomorphisme* de U sur V. Mais un homéomorphisme différentiable d'un ouvert U de E sur un ouvert V de F n'est pas nécessairement un difféomorphisme. Ainsi par exemple, l'application de $\mathbb{R}$ sur lui-même $x \mapsto f(x) = x^3$, qui est différentiable, est un homéomorphisme. Cependant, l'application réciproque $y \mapsto f^{-1}(y) = y^{1/3}$ n'est pas différentiable à l'origine.

■ **Exemple 6.4** Si $A \in \mathscr{L}(E,F)$ est inversible (càd bijective et $A^{-1} \in \mathscr{L}(F,E)$) et $y_0 \in F$, alors l'application affine $x \to y_0 + Ax$ est un difféomorphisme de E dans F. ■

Exercice 6.5 Montrer que le disque unité ouvert D dans un espace euclidien E est difféomorphe à E entier. (C'est-à-dire qu'il existe un difféomorphisme de D dans E.)
Indication : commencer par le cas $E = \mathbb{R}$. ■

R **Remarque 6.6** Attention, l'image d'un ouvert par une application différentiable n'est pas forcément un ouvert, *cf.* x^2.

Proposition 6.7 Si f est un difféomorphisme $U \to V$, alors sa différentielle est inversible

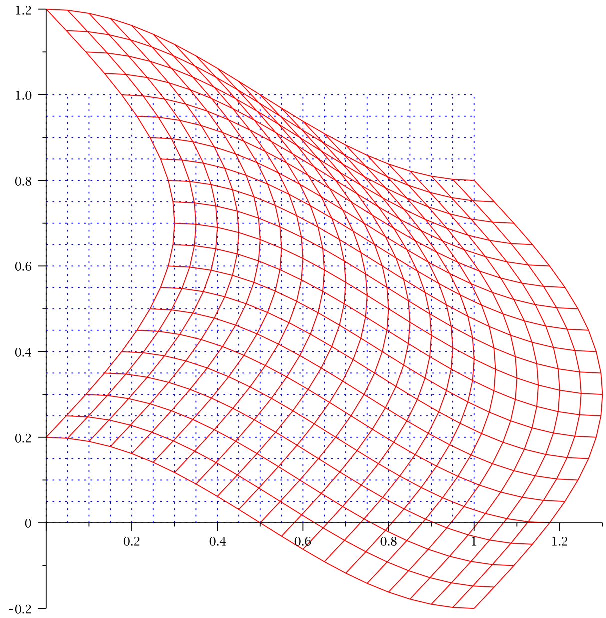

FIGURE 6.1 – Image du carré $[0,1]^2$ par un difféomorphisme C^∞.

et la différentielle de f^{-1} est

$$D(f^{-1})(f(x)) = Df(x)^{-1}$$

Démonstration. On écrit $f^{-1}(f(x)) = x$ et on différencie en x :

$$D(f^{-1})(f(x)) \cdot Df(x) = \mathrm{Id}\,.$$

Donc $Df(x)$ est inversible à gauche (=injective).

De même on écrit $f(f^{-1}(y)) = y$ et on différencie en $y = f(x)$:

$$Df(x) \cdot D(f^{-1})(y) = \mathrm{Id}\,.$$

Donc $Df(x)$ est inversible à droite (surjective). Donc $Df(x)$ est bijective. Et on a nécessairement

$$D(f^{-1})(f(x)) = Df(x)^{-1}\,.$$

■

R **Remarque 6.8** Si f est un difféo, on a supposé f^{-1} différentiable, donc $D(f^{-1})(f(x))$ est une application linéaire continue. Donc $Df(x)^{-1}$ est une ALC aussi. Ce n'est pas automatique *a priori*, dans le sens où si f est seulement différentiable et bijective (et si E, F ne sont pas des Banach), alors $Df(x)$ est une ALC mais pas forcément inversible, et même si on suppose que $Df(x)$ est inversible, son inverse $Df(x)^{-1}$ est une application linéaire pas forcément continue.

Théorème 6.9 — Rappel TOPG. Dans un Banach E, si $A \in \mathscr{L}(E) = \mathscr{L}(E,E)$ est de norme subordonnée < 1, alors $\mathrm{Id} - A$ est inversible (c'est-à-dire admet un inverse dans $\mathscr{L}(E)$). On en déduit que l'ensemble des **isomorphismes** (applications linéaires continues **inversibles**) forme un ouvert de $\mathscr{L}(E)$.

Démonstration. Montrer que l'inverse de $\mathrm{Id} - A$ est $\sum_{k\geqslant 0} A^k$ (série ACV donc convergente dans le Banach E). Cette série est appelée **série de Neumann**. ■

Proposition 6.10 Si E,F sont des Banach, l'application $f : A \to A^{-1}$ de $\mathscr{L}(E,F)$ dans $\mathscr{L}(F,E)$ est continue en tout A inversible.

Démonstration. Voir aussi TOPG. On fixe $A_0 \in \mathscr{L}(E,F)$ inversible. (Donc A_0^{-1} est dans $\mathscr{L}(F,E)$.) Il suffit alors de considérer l'application $g : B \to B^{-1}$ sur un voisinage de Id dans $\mathscr{L}(E)$. En effet si g est continue, alors $f(A) = g(A_0^{-1}A)A_0^{-1}$ sera continue comme composée d'applications continues.

Montrons donc que g est continue au voisinage de Id. On a $g(\mathrm{Id} - C) = (\mathrm{Id} - C)^{-1} = \sum_{k\geqslant 0} C^k$ pour $\|C\| < 1$. Chaque $C \to C^k$ est continue sur $\mathscr{L}(E)$ (rappel : on sait que $\mathscr{C} : (A,B) \to AB$ est continue, voir la proposition 3.14). Puisque $\left\|C^k\right\| \leqslant \|C\|^k$, la série est normalement convergente sur $B(0,r)$ pour $r < 1$, donc uniformément convergente. La somme est donc continue. ■

R **Remarque 6.11** En utilisant le théorème de différentiation sous le signe somme (utiliser le théorème 5.34 avec une série), on voit que $A \to A^{-1}$ est C^1 (et même C^∞).

Corollaire 6.12 Si f est un difféo entre ouverts d'espaces de Banach et f est C^1, alors f^{-1} est C^1 (c'est donc ce qu'on a appellé un « difféomorphisme C^1 »).

Démonstration. On applique la formule $D(f^{-1})(y) = Df(f^{-1}(y))^{-1}$ (proposition 6.7) et le corollaire, pour obtenir la continuité de $y \to D(f^{-1})(y)$ comme composée d'applications continue. ■

Proposition 6.13 — Composée de difféomorphismes. Si $f : U \to V$ et $g : V \to W$ sont des difféos, alors $g \circ f : U \to W$ est un difféo.

Démonstration. La composée d'homéos et un homéo et la composée d'applications différentiables est différentiable... ■

6.2 Énoncé du théorème d'inversion locale

Le théorème d'inversion locale est peut-être le résultat le plus important de ce cours ! Il nécessite la notion (un peu subtile) de difféomorphisme *local*.

Définition 6.14 On dit que $f : U \to F$ est un **difféo local** en $a \in U$ s'il existe un ouvert U' contenant a et un ouvert $V \subset F$ tel que la restriction de f à U' soit un difféo $U' \to V$.

Autrement dit, un difféo local est une application qui devient un difféomorphisme *si on réduit convenablement l'ensemble de départ **et** l'ensemble d'arrivée*. Par exemple, la fonction $t \mapsto \sin t$ est un difféo local en 0.

Le fameux *théorème d'inversion locale* dit que, pour être un difféo C^1 local en a, il suffit que la différentielle en a soit inversible (et on a vu plus haut que c'est nécessaire.)

Théorème 6.15 — inversion locale. Soit $f : U \to F$ une application de classe C^1 d'un ouvert U d'un espace de Banach E dans un espace de Banach F. On suppose qu'en un point $a \in U$, la différentielle $Df(a)$ est un isomorphisme de E sur F. Alors f est un C^1-difféo local en a.

Remarque 6.16 Évidemment, ça ne marchera jamais si E n'est pas isomorphe à F !

Remarque 6.17 J'ai énoncé le théorème dans des EVN, mais on peut (évidemment) prendre $U \subset X$ et $F : U \to Y$, espaces *affines* associés à des Banach.

Ce théorème repose sur un résultat de **point fixe**.

6.3 Points fixes et applications contractantes

Définition 6.18 Une application $F : X \to X$ admet x comme **point fixe** lorsque $F(x) = x$.

(Penser au « système dynamique » $x_{n+1} = F(x_n)$.)

Définition 6.19 Une application **contractante** (ou « contraction stricte ») sur un espace métrique X est une application $F : X \to X$ qui est lipschitzienne de rapport $\operatorname{Lip} F < 1$.

Le théorème suivant s'appelle aussi le « lemme de contraction ». Voir TOPG.

Théorème 6.20 — Théorème du point fixe de Picard. Soient X un espace métrique *complet* et $F : X \to X$ une application *contractante*. Alors F a un unique point fixe $a \in X$, et pour tout $x \in X$, la suite $F^n(x)$ converge vers a quand $n \to \infty$.

Le théorème du point fixe peut servir à inverser des applications ! En effet, on remarque que, étant donné un point y, trouver un point fixe de l'application $x \to y + F(x)$ revient à trouver un x tel que

$$y = x - F(x) = (\mathrm{Id} - F)(x),$$

et donc de montrer que $\mathrm{Id} - F$ est injective. Si en outre ce point fixe x est unique, alors $\mathrm{Id} - F$ est forcément bijective.

Remarquons que $x \to y + F(x)$ est tout aussi contractante que F. On va ainsi obtenir :

Théorème 6.21 — Inversion globale lipschitzienne. Pour toute F contractante (pas forcément linéaire !) sur un espace de Banach E, l'application $\mathrm{Id} - F$ est bijective et son inverse est lipschitzienne. Plus précisément on a :

$$\operatorname{Lip}((\mathrm{Id} - F)^{-1}) \leqslant (1 - \operatorname{Lip} F)^{-1}$$

Démonstration. puisque F est contractante, en posant

$$F_y(x) = y + F(x),$$

on voit que $\forall y \in E$, on a les mêmes constantes de Lipschitz : $\operatorname{Lip} F_y = \operatorname{Lip} F$. Par la remarque précédente, on peut appliquer le théorème du point fixe, ce qui donne la bijectivité de $\mathrm{Id} - F$, et la formule de la limite.

Si $y = (\mathrm{Id} - F)(x)$ et $y' = (\mathrm{Id} - F)(x')$ on a donc $x = y + F(x)$ et de la même façon, $x' = y' + F(x')$.

Donc $|x - x'| = |y - y' + F(x) - F(x')| \leqslant |y - y'| + Lip(F)\,|x - x'|$, ce qui donne $|x - x'| \leqslant (1 - Lip(F))^{-1}\,|y - y'|$. ■

6.4 Preuve du théorème d'inversion locale

Étape 1. Commençons par simplifier : on va voir qu'on peut supposer $a = 0$, $f(a) = 0$ et $Df(a) = \mathrm{Id}$ (et donc $F = E$). En effet, $T_a : x \mapsto x + a$, la translation de vecteur a, est un difféo, donc il suffit de montrer que $f \circ T_a$ est un difféo g ; dans ce cas, $f = g \circ T_{-a}$ sera un difféo par composition. Cette manip permet de remplacer a par 0.

De même, on compose à gauche par $T_{-f(a)}$: il suffit de montrer que $T_{-f(a)} \circ f$ est un difféo. Cette manip permet de remplacer $f(a)$ par 0.

Enfin, on compose par $x \mapsto Ax$, où $A = (Df(a))^{-1} : F \to E$. Il suffit de montrer que $A \circ f$ est un difféo, d'un voisinage de 0 dans E dans un voisinage de 0 dans E. Cette manip permet de remplacer $Df(a)$ par Id.

Étape 2. Montrons que f est un *homéo local*. C'est l'étape clef, et c'est un peu subtil. Par l'étape 1, on peut supposer que U est un voisinage de 0, et, par différentiabilité en 0, puisque $Df(0) = \mathrm{Id}$,

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

Posons $F(x) := x - f(x)$. Du coup $f = \mathrm{Id} - F$, et on va essayer d'appliquer l'idée du théorème précédent 6.21, à la différence qu'on va devoir ici se limiter à une petite boule, à cause du terme $o(x)$.

Fixons $\rho > 0$ (à déterminer plus tard), fixons un y tel que $\|y\| < \rho$, et considérons l'application

$$F_y : x \mapsto y + F(x)\,.$$

Est-elle contractante ? $F_y(x) - F_y(x') = F(x) - F(x')$, et par les accroissements finis,

$$\left\|F(x) - F(x')\right\| \leqslant \sup_{[x,x']} \|DF\| \left\|x - x'\right\|\,. \tag{6.1}$$

Comme $DF = \mathrm{Id} - Df$ et $Df(0) = \mathrm{Id}$, par continuité de Df, et donc de DF, pour tout $\varepsilon > 0$, il existe $\alpha > 0$ tel que dès que $\|x\| \leqslant \alpha$, on a $\|DF(x)\| \leqslant \varepsilon$. Il découle donc de (6.1) que F (et, par le même calcul, F_y) est ε-lipschitzienne sur la boule fermée $B_f(0,\alpha)$.

Quelle est l'image de F_y ? On a $\left\|F_y(x)\right\| \leqslant \|y\| + \|F(x)\| \leqslant \|y\| + \varepsilon\,\|x\|$ (car $F(0) = 0$). Donc $\leqslant \rho + \varepsilon\alpha$. Donc F_y envoie la boule $B_f(0,\alpha)$ dans la boule $B_f(0,\rho+\varepsilon\alpha)$.[1] Prenons par exemple $\varepsilon = 1/2$ (ce qui détermine α), puis $\rho = \alpha/2$. On obtient que $F_y : B_f(0,\alpha) \to B_f(0,\alpha)$ et est $\frac{1}{2}$-contractante sur cette boule.

Cette boule étant fermée dans E Banach, elle est un espace métrique complet. On peut appliquer le point fixe de Picard : dans cette boule, F_y admet un unique point fixe x : écrivons $F_y(x) = x$; comme précédemment, on obtient $y = x - F(x) = f(x)$: tous les y

1. On aurait pu également obtenir ceci en disant que, puisque F_y est ε-lipschitzienne, elle envoie $B_f(0,\alpha)$ dans $F_y(0) + \varepsilon B_f(0,\alpha) = \{y\} + B_f(0,\varepsilon\alpha) \subset B_f(0,\rho+\varepsilon\alpha)$.

dans $B(0,\rho)$ admettent un unique antécédent par f dans $B_f(0,\alpha)$. Bref, par restriction, f réalise une bijection $\hat{f}$ de

$$X := f^{-1}(B(0,\rho)) \cap B_f(0,\alpha)$$

sur $B(0,\rho)$.

Pour y, y' dans $B(0,\rho)$ et x, x' leurs antécédents respectifs par $\hat{f}$, on a, comme au théorème précédent :

$$\|x-x'\| \leqslant \|y-y'\| + \varepsilon \|x-x'\| ,$$

donc $\|x-x'\| \leqslant (1-\varepsilon)^{-1} \|y-y'\|$, donc l'inverse $\hat{f}^{-1} : y \mapsto x$ est Lipschitz donc continue. Donc $\hat{f}$ est un homéo de X dans $B(0,\rho)$.

Il reste un point à éclaircir : X, qui est ouvert dans $B_f(0,\alpha)$, est-il ouvert dans E ? Il y a deux façons de gérer ceci : la première est de montrer que, effectivement, X est ouvert dans E. En effet, avec $\varepsilon = \frac{1}{2}$, on obtient que $\hat{f}^{-1}$ est 2-lipschitzienne. Ainsi, l'image de $B(0,\rho)$ par $\hat{f}^{-1}$ (qui doit être égale à X) est en fait contenue dans $B(0,2\rho) = B(0,\alpha)$. Donc, en réalité, on a $X = f^{-1}(B(0,\rho)) \cap B(0,\alpha)$, qui est ouvert (intersection de deux ouverts). On peut donc choisir $U' = X$, $V = B(0,\rho)$.

L'autre réponse est de dire qu'en fait, ce n'est pas important ! En effet, il nous suffit de choisir U' comme l'intersection de X avec n'importe quelle boule ouverte de rayon $\leqslant \alpha$ centrée en 0, par exemple $U' := f^{-1}(B(0,\rho)) \cap B(0,\alpha) \subset X$. C'est un ouvert de E (intersection de deux ouverts), contenant 0 (puisque $f(0) = 0$), sur lequel la restriction

$$\hat{f}|_U : U \to f(U) = \hat{f}(U) \subset B(0,\rho)$$

est encore un homéo. Puisque $\hat{f}^{-1}$ est continue sur $B(0,\rho)$, $V := f(U) = (\hat{f}^{-1})^{-1}(U)$ est ouvert dans $B(0,\rho)$, et donc dans E.

Étape 3. On note f à la place de $\hat{f}|_U$ pour simplifier. Il reste à montrer que f^{-1} est différentiable sur $f(U)$.

Écrivons, pour $\|h\|$ assez petit :

$$\begin{aligned} h &= y + h - y && (6.2)\\ &= f(f^{-1}(y+h)) - f(f^{-1}(y)) && (6.3)\\ &= Df(x) \cdot (f^{-1}(y+h) - f^{-1}(y)) + o(\|f^{-1}(y+h) - f^{-1}(y)\|) . && (6.4) \end{aligned}$$

Comme f^{-1} est Lipschitz, le o est $o(h)$.

Puisque $Df(x) = \mathrm{Id} - DF(x)$ et $\|DF(x)\| \leqslant 1/2$, on voit que $Df(x) : E \to E$ est inversible (théorème 6.9). On peut donc composer par $Df(x)^{-1}$ et on obtient

$$f^{-1}(y+h) - f^{-1}(y) = Df(x)^{-1} h + o(h) ,$$

ce qui prouve que f^{-1} est différentiable en y (et de différentielle $Df(x)^{-1}$, ce qu'on savait déjà). Comme remarqué plus haut, cette différentielle est continue en $x = \hat{f}^{-1}(y)$, donc en y, donc f^{-1} est C^1.

Fin de la preuve !

6.5 Théorème des fonctions implicites

D'un point de vue technique, le théorème des fonctions implicites est une simple reformulation du théorème d'inversion locale 6.15 ; cependant, il est plus naturel à utiliser dans la situation (assez courante) où on *cherche à résoudre une équation dépendant de façon C^1 d'un paramètre*. Voici son énoncé.

Théorème 6.22 — Fonctions implicites. Soient E, F, G des espaces de Banach. Soit $f : \Omega \to G$ une application C^1 définie sur un ouvert $\Omega \subset E \times F$. Soit $a_0 \in G$. Supposons qu'on veuille résoudre l'équation

$$f(x, y) = a_0 \tag{6.5}$$

d'inconnue y, dépendant d'un paramètre x.

1. Soient $x_0 \in E$, $y_0 \in F$ tels que $f(x_0, y_0) = a_0$. (Autrement dit, pour $x = x_0$, l'équation (6.5) est résolue avec $y = y_0$).
2. Supposons que $\partial_y f(x_0, y_0) : F \to G$ soit inversible.

Alors pour tout x proche de x_0, on peut résoudre l'équation (6.5) pour un unique $y = y(x)$ proche de y_0. Et l'application $x \mapsto y(x)$ est C^1.

Remarque 6.23 Les espaces F et G doivent donc être isomorphes.

Bien entendu, dans cet énoncé, le mot « proche » est utilisé au sens des voisinages : il existe un voisinage ouvert U de x_0 et un voisinage ouvert V de y_0, et une fonction $g : U \to V$ de classe C^1 telle que pour tout $x \in U$, $y = g(x) \in V$ est l'unique solution dans V de l'équation (6.5), c'est-à-dire telle que

$$f(x, g(x)) = a_0 .$$

En outre, on peut également varier a, et on a encore une unique solution locale $(x, a) \mapsto y(x, a)$, qui est également C^1 :

Théorème 6.24 — Fonctions implicites — suite. Sous les mêmes hypothèses, pour tout x proche de x_0, et tout a proche de a_0, on peut résoudre

$$f(x,\ y(x, a)) = a ,$$

pour une unique fonction $(x, a) \mapsto y(x, a)$ à valeurs dans un voisinage de y_0, et cette fonction est de classe C^1.

Remarquez que ce dernier énoncé découle directement du premier (théorème 6.22), en considérant la fonction $\tilde{f}((x, a), y) = f(x, y) - a$ avec le nouveau « paramètre » $\tilde{x} = (x, a)$. Cependant, pour ce qui est de la preuve du théorème 6.22, il est aussi simple d'incorporer dès le début la dépendance en a (et bien sûr, l'énoncé pour a_0 fixé en découle immédiatement).

Preuve directe du théorème 6.24. L'idée est d'appliquer le *théorème d'inversion locale* à l'application

$$F(x, y) := (x, f(x, y)) : U \to E \times G , \qquad F(x_0, y_0) = (x_0, a_0) . \tag{6.6}$$

Voyons les détails.

1. Sous forme « matricielle » (par blocs dans $E\times F$, comme une matrice jacobienne) on voit bien que $DF(x_0,y_0)$ est inversible, car
$$DF(x_0,y_0)=\begin{pmatrix}\mathrm{Id} & 0\\ \partial_x f(x_0,y_0) & \partial_y f(x_0,y_0)\end{pmatrix}\in\mathscr{L}(E\times F,\ E\times G),$$
et $\partial_y f(x_0,y_0)$ est inversible par hypothèse. On peut aussi donner explicitement l'inverse : Si $DF(x_0,y_0)\cdot(h,k)=(u,v)$, alors l'écriture composante par composante donne :
$$\begin{aligned}&h=u \quad \text{et}\\ &\partial_x f(x_0,y_0)\cdot h+\partial_y f(x_0,y_0)\cdot k=v,\end{aligned}$$
donc $k=[\partial_y f(x_0,y_0)]^{-1}\cdot(v-\partial_x f(x_0,y_0)\cdot h)$.
2. Le théorème d'inversion locale fournit un inverse G de classe C^1 :
$$F\circ G(x,a)=(x,a) \quad \text{dans un voisinage de } (x_0,a_0)\,.$$
De la formule de F (6.6), on déduit que l'inverse est forcément de la forme
$$G(x,a)=(x,\ g(x,a))\,,$$
pour une fonction g de classe C^1.
3. La deuxième composante de l'équation $F(G(x,a))=(x,a)$ s'écrit $f(x,g(x,a))=a$. L'unique solution cherchée est donc $y(x)=g(x,a)$.

■

R **Remarque 6.25** *Pour bien se souvenir* de la bonne hypothèse dans le théorème des fonctions implicites (est-ce $\partial_x f$ ou $\partial_y f$ qui doit être inversible ?) pensez à remplacer f par une fonction linéaire : $f(x,y)=Ax+By$. On a l'équation $Ax+By=a$: quelle est la condition pour « sortir y », càd exprimer y en fonction de x ? Clairement, il suffit que B soit inversible. L'hypothèse du théorème est donc que $\partial_y f$ soit inversible.

7. Différentielles d'ordre supérieur

7.1 Applications deux fois différentiables

Il est facile de comprendre pourquoi, sur un espace vectoriel (ou affine) normé général, la notion de dérivées d'ordre supérieur n'est pas aussi simple que dans le cas d'une variable réelle. En effet, soient U un ouvert de E, $a \in U$ et $f : U \to F$ une application différentiable sur un voisinage V de a.

On rappelle que la différentielle Df est une application de V dans $\mathscr{L}(E,F)$, qui est lui-même un evn. On pourra donc dire (et on n'a pas le choix !) que f est **deux fois différentiable** en $a \in U$ si Df est différentiable en a.

Mais, dans ce cas, $D(Df)(x) \in \mathscr{L}(E,\mathscr{L}(E,F))$. Quel est cet espace compliqué ?

Lemme 7.1 $\mathscr{L}(E,\mathscr{L}(E,F))$ est isomorphe (par une isométrie linéaire) à l'espace $\mathscr{L}^2(E,F)$ des applications *bilinéaires continues* de $E \times E$ dans F.

Démonstration. Si $A \in \mathscr{L}(E,\mathscr{L}(E,F))$, on définit l'application bilinéaire $a : E \times E \to F$ par

$$a(u,v) = A(u)(v)$$

Réciproquement, si a bilinéaire continue est donnée, on définit $A(u) = (v \mapsto a(u,v))$. On obtient ainsi une relation $A \mapsto a$ et $a \mapsto A$, l'une étant clairement l'inverse de l'autre.

Montrons que cette relation est une isométrie : on a (par définition, *cf.* proposition 3.13)

$$\|a\| = \sup_{(u,v)} \frac{\|a(u,v)\|}{\|u\|\,\|v\|}$$

donc $\|a\| = \sup_{(u,v)} \|A(u)(v)\| / \|u\|\,\|v\|$: c'est la plus petit constante c telle qu'on ait :

$$\|A(u)(v)\| \leqslant c\,\|u\|\,\|v\| \qquad \text{pour tous } (u,v)\,. \tag{7.1}$$

D'autre part, $\|A\|$ est la plus petite constante C telle que pour tous u, $\|Au\| \leqslant C\,\|u\|$. Pour tout v, $\|A(u)(v)\| \leqslant \|Au\|\,\|v\| \leqslant C\,\|u\|\,\|v\|$, donc $c \leqslant C$.

Mais par (7.1), en fixant u, on voit que $\|Au\| \leqslant c\,\|u\|$, donc $C \leqslant c$. ∎

Définition 7.1 Si f est deux fois différentiable en a, on appelle **différentielle seconde** de f en a l'application bilinéaire continue $D^2 f(a) = D(Df)(a)$ donnée par l'isomorphisme ci-dessus : càd :

$$D^2 f(a)(u,v) = (D(Df)(a) \cdot u)(v)\,.$$

Il se trouve que ce n'est pas n'importe quelle application bilinéaire : elle est toujours *symétrique* :

Théorème 7.2 — Lemme de Schwarz. Si f est deux fois différentiable au point a, sa dérivée seconde $D^2 f(a)$ appartient à l'espace $\mathscr{L}^2_S(E,F)$ des *applications bilinéaires continues et symétriques* de $E \times E$ dans F :

$$D^2 f(a)(u,v) = D^2 f(a)(v,u) \qquad \forall u, v \in E\,.$$

Démonstration. Elle est un peu technique...

Montrons-le d'abord pour des fonctions d'un voisinage de $(0,0)$ dans $\mathbb{R}^2$, à valeurs dans F.

Lemme 7.2 Si $g(x,y)$ est deux fois dérivable en $(0,0) \in \mathbb{R}^2$ alors, quand $h \to 0$,

$$g(h,h) - g(h,0) - g(0,h) + g(0,0) = h^2 \partial_x \partial_y g(0,0) + o(h^2)$$

Si on prouve ce lemme, alors en l'appliquant à la fonction $G(x,y) = g(y,x)$, on voit que la même égalité est vraie avec $\partial_y \partial_x g(0,0)$, ce qui montre que

$$\partial_x \partial_y g(0,0) = \partial_y \partial_x g(0,0) + o(1)$$

et donc

$$\partial_x \partial_y g(0,0) = \partial_y \partial_x g(0,0)\,.$$

Pour revenir au théorème, il ne reste plus qu'à appliquer ça à la fonction

$$g(x,y) := f(a + x \cdot u + y \cdot v).$$

(Avec u et v quelconques dans E, (x,y) proche de 0 dans $\mathbb{R}^2$). On a (pour $a + x \cdot u \in V$)

$$\partial_y g(x,0) = Df(a + x \cdot u) \cdot v\,;$$

c'est donc différentiable par rapport à x en $x = 0$, et

$$\partial_x \partial_y g(0,0) = [D(Df)(a) \cdot u] \cdot v = D^2 f(a)(u,v)\,.$$

De même,

$$\partial_y \partial_x g(0,0) = [D(Df)(a) \cdot v] \cdot u = D^2 f(a)(v,u)\,,$$

ce qui montre le théorème. ■

Preuve du lemme 7.2. Posons

$$F(x,y) := g(x,y) - g(x,0) - g(0,y) + g(0,0) - y(\partial_y g(x,0) - \partial_y g(0,0)).$$

On a $F(x,0) = 0$ et

$$\partial_y F(x,y) = \partial_y g(x,y) - \partial_y g(0,y) - \partial_y g(x,0) + \partial_y g(0,0). \tag{7.2}$$

On voit que seule la fonction $\partial_y g$ intervient. Cette fonction est différentiable en $(0,0)$, et donc $\forall (x,y)$ on a

$$\partial_y g(x,y) = \partial_y g(0,0) + D(\partial_y g)(0,0) \cdot (x,y) + o(x,y) \quad \text{quand } (x,y) \to (0,0).$$

On applique ça aux 4 termes de (7.2) et on obtient donc

$$\begin{aligned}\partial_y F(x,y) &= [1-1-1+1]\partial_y g(0,0) \\ &\quad + D(\partial_y g)(0,0) \cdot [(x,y) - (0,y) - (x,0) - (0,0)] \\ &\quad + o(x,y) \\ &= 0 + 0 + o(x,y).\end{aligned}$$

Donc par les accroissements finis, appliqués à la fonction différentiable $y \mapsto F(x,y)$, on a, au voisinage de $(0,0)$,

$$\|F(x,y)\| = \|F(x,y) - F(x,0)\| \leqslant \sup \|\partial_y F\| \, |y| = \|o(x,y)\| \, |y| .$$

Donc $F(h,h) = o(h^2)$.

En remarquant que

$$\partial_y g(x,0) - \partial_y g(0,0) = x\partial_x \partial_y g(0,0) + o(x)$$

et en remplaçant le terme correspondant dans F, on obtient bien le lemme. ■

7.2 Applications de classe C^k

On rappelle que f est deux fois différentiable en a si f est différentiable *au voisinage de* a et Df est différentiable *en* a.

On peut ensuite se poser la question : est-ce que Df est *deux fois* différentiable en a, au sens du paragraphe précédent ? Autrement dit, est-ce que Df est différentiable au voisinage de a et que $D(Df)$ est différentiable en a ?

On dira alors que f est 3 fois différentiable en a.

Ainsi, par récurrence :

Définition 7.3 On dit que f est ***k* fois différentiable en** a si f est différentiable *au voisinage de* a et que Df est $(k-1)$ fois différentiable en a.

Dans ce cas, on définit ensuite naturellement les différentielles (ou dérivées) successives :

Définition 7.4

1. $D^0 f = f$;
2. $D^j f = D(D^{j-1} f)$, pour $0 < j < k$;
3. $D^k f(a) = D(D^{k-1} f)(a)$.

On dira bien sûr que f est k fois différentiable si elle l'est en tout point de U.

Définition 7.5 On dit que f est C^k si elle est k fois différentiable et toutes ses dérivées $D^j f$, $j =, \ldots, k$ sont continues. On dit que f est C^∞ si elle est C^k pour tout k.

Quid des espaces où vivent ces différentielles multiples ? En appliquant par récurrence la symétrie de la dérivée seconde, on obtient (voir par exemple le théorème 2.3 dans [4])

Théorème 7.6 Si $f : U \subset E \to F$ est n fois différentiable en $a \in U$, alors $D^n f(a)$ est une application **n-linéaire continue symétrique** de E^n dans F.

Ici, l'adjectif *symétrique* veut dire

$$D^n f(a)(u_1, \ldots, u_n) = D^n f(a)(u_{\sigma(1)}, \ldots, u_{\sigma(n)})$$

pour toute *permutation* σ de $\{1, \ldots, n\}$.

Notation 7.7 On notera $D^k f(a) \cdot (h^k)$ pour $D^k f(a) \cdot (h, \ldots, h)$.

Exemple 7.8 Une application linéaire continue $f(x) = Ax$ est C^∞, et on a :
1. $D^1 f(x) = A \quad (\forall x)$;
2. $D^k f = 0$ pour $k \geqslant 2$.

■

Proposition 7.9 Une application *bilinéaire continue* B est toujours C^∞ et ses dérivées sont :
1. $DB(x,y)(u,v) = B(u,y) + B(x,v)$;
2. $D^2 B(x,y)[(u_1,v_1),(u_2,v_2)] = B(u_1,v_2) + B(u_2,v_1)$;
3. $D^k B = 0$ pour $k > 2$.

Démonstration. Exercice ■

Proposition 7.10
1. Si f est à valeurs dans un espace produit $F_1 \times \cdots \times F_p$, alors f est C^k si et seulement si toutes ses composantes f_j sont C^k. Dans ce cas, les composantes de $D^m f(a)$ sont les $D^m f_j(a)$, pour $m = 0, \ldots, k$.
2. La composée de deux applications C^k est C^k.
3. Si f est définie sur un espace produit $E_1 \times \cdots \times E_n$, alors f est C^{k+1} si et seulement si ses dérivées partielles sont C^k.

Démonstration.
1. Exercice.
2. Utiliser la proposition précédente, car

$$D(f \circ g)(x) = Df(g(x)) \circ Dg(x)$$

et utiliser l'application bilinéaire continue « composition linéaire » $\mathscr{C} : (A,B) \to A \circ B$ (proposition 3.14), puis procéder par récurrence.

3. Le cas $k = 0$ est traité au théorème 4.31. Pour $k > 0$, on sait alors (pour les 2 implications) que f est différentiable, on peut donc appliquer la formule

$$Df(a)(u_1, \ldots, u_n) = \sum_{j=1}^{n} \partial_j f(a) \cdot u_j \, .$$

■

Théorème 7.11 — Formule de Taylor avec reste intégral sur un segment. Soit $f : U \subset E \to F$ avec F Banach. On suppose f de classe C^n. Soit $h \in E$ tel que le segment $[a, a+h]$ soit entièrement contenu dans U. Alors

$$f(a+h) = f(a) + \sum_{k=1}^{n} \frac{1}{k} D^k f(a) \cdot (h^k) + \int_0^1 \frac{(1-t)^n}{n!} D^{n+1} f(a+th) \cdot (h^{n+1}) dt \, .$$

Démonstration. Se ramener à la formule à une variable t (théorème 5.29) en posant $g(t) = f(a+th)$. ■

8. Sous-variétés et extrema liés

Ce chapitre est une introduction à une notion, de type géométrique, mais très utile dans toutes les branches des mathématiques, cette de *sous-variétés*. Attention, on se restreindra ici à la dimension finie. Les idées s'étendent sans trop de difficulté à la dimension infinie (dans les espaces de Banach, en faisant souvent attention à ce que les sous-espaces considérés admettent des supplémentaires fermés).

Dans ce chapitre, on supposera que toutes les applications sont C^k avec $k \geqslant 1$. (On dira juste « lisse » pour signifier « qui appartient à la classe de régularité choisie »).

8.1 Sous-variétés

Soit $E = \mathbb{R}^n$. (Ou, plus généralement, un espace affine de dimension finie.) On souhaite définir des sous-ensembles de E qui soient « lisses » (c'est-à-dire C^k). On les appellera des *sous-variétés* de E. Qu'est-ce que ça peut bien vouloir dire ?

Par exemple, un sous-espace *vectoriel* est certainement lisse. Mais ce n'est pas le seul cas ! Ainsi, on a également envie de dire qu'un *cercle* dans le plan $\mathbb{R}^2$, doit être une partie lisse.

Une définition tentante serait de dire qu'un ensemble de $\mathbb{R}^2$ est lisse si c'est le *graphe* d'une application lisse $y = f(x)$. Mais ça ne fonctionne pas pour le cercle : certes, un demi-cercle peut être décrit comme un graphe, mais pas un cercle complet.

Une autre approche est de décrire les ensembles lisses comme les solutions d'une « bonne » équation. Le problème est donc déplacé : quelles sont les bonnes équations ? Remarquons par exemple qu'un sous espace vectoriel est toujours le noyau d'une application linéaire surjective sur un certain espace vectoriel (par exemple la projection sur l'orthogonal). Les surjections linéaires seront de bonnes équations ; mais alors, suivant la philosophie que les isomorphismes des « espaces lisses » sont les difféomorphismes, on doit pouvoir d'autoriser à « déformer » les surjections linéaires par des difféomorphismes. C'est ce qu'on appelera une *submersion* (locale). Précisément :

Définition 8.1 Une application $g : U \subset E \to F$, où F est un evn de dimension m, est une **submersion locale** en $a \in U$ si, à difféomorphismes près en a et $g(a)$, c'est une surjection linéaire : il existe $k \geqslant 0$ et $\phi : U \to \mathbb{R}^{m+k}$ difféo local en a avec $\phi(a) = 0$, et

$\psi : g(U) \to \mathbb{R}^m$ difféo local en $b = g(a)$ avec $\psi(b) = 0$, tels que

$$\psi \circ g \circ \phi^{-1}(x,y) = x \qquad x \in \mathbb{R}^m, y \in \mathbb{R}^k.$$

Examinons cette définition. D'abord, si $n = \dim E$, on a nécessairement $n = m+k$, puisque ϕ est un difféo local. L'équation $g(z) = g(a)$ d'inconnue z est (localement au voisinage de a) équivalente à $\psi(g(z)) = \psi(g(a)) = 0 \in \mathbb{R}^m$; ainsi, en posant $\phi(z) = (x,y) \in \mathbb{R}^m \times \mathbb{R}^k$, on obtient l'équation $x = 0$. Dans les variables (x,y), cette équation décrit donc (localement) le sous-espace vectoriel $\{0\} \times \mathbb{R}^k$; et dans la variable d'origine z, cet ensemble est donné par $z = \phi^{-1}(0,y)$; c'est donc un ensemble (*a priori*) *non-linéaire*, paramétré par la variable y dans un voisinage de 0 dans $\mathbb{R}^k$.

Définition 8.2 Une partie $S \subset E$ est une **sous-variété** de E si, pour chaque point $a \in S$, il existe un evn F et une submersion locale $g : U \to F$ en a telle que, pour un voisinage U' de a, $S \cap U'$ soit exactement donné par l'équation $g(x) = g(a)$, $x \in U'$.

Exercice 8.3 Soit S une sous-variété *connexe* ; montrer que l'entier k de la définition de submersion locale est constant. ■

Définition 8.4 Si S est connexe, l'entier k de la définition de submersion locale est appelé la **dimension** de S.

Heureusement, tester le fait d'être une submersion locale est très facile, grâce au théorème d'inversion locale :

Théorème 8.5 — théorème de submersion locale. Une application lisse g est une submersion locale en a si et seulement si $Dg(a)$ est surjective.

R **Remarque 8.6** Pour une application linéaire (en dimension finie), le fait d'être surjective est équivalent au fait d'être de rang maximal.

■ **Exemple 8.7** Par exemple, le cercle unité dans $\mathbb{R}^2$ est bien une sous-variété : on peut choisir la submersion $g(x,y) = x^2 + y^2$ (la même va fonctionner en tout point du cercle), car $Dg(x) \neq 0$ lorsque x est sur le cercle. Bien sûr, $Dg(0,0) = 0$, donc g n'est pas une submersion locale en 0, mais ça ne l'empêche pas d'être une submersion locale en tout point du cercle, et donc de prouver que le cercle est une sous-variété. ■

Preuve du théorème de submersion locale.

1. Si g est une submersion locale, $\psi \circ g \circ \phi^{-1}(x,y) = x$, donc $g = \psi^{-1} \circ \pi_1 \circ \phi$ où $\pi_1(x,y) = x$.
   En différenciant en a : (posons $b = g(a)$)
   $$Dg(a) = D\psi(b)^{-1} \circ \pi_1 \circ D\phi(a).$$
   L'image de $Dg(a)$ est donc celle de $D\psi(b)^{-1} \circ \pi_1$ puisque $D\phi(a)$ est inversible. C'est donc l'image de $\mathbb{R}^m = \pi_1(\mathbb{R}^n)$ par l'isomorphisme $D\psi(b)^{-1}$, c'est-à-dire F. Donc $Dg(a)$ est bien surjective.
2. Si $Dg(a)$ est surjective, soit H un supplémentaire de $\ker Dg(a)$: on a $E = H \oplus \ker Dg(a)$, et $Dg(a)$ induit un isomorphisme de H sur F (espaces de dimension m) (voir aussi l'exercice 2 de la feuille TD11). En écrivant $g(z) = g(x,y)$ dans cette décomposition, on voit que $\partial_x g : H \to F$ est un isomorphisme ; on est ainsi dans

la situation du théorème d'inversion locale (théorème 6.15, voir aussi la preuve du théorème 6.24), et donc l'application

$$\phi : (x,y) \to (g(x,y),\, y) \in \mathbb{R}^m \times \mathbb{R}^k$$

est un difféo local. On a

$$g \circ \phi^{-1}(x,y) = x$$

ce qui prouve que g est une submersion locale.

■

8.2 Espace tangent

Soit $S \subset E$ une sous-variété, et $a \in S$.

Définition 8.8 L'**espace tangent** à S en a est l'espace noté $T_aS \subset E$ égal à l'ensemble des $\gamma'(0)$ où $\gamma :]-1,1[$ est un chemin C^1 tracé dans S avec $\gamma(0) = a$.

Cette définition très géométrique paraît difficile à manipuler. Mais dès qu'on dispose d'une équation de la sous-variété (une submersion locale), c'est très facile :

Proposition 8.9 Soit g une submersion locale en a qui définit S au voisinage de a. Alors $T_aS = \ker Dg(a)$.

Démonstration. Puisque g est constant sur S, $g \circ \gamma$ est constant, donc de différentielle nulle : $Dg(a) \cdot \gamma'(0) = 0$. Donc $T_aS \subset \ker Dg(a)$.

Réciproquement, en conjuguant par les difféos de la définition, g s'écrit $(x,y) \mapsto x$, donc un vecteur dans $\ker Dg(a)$ est (modulo ces difféos) de la forme $(0,v)$, pour $v \in \mathbb{R}^k$, où k est la dimension de S en a. On choisit alors le chemin $t \to (0, t \cdot v)$. ■

Exercice 8.10 Dans la preuve de la proposition 8.9 ci-dessus, expliciter ce qu'on a voulu dire par « modulo ces difféomorphismes ». ■

8.3 Points critiques et extrema

Soit $f : U \subset E \to F$.

Définition 8.11 On dit que $a \in U$ est un **point critique** de f si $Df(a)$ n'est pas surjective.

Proposition 8.12 Supposons $F = \mathbb{R}$. Si f admet un maximum local ou un minimum local en a, alors $Df(a) = 0$.

(Bien sûr, la réciproque n'est pas vraie, cf $f(x) = x^3$, il faudrait au moins regarder le signe de la forme quadratique associée à la dérivée seconde.)

Démonstration. On a déjà vu la preuve à une variable. On va s'y ramener. Soit $u \in E$. Prenons $F(t) := f(a + t \cdot u)$. Alors F admet un extremum en $t = 0$, donc $F'(0) = 0$, donc $Df(a) \cdot u = 0$, CQFD. ■

Soit maintenant S une sous-variété de E. Supposons que x soit *lié à* S, c'est-à-dire qu'on considère uniquement la restriction de f à S. Il est clair que les extrema de f restreinte à S sont en général très différents des extremas de f dans E. Par exemple, si x est une perle contrainte à se déplacer le long d'un collier S, la hauteur maximale de la perle liée au collier n'est certainement pas la hauteur maximale de la perle non contrainte (qui est probablement infinie...)

A-t-on une condition de type différentiel (comme la proposition 8.12) nécessaire pour avoir un extremum de $f_{\upharpoonright S}$?

FIGURE 8.1 – *La Dame au collier de perles* (Vrouw met parelsnoer), partie haute, Johannes Vermeer, 1664 (Gemäldegalerie, Berlin). Les perles sont liées au collier : elles sont donc contraintes de rester sur une sous-variété de dimension 1.

Proposition 8.13 — extrema liés. Si la restriction de $f : U \to \mathbb{R}$ à une sous-variété S admet un extremum local, alors $Df(a)$ s'annule sur l'espace tangent T_aS. (Autrement dit, $T_a \subset \ker Df(a)$.) Si g est une submersion locale en a qui définit S, alors cette condition s'écrit $\ker Dg(a) \subset \ker Df(a)$.

■ **Exemple 8.14** Soit S le cercle unité (noté souvent S^1) dans $\mathbb{R}^2$. Soit $f(x,y) = x+y$. Quels sont les extrema de f restreint au cercle ?
Solution : le cercle est défini par la submersion $g(x,y) = x^2+y^2 = 1$. L'espace tangent en $a = (x,y)$ est $\ker Dg(a)$, c'est-à-dire les vecteurs (u,v) tels que $Dg(a)\cdot(u,v) = 0$: cette équation s'écrit

$$xu + yv = 0$$

ou encore : (u,v) est orthogonal à (x,y). Donc $(u,v) = \lambda\cdot(-y,x)$, pour un $\lambda \in \mathbb{R}$. Par la proposition 8.13, ce vecteur doit annuler $Df(a) = Dx + Dy$, ce qui donne $u + v = 0$. Donc $x - y = 0$. On obtient les deux points intersection de la diagonale $x = y$ avec le cercle.

Réciproquement, il reste à vérifier que ce sont bien des extrema. ■

Définition 8.15 Soit $f : U \to \mathbb{R}$, $a \in U \cap S$. Si $Df(a)$ s'annule sur l'espace tangent T_aS, on dit que a est un **point critique** de $f_{\restriction S}$.

Théorème 8.16 — Multiplicateurs de Lagrange. Soit S une sous-variété de E, et soit $g : U \subset E \to F$ une submersion locale en a, telle que $g(z) = g(a)$ soit une équation de S au voisinage de a. Soit U un ouvert contenant S et $f : U \to \mathbb{R}$.

Les deux propriétés suivantes sont équivalentes

1. Le point a est un point critique de $f_{\restriction S}$.
2. Il existe une forme linéaire $\lambda \in F'$ telle que $Df(a) = \lambda \circ Dg(a)$.

La forme linéaire λ est alors unique et s'appelle un **multiplicateur de Lagrange**.

Démonstration. C'est une caractérisation purement linéaire sur l'espace tangent T_aS.

— $2 \Rightarrow 1$ est clair puisque $T_aS = \ker Dg(a)$.

— $1 \Rightarrow 2$: λ est unique car l'image de $Dg(a)$ est F. Pour l'existence : $Df(a)$ est une forme linéaire qui s'annule sur $\ker Dg(a)$. On écrit $E = H \oplus \ker Dg(a)$, et $Dg(a)$ induit un isomorphisme A de H sur F. Posons $\lambda = Df(a) \circ A^{-1}$. On a bien $Df(a) = \lambda \circ Dg(a)$.

■

Bibliographie

[1] A. AVEZ, *Calcul différentiel* (Collection Maîtrise de Mathématiques Pures.). Masson, Paris, 1983, page 148, ISBN : 2-225-79079-5 (cf. page 25).

[2] H. CARTAN, *Calcul différentiel*. Hermann, Paris, 1967, page 178 (cf. page 9).

[3] M. CHAPERON, *Calcul différentiel et calcul intégral* (Sciences Sup), 2ème édition. Dunod, 2008 (cf. page 9).

[4] G. CHRISTOL, A. COT et C.-M. MARLE, *Calcul différentiel* (Cours et exercices corrigés). ellipses, 1997 (cf. pages 8, 56, 70).

[5] D. KAHNEMAN, *Système 1, système 2 : les deux vitesses de la pensée*. Flammarion, 2012 (cf. page 8).

[6] L. SCHWARTZ, *Analyse II — Calcul différentiel et équations différentielles* (Enseignement des sciences 43), Nouvelle édition corrigée. Hermann, 1997 (cf. page 15).

Index