

Notes cours
«Fonctions de plusieurs variables»

Bernhard Haak

Version du 13 octobre 2023

Table des matières

1	Normes et topologie	4
1.1	L'espace \mathbb{R}^d , normes, produit scalaire	4
1.2	Exemples de normes	5
1.3	La topologie de \mathbb{R}^d	6
1.3.1	Intérieur, frontière (bord), adhérence d'un ensemble	8
1.4	Limites et continuité	9
1.5	Compacité (séquentielle)	17
1.6	Équivalence de normes	18
1.7	Courbes paramétrées	22
1.8	Reparamétrisation	27
2	Fonctions de plusieurs variables	28
2.1	Continuité dans \mathbb{R}^m	28
2.2	Dérivées partielles	29
2.3	Dérivées partielles supérieures	31
2.4	La différentielle	33
2.5	Propriétés de la différentielle	36
2.5.1	Accroissements finis	40
2.6	La formule de Taylor	45
2.6.1	Le cas 1D	45
2.6.2	Le cas de plusieurs variables	46
2.6.3	Le polynôme de Taylor d'ordre 2 et la matrice Hessienne	47
2.7	Quelques résultats d'algèbre linéaire	48
2.8	Extrema locaux	51
3	Les grands théorèmes du calcul différentiel	53
3.1	Le théorème du point fixe	53
3.1.1	Application : google page rank	55
3.2	Le théorème de l'inverse locale	56
3.3	Le théorème des fonctions implicitement définis	61
3.3.1	Dérivation implicite	63
3.4	Le théorème de l'immersion locale	65
4	Applications	66
4.1	Extrema sous contrainte	66
4.2	Intégrales curvilignes	70
4.2.1	Intégrales sur une courbe	70

4.2.2	Champs vectoriels	71
4.2.3	Application : Primitives de champs de vecteurs	72
4.3	Intégrales de surface	75
4.3.1	La vérité sur les déterminants	75
4.3.2	Intégrale de fonctions scalaires sur une surface	79
4.3.3	Intégration d'un champ de vecteurs sur une surface	81
4.4	Généralisation à la dimension n	82
4.4.1	Intégration par parties sur un domaine "pavé"	83
4.4.2	Formules de Green et le théorème de la divergence	84


Avant-propos

Ce document est basé sur des **notes de cours** du semestre d'automne 2020 prises par Cécile Touzet dans le cours de Bernhard Haak. Aucune garantie d'être correct, complet, etc. Prière de communiquer toute sorte de faute (grammaire, orthographe, maths) ou passage incompréhensible à

`bernhard.haak@math.u-bordeaux.fr`

Il est important de souligner que la lecture d'un document quelconque ne peut pas remplacer le travail personnel. Une bonne façon de travailler le cours consiste à rédiger pour chaque théorème (au moins) la stratégie de la preuve en quelques phrases en prose, puis d'être capable de re-transformer ce résumé à nouveau en preuve rigoureuse.

Ce document est mis à disposition selon les termes de la Licence Creative Commons

Attribution "creative commons licence CC-BY-SA 4.0" 

Vous trouvez les détails de ces règles sur <https://creativecommons.org>

1 Normes et topologie

Nous utilisons la **notation** suivante pour des intervalles réels :

- a) $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$.
- b) $[a, b[= \{x \in \mathbb{R} : a \leq x < b\}$.
- c) $(a, b] = \{x \in \mathbb{R} : a < x \leq b\}$.
- d) $(a, b[= \{x \in \mathbb{R} : a < x < b\}$.

Si on écrit $f \stackrel{\text{def.}}{=} g$, ou $f \stackrel{\text{def.}}{=} g$, le double point : est du côté de l'objet à définir.

1.1 L'espace \mathbb{R}^d , normes, produit scalaire

La notion de distance entre deux points est un des concepts basiques de l'analyse. Rappelons par exemple la notion d'une suite réelle convergente : pour tout $\varepsilon > 0$ il existe un rang N à partir duquel (donc pour $n \geq N$), $|x_n - \ell| < \varepsilon$. Il est utile de se rendre compte que la valeur absolue ici sert à mesurer la *distance* entre x_n et la limite ℓ . La valeur absolue induit une distance sur \mathbb{R} par $d(x, y) = |x - y|$. La bonne façon de définir la convergence serait de dire que les distances entre x_n et ℓ deviennent uniformément petites à partir d'un certain rang. Voici une définition rigoureuse de distance.

Définition 1.1. Une distance sur un ensemble X est une application $d : X \times X \rightarrow \mathbb{R}$ qui vérifie

(positivité) $d(x, y) \geq 0$ pour tout $x, y \in X$, et $d(x, y) = 0$ si, et seulement si $x = y$.

(symétrie) Pour tout $x, y \in X$, on a $d(x, y) = d(y, x)$

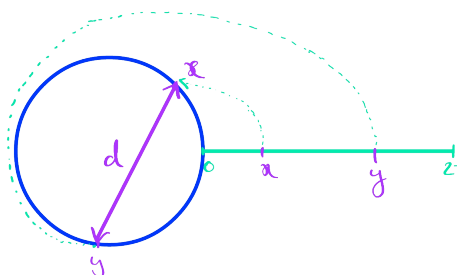
(inégalité triangulaire) $\forall x, y, z \in X : d(x, y) \leq d(x, z) + d(y, z)$

Mais on peut utiliser d'autres distances !

Sur $(0, 2\pi]$, par exemple

$$d(x, y) = |e^{ix} - e^{iy}|$$

est une autre distance (exercice!) : le segment $(0, 2\pi]$ est "enroulé" sur le cercle d'unité ; puis on mesure la distance (Euclidienne) de \mathbb{R}^2 - qui coïncide avec le module complexe. Voir image à côté.



Observons que pour cette distance la suite (x_n) défini par $x_n = \frac{1}{n}$ tend vers 2π et pas vers 0 ! L'étude d'espaces métriques se fera dans le cours de topologie ; dans ce cours nous allons travailler avec des structures plus riches, notamment avec des *espaces vectoriels* et

des normes. Rappelons que $\mathbb{R}^d = \{(x_1, \dots, x_d) \text{ tel que } x_i \in \mathbb{R}\}$ est un espace vectoriel pour la loi interne (addition)

$$(x_1, \dots, x_d) + (y_1, \dots, y_d) = (x_1 + y_1, \dots, x_d + y_d)$$

et la multiplication externe par des scalaires

$$\lambda(x_1, \dots, x_d) = (\lambda x_1, \dots, \lambda x_d)$$

Définition 1.2. Une norme sur un espace vectoriel E est une application $N : E \rightarrow \mathbb{R}$ qui vérifie pour tout vecteur x de \mathbb{R} :

(positivité) $N(x) \geq 0$ pour tout $x \in E$. Si $N(x) = 0$ alors $x = 0$

(homogénéité) $\forall \lambda \in \mathbb{R} \quad N(\lambda x) = |\lambda|N(x)$

(inégalité triangulaire) $\forall x, y \in E : \quad N(x + y) \leq N(x) + N(y)$

Toute norme induit une distance par $d(x, y) = N(x - y)$ (exercice). La réciproque est fautive : par exemple

$$d(x, y) = \begin{cases} 1 & \text{si } x \neq y \\ 0 & \text{si } x = y \end{cases}$$

est une distance qui ne satisfait clairement pas l'homogénéité. Si N est une norme, on a l'**inégalité triangulaire inférieure** : $|N(x) - N(y)| \leq N(x - y)$.

Démonstration. $N(x) = N(x - y + y) \leq N(x - y) + N(y)$ donne $N(x) - N(y) \leq N(x - y)$. On peut intervertir les rôles de x et y . Alors

$$|N(x) - N(y)| = \max(N(x) - N(y), N(y) - N(x)) \leq N(x - y).$$

1.2 Exemples de normes

- La **valeur absolue** dans \mathbb{R} est une norme (ce qui fait de $(\mathbb{R}, |\cdot|)$ un e.v.n.)
- On peut définir dans \mathbb{R}^d ou \mathbb{C}^d la **norme sup** :

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

- Soit dans \mathbb{R} ou \mathbb{C} le **produit scalaire** défini par \mathbb{C} :

$$\langle x, y \rangle = \langle x | y \rangle = \sum_{k=1}^d x_k \bar{y}_k$$

Définition 1.3. Un produit scalaire sur E est une application de $E \times E$ dans son corps scalaire avec

(PS 1) $\forall x \in \mathbb{R}^d : \langle x, x \rangle \geq 0$ et $\langle x, x \rangle = 0$ si et seulement si $x = 0$

(PS 2) $\forall x, y \in \mathbb{R}^d : \langle x, y \rangle = \overline{\langle y, x \rangle}$

(PS 3) $\forall \alpha, \beta \in \mathbb{R}, \forall x, y, z \in \mathbb{R}^d : \langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$.

On peut définir une norme associée à un produit scalaire : $\|x\|^2 = \langle x, x \rangle$.

Lemme 1.4. (Inégalité de Cauchy-Schwarz*) $|\langle x, y \rangle| \leq \|x\| \cdot \|y\|$

Demo dans le cahier de TD : $\forall \lambda \in \mathbb{R}, \langle x - \lambda y, x - \lambda y \rangle \geq 0$ ce qui nous donne un polynôme de degré 2 en λ qui est toujours positif donc son discriminant est toujours négatif

- d) $\|x\|_2 = \sqrt{\sum |x_i|^2}$ définit la **norme euclidienne**, cela permet d'étendre Pythagore dans \mathbb{R}^d : $\|x - y\|_2 = \sqrt{\sum |x_i - y_i|^2}$. On peut généraliser cette norme :

$$\|x\|_p = \left(\sum |x_i|^p \right)^{\frac{1}{p}}$$

Démonstration (qu'il s'agit d'une norme) :

La séparation et l'homogénéité sont faciles à voir : on additionne des termes positifs donc le résultat est positif, et pour qu'il soit nul, tous les termes doivent l'être. De plus $(\sum |\lambda x_i|^p)^{\frac{1}{p}} = |\lambda| (\sum |x_i|^p)^{\frac{1}{p}}$. reste à montrer l'inégalité triangulaire. On introduit la fonction f définie sur $(0,1)$ comme suit : Soit $a, b \geq 0$ et

$$f(t) = t^{1-p}a^p + (1-t)^{1-p}b^p$$

Cette fonction est positive, de plus sa limite en 0^+ et 1^- est $+\infty$. Elle admet donc un minimum, atteint lorsque sa dérivée s'annule, en un point t_0 . Après calcul (cf. cahier TD), on trouve $t_0 = \frac{a}{a+b}$. Résultat : $\inf\{f(t), t \in (0, 1)\} = f\left(\frac{a}{a+b}\right) = (a+b)^p$. De ce fait $|x_i + y_i|^p \leq t^{1-p}|x_i|^p + (1-t)^{1-p}|y_i|^p$ pour tout $t \in (0, 1)$. En sommant pour $i = 1 \dots n$ on obtient

$$\|x + y\|_p^p \leq t^{1-p}\|x\|_p^p + (1-t)^{1-p}\|y\|_p^p$$

On reconnaît la fonction f à droite pour $a = \|x\|_p$ et $b = \|y\|_p$. L'inégalité étant vraie pour tout $t \in (0, 1)$, on peut prendre l'infimum sur $t \in (0, 1)$. On obtient alors

$$\|x + y\|_p^p \leq (\|x\|_p + \|y\|_p)^p$$

ce qui donne le résultat en prenant la racine p -ième.

1.3 La topologie de \mathbb{R}^d

Définition 1.5. Soit $(E, \|\cdot\|)$ un e.v.n ; On appelle alors

- a) **Boule ouverte** : La boule ouverte de rayon r et de centre x , notée $B(x, r)$ définie par

$$B(x, r) \stackrel{\text{def.}}{=} \{y \in E \text{ tel que } \|x - y\| < r\}$$

*. D'après Augustin Louis, baron Cauchy, né à Paris le 21 août 1789 et mort à Sceaux le 23 mai 1857, mathématicien français, membre de l'Académie des sciences et professeur à l'École polytechnique, et Hermann Amandus Schwarz (né le 25 janvier 1843 à Hermsdorf en Silésie et mort le 30 novembre 1921 à Berlin), un mathématicien allemand.

b) **Boule fermée** : La boule fermée de rayon r , de centre x notée $B[x, r]$ définie par

$$B[x, r] = \{y \in E \text{ tel que } \|x - y\| \leq r\}$$

c) **Partie bornée** : Une partie $M \subset E$ est appelée bornée si

$$\exists C > 0 \text{ tel que } \forall m \in M, \|m\| \leq C$$

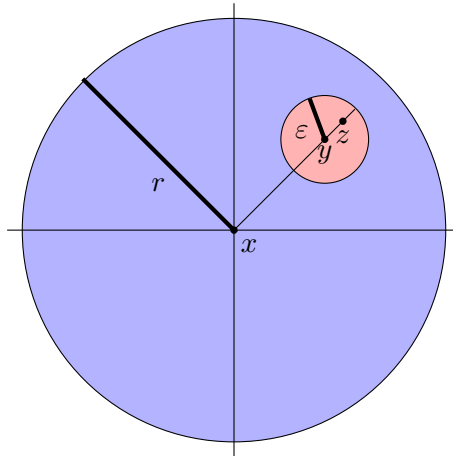
d) **Partie ouverte** : Une partie $O \subset E$ est appelée ouvert de E si :

$$\forall x \in O, \exists r > 0 \text{ tel que } B(x, r) \subset O$$

e) **Partie fermée** : Une partie $F \subset E$ est appelée fermé de E si son complémentaire dans E est un ouvert i.e. F^c est un ouvert de E .

Lemme 1.6. Une boule ouverte est un ouvert de E . Une boule fermée est un fermé de E .

Démonstration. Soit $y \in B(x, r) \Leftrightarrow \|x - y\| < r$. Il existe alors un $\varepsilon > 0$ tel que $\|x - y\| + \varepsilon < r$ (il suffit de prendre $\varepsilon = \frac{r - \|x - y\|}{2}$). Alors si $z \in B(y, \varepsilon)$ on a $\|y - z\| < \varepsilon$ alors : $\|x - y + y - z\| \leq \|x - y\| + \|y - z\| < r$. C'est à dire $\|x - z\| < r$ donc $z \in B(x, r)$.



Ce qui montre que $B(y, \varepsilon) \subset B(x, r)$: c'est à dire qu'on peut toujours trouver une boule ouverte incluse dans une boule ouverte, donc c'est un ouvert !

De même pour la boule fermée : on prend son complémentaire dans E : Soit $y \in (B[x, r])^c \Leftrightarrow \|x - y\| > r$ Il existe alors $\varepsilon > 0$ tel que $r + \varepsilon < \|x - y\|$. Ainsi, $\forall z$ tel que $\|z - y\| \leq \varepsilon$ on a $\|x - z\| \geq \|x - y\| - \|y - z\| \geq \|x - y\| - \varepsilon > r$ donc $B(y, \varepsilon) \subset (B[x, r])^c$, ce qui montre que $(B[x, r])^c$ est un ouvert. \square

Lemme 1.7. Soit E un e.v.n. $\mathcal{T} \stackrel{\text{def.}}{=} \{O \subset \mathbb{R}^d \text{ tel que } O \text{ est ouvert}\}$. Alors on a

(Topo 1) $\emptyset \in \mathcal{T}, \mathbb{R}^d \in \mathcal{T}$.

(Topo 2) $A, B \in \mathcal{T} \Rightarrow A \cap B \in \mathcal{T}$ (les ouverts sont stables par intersections finies)

(Topo 3) soit Λ un ensemble d'indices, et $\forall \lambda \in \Lambda : O_\lambda \in \mathcal{T}$. Alors $\bigcup_{\lambda \in \Lambda} O_\lambda \in \mathcal{T}$. (les ouverts sont stables par unions quelconques)

Démonstration. a) L'assertion $\forall x \in \emptyset$ est fausse, on ne peut pas choisir un x dans l'ensemble vide. De ce fait, $\forall x \in \emptyset : B(x, 1) \subset \emptyset$ est valide, car la proposition

$$\text{faux} \Rightarrow (\text{quoi que ce soit})$$

est toujours vraie. Pour la deuxième, observez que $\forall x \in \mathbb{R}^d : B(x, 42) \subset \mathbb{R}^d$.

b) Soit $A, B \in \mathcal{T}$; soit $x \in A \cap B$. Alors $\exists r > 0$ tel que $B(x, r) \subseteq A$ et $\exists \varepsilon > 0$ tel que $B(x, \varepsilon) \subseteq B$. Ainsi

$$B(x, \min(r, \varepsilon)) = B(x, r) \cap B(x, \varepsilon) \subseteq A \cap B$$

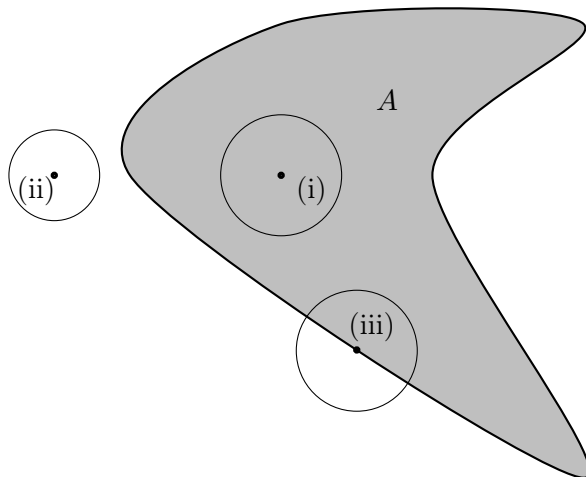
c) Soit $x \in \bigcup_{\lambda \in \Lambda} O_\lambda$. Alors $\exists \lambda_0 \in \Lambda$ tel que $x \in O_{\lambda_0}$, or O_{λ_0} est un ouvert donc $\exists r > 0$ tel que $B(x, r) \subseteq O_{\lambda_0} \subseteq \bigcup_{\lambda \in \Lambda} O_\lambda$. □

Définition 1.8. Soit \mathcal{T} une collection de parties d'un ensemble X vérifiant les propriétés (Topo1)–(Topo3). Alors on appelle \mathcal{T} une topologie sur X . L'étude de topologies "abstraites" sera faite en L3 dans un module "topologie".[†]

Remarque 1.9. On se rappelle que F fermé $\Leftrightarrow F^c$ ouvert. Par conséquent, les fermés sont stables par intersections quelconques et unions finies.

1.3.1 Intérieur, frontière (bord), adhérence d'un ensemble

Soit A un ensemble et x un point. Alors exactement un des trois cas suivants s'applique :



- i) Soit, $x \in A$ et $B(x, r) \subset A$ pour un certain $r > 0$.
- ii) Soit, $x \in A^c$ et $B(x, r) \subset A^c$ pour un certain $r > 0$.
- iii) Toute boule $B(x, r)$ intersecte à la fois A et A^c .

[†]. Mot dérivée du Grec "τόπος" (lieux) et "λόγος" (étude). La topologie étudie les propriétés d'objets géométriques préservées par déformation continue sans arrachage ni recollement, comme un élastique que l'on peut tendre sans le rompre. L'étude des applications continues qui est, comme on verra bientôt, est lié aux parties ouvertes, donne par extension le nom "topologie" à \mathcal{T} .

Exercice 1.10. Démontrer que si les situations (i) et (ii) sont fausses, alors (iii) est vraie.

Définition 1.11. a) $\overset{\circ}{A} = \text{int}(A)$, l'intérieur de A , formé des points de type (i)
 b) ∂A , la frontière ou le bord de A est l'ensemble formé des points de type (iii).
 c) \overline{A} adhérence de A , l'ensemble formé des points de type (i) ou (iii).

Exercice 1.12. Montrer que $x \in \overline{A}$ si et seulement si toute boule centrée en x rencontre A ("rencontrer" veut dire : a une intersection non-vide).

Il suit de la définition que O est ouvert si et seulement tous ses points sont de type (i). Les points de type (i) de F^{\complement} sont les points de type (ii) de F . Il suit que F est fermé si et seulement si tous ses points sont des types (i) ou (iii).

En particulier $\overset{\circ}{A}$ est un ouvert, ainsi que $(\overline{A})^{\complement}$. De cela suit immédiatement que (\overline{A}) est fermé! On a donc les suivantes propriétés simples :

- a) $A \subseteq B$ implique $\overset{\circ}{A} \subseteq \overset{\circ}{B}$ et $\overline{A} \subseteq \overline{B}$.
- b) $\overset{\circ}{A}$ est ouvert
- c) $\overset{\circ}{A} \subseteq A$ de plus : $\overset{\circ}{A} = A$ si et seulement si A est ouvert
- d) \overline{A} est fermé
- e) $A \subseteq \overline{A}$, de plus : $\overline{A} = A$ si et seulement si A est fermé.

Lemme 1.13.

$$\overset{\circ}{A} = \bigcup_{\substack{O \text{ ouvert} \\ O \subseteq A}} O \quad \text{et} \quad \overline{A} = \bigcap_{\substack{F \text{ fermé} \\ A \subseteq F}} F$$

Démonstration. Les deux par double inclusion!

- a) Première identité " \subseteq " : Soit U l'union à droite. Pour $x \in \overset{\circ}{A}$, il existe $r > 0$: $B(x, r) \subset A$. Or $B(x, r)$ ouvert, on a $x \in U$.
- b) Première identité " \supseteq " : Soit $x \in U$, il existe un O ouvert contenant x : Or, $x \in O$, il existe $r > 0$ tel que $B(x, r) \subset O \subset A$ donc $x \in \overset{\circ}{A}$.
- c) Deuxième identité " \subseteq " : Soit F un fermé contenant A . Alors $x \notin F$ implique $x \in F^{\complement}$ (un ouvert!), alors x serait de type (ii). Par contraposé, si x n'est pas de type (ii), nécessairement $x \in F$. On a donc $\overline{A} \subset F$ pour tout F fermé contenant A .
- d) Deuxième identité " \supseteq " : \overline{A} contient les points de type (i) ou (iii). Il suit que $\overline{A}^{\complement}$ est constituée des points de type (ii) – et donc un ouvert. Ainsi, on reconnaît \overline{A} comme un fermé qui contient A – d'où l'inclusion! \square

Exemple 1.14. Soit $I = (a, b]$ et $J = [a, b]$. Alors $\overset{\circ}{I} = (a, b)$ et $\overline{I} = [a, b]$ et $\overset{\circ}{J} = (a, b)$ et $\overline{J} = [a, b]$.

1.4 Limites et continuité

Définition 1.15. Soit $(E, \|\cdot\|)$ un e.v.n. et $(x_n)_n$ une suite dans E . On dit que $x_n \rightarrow x$ quand $n \rightarrow \infty$ ou bien que $\lim_{n \rightarrow \infty} x_n = x$, si

$$\forall \varepsilon > 0 \quad \exists N : \quad \forall n \geq N : \quad \|x_n - x\| \leq \varepsilon$$

C'est à dire : c'est la définition "bien connue" de la limite où on se sert de la norme (à la place de la valeur absolue) pour mesurer des distances entre x_n et ℓ .

La convergence $x_n \rightarrow x$ veut dire que toute boule ouverte de centre x contient tous les termes de la suite à partir d'un certain rang. C'est à dire que toute (!) boule autour de x contient tous les termes de la suites sauf, peut-être, un nombre fini d'exceptions.

Les assertions en quantificateurs, comme dans la définition ci-dessus s'apprennent avec une "phrase en prose" (certes, pas candidat pour le prix Goncourt) : ici, par exemple :
"Pour tout epsilon il existe un rang à partir duquel les distances de x_n à ℓ deviennent plus petites que epsilon"

Lemme 1.16. Soit $A \subset E$ un ensemble, et $L = \{\lim_{n \rightarrow \infty} a_n : (a_n) \in A^{\mathbb{N}} \text{ qui convergent}\}$ (i.e. l'ensemble des limites des suites de A qui convergent dans E) alors,

$$L = \bar{A} = A \cup \partial(A)$$

Démonstration. \supseteq : Il suffit de montrer que $A \subset L$ et L fermé. $\forall a \in A$; $a_n = a$ (la suite constante égale à a) montre que $A \subset L$ (c'est à dire on prend chaque élément de A et on l'associe comme limite d'une suite constante $a_n = a$, donc a est forcément dans L). On va montrer que L est un fermé qui contient A , donc L est à priori plus grand que l'adhérence de A : $\bar{A} \subset L$: Montrons que L est fermé. Soit $x \in L^c$, alors $r = \inf\{\|x - a\|; a \in A\}$ est strictement positif. En effet, si on avait $r = 0$, $\forall n, \exists a_n \in A$ tel que $\|a_n - x\| < 1/n$, ce qui voudrait dire que $a_n \rightarrow x$ mais x n'est pas dans L . Il suit que $B(x, R/2) \subset L^c$ i.e. L^c est un ouvert. Et L est fermé.

\subseteq : Soit $x \in L$, (a_n) une suite de A qui converge vers x . Si x était de type 2 (en dehors), on aurait une boule $B(x, r)$ autour, dans $L^c \subset A^c$: ce n'est pas possible car $a_n \rightarrow x$, donc $\exists n$ tel que $a_n \in B(x, r)$, donc x est soit intérieur soit "sur le bord" de A , donc $\bar{A} \subset L$. D'où l'égalité recherchée. \square

Définition 1.17. (Suites de Cauchy) Une suite est appelée de Cauchy si :

$$\forall \varepsilon > 0, \exists N \text{ tel que } n, m \geq N \Rightarrow \|u_n - u_m\| < \varepsilon$$

Attention : ce doit être vrai pour tous les n, m au-delà d'un certain rang. Par exemple pour $x_n = \sqrt{n}$ on a : $|x_{n+1} - x_n| \leq \frac{1}{2\sqrt{n}} \rightarrow 0$, pourtant clairement ce n'est pas une suite de Cauchy, elle ne converge pas.

Lemme 1.18. *Toute suite de Cauchy est bornée. La réciproque est fausse en général.*

Démonstration. Soit $\varepsilon = 42$. Il existe un rang N à partir duquel la distance entre les termes est inférieure à ε . Donc la boule $B[x_N, 42]$ contient presque tous les termes de la suite (x_n) (tous sauf un nombre fini). Donc $\forall k \in \mathbb{N}, \|x_k\| \leq \max(\|x_1\|, \dots, \|x_N\|) + 42$. Elle est donc bornée.

La suite $(\sin(n))_{n \geq 1}$ est bornée, mais pas Cauchy : il existe une infinité d'indices avec $\sin(n) > 1/2$ et une (autre) infinité d'indices avec $\sin(n) < -1/2$. \square

Lemme 1.19. *Toute suite convergente est de Cauchy. La réciproque est fautive en général.*

Démonstration. Soit (α_n) une suite convergente, $\forall \varepsilon > 0, \exists N$ tel que $\forall m, n \geq N, \|\alpha_n - \alpha_m\| < \frac{\varepsilon}{2}$ et $\|\alpha_m - l\| < \frac{\varepsilon}{2}$. Alors $\|\alpha_n - \alpha_m\| \leq \|\alpha_n - l\| + \|\alpha_m - l\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$. \square

Exemple 1.20. (\star). Vu le Théorème 1.24 en bas, il est nécessaire de construire un contre-exemple à la réciproque en dimension infinie (ou bien dans un espace métrique). Les espaces vectoriels de dim. infinie les plus simples sont probablement des espaces de suites : Soit donc

$$E = \{(x_n) \in \mathbb{R}^{\mathbb{N}} : \exists N > 0 : \forall n \geq N : x_n = 0\}$$

l'espace vectoriel des suites nulles à partir d'un certain rang, muni de la norme sup (exercice : c'est une norme!). On considère les suites

$$x_1 = (1, 0, \dots), \quad x_2 = (1, 1/2, 0, \dots), \quad x_3 = (1, 1/2, 1/3, 0, \dots), \quad x_4 = (1, 1/2, 1/3, 1/4, 0, \dots), \quad \dots$$

Soit $\min(n, k) \stackrel{\text{def.}}{=} N$ et $\max(n, k) \stackrel{\text{def.}}{=} M$. On voit que

$$\|x_n - x_k\| = \max\{1/N+1, 1/N+2, \dots, 1/M\} = 1/N+1$$

Ainsi, (x_n) est une suite de Cauchy (de suites). Car convergence pour la norme sup entraîne convergence coordonnée par coordonnée, la seule limite possible est la suite $(1/n)_{n \geq 1}$ — qui n'appartient pas à E ! On a donc une suite de Cauchy qui ne converge pas. Voici une version "fonctions" au lieu de "suites" de l'exemple :

Exercice 1.21. ($\star\star$) Soit $E = \{f : \mathbb{R} \rightarrow \mathbb{R} \text{ continues, telles que } \exists M > 0 : \forall |x| > M : f(x) = 0\}$. (le M dépend, bien entendu, de f , il n'est pas uniforme! On peut munir E de la norme $\|f\| = \int_{\mathbb{R}} |f(x)| dx$. Soit $f_n(x) = \exp(-x^2) \mathbb{1}_{[-n, n]}(x)$. Montrer que (f_n) est Cauchy, mais ne converge pas.

Définition 1.22. Un e.v.n. dans lequel toute suite de Cauchy converge est appelé un espace **complet**. Un espace vectoriel normé complet est appelé un espace de Banach[‡]

L'importance d'espaces complets réside dans le fait qu'on peut montrer la convergence *sans connaître un "candidat" pour la limite!*

Sous-suites extraites

Soit $(x_n)_{n \geq 0}$ une suite dans E . Soit $\varphi : \mathbb{N} \rightarrow \mathbb{N}$ une application strictement croissante. On appelle $(x_{\varphi(n)})$ la sous-suite extraite de (x_n) . Si $(x_n) \rightarrow x$ alors toute sous-suite $(x_{\varphi(n)}) \rightarrow x$. La réciproque est vraie aussi!

[‡]. d'après Stefan Banach, mathématicien polonais né en 1892 à Cracovie et mort en 1945 - d'un cancer du poumon. Banach est un des mathématiciens les plus influents du XXe siècle.

Rappel notations : On peut concevoir un élément $x \in \mathbb{R}^d$ comme une fonction

$$f : \{1, \dots, d\} \rightarrow \mathbb{R} \quad i \mapsto x_i$$

De même, une suite dans \mathbb{R} est une fonction f :

$$f : \mathbb{N} \rightarrow \mathbb{R} \quad n \mapsto x_n$$

Lemme 1.23. Soit (f_n) une suite de \mathbb{R}^d (i.e. $f_n : \{1, \dots, d\} \rightarrow \mathbb{R}$). Alors $f_n \xrightarrow{\|\cdot\|_2} f$ signifie : $f_n(i) \xrightarrow{n \rightarrow \infty} f(i)$, $\forall i = 1, \dots, d$.

Dans \mathbb{R}^n , la convergence en norme (Euclidienne) est équivalente à la convergence coordonnée par coordonnée.

Démonstration. \Rightarrow : $|f_n(i) - f(i)| \leq (\sum_{j=1}^d |f_n(j) - f(j)|^2)^{1/2}$ (si f_n converge, tous les $f_n(i)$ aussi). \Leftarrow : Soit $\varepsilon > 0$, $\forall j \in \{1, 2, \dots, d\}$, $\exists N_j$ tel que $\forall n \geq N_j$ on a $|f_n(j) - f(j)| \leq \frac{\varepsilon}{\sqrt{d}}$ (= convergence par coordonnée). Pour $n \geq \max\{N_1, N_2, \dots, N_d\}$, $(\sum_{j=1}^d |f_n(j) - f(j)|^2)^{1/2} \leq \varepsilon$ (Il suffit de sommer de chaque côté de l'inégalité puis de prendre la racine). \square

Théorème. 1.24. Pour tout $d \geq 1$, \mathbb{R}^d est complet. Par conséquent tout e.v.n. de dimension finie est complet.

Démonstration. On utilise que \mathbb{R} est complet. On peut plonger davantage dans cette propriété, et la construction de \mathbb{R} , mais ce n'est pas le moment.

Soit donc (x_n) Cauchy dans \mathbb{R}^n . Alors $(\pi_k(x_n))_{n \geq 1}$ est Cauchy pour tout $k = 1 \dots d$. Il suit que $\pi_k(x_n) \rightarrow \ell_k$ et donc, vu le lemme précédent, $x_n \rightarrow (\ell_1, \dots, \ell_d)$. \square

Continuité

Définition 1.25. (Fonction continue) Soient E, F , des e.v.n., $\Omega \subset E$ ouvert et $f : \Omega \rightarrow F$ une application. On dit que f est continue en $x \in \Omega$ si pour toute suite (x_n) dans Ω qui converge vers x on a :

$$\lim_{n \rightarrow +\infty} f(x_n) = f(x) = f(\lim_{n \rightarrow +\infty} x_n)$$

Voici trois exemples importants :

Exemple 1.26. a) $\frac{xy}{x^2+y^4} = f(x, y)$ $\Omega = \mathbb{R}^2 \setminus \{(0, 0)\}$, $f(\frac{1}{n}, 0) = 0$ quand $n \rightarrow \infty$

et $f(\frac{1}{n^2}, \frac{1}{n}) = \frac{\frac{1}{n^2} \cdot \frac{1}{n}}{\frac{1}{n^4} + \frac{1}{n^4}} = \frac{n}{2} \rightarrow +\infty$. Donc : on ne peut pas étendre f par continuité en 0.

b) $g(x, y) = \frac{xy^2}{x^2+y^2}$ $\Omega = \mathbb{R}^2 \setminus \{(0, 0)\}$.

i) Parfois la technique du "laisser tomber" s'avère efficace : $|g(x, y)| = |\frac{xy^2}{x^2+y^2}| \leq |\frac{xy^2}{0+y^2}| = |x|$ et $|x| \rightarrow 0$ si $(x, y) \rightarrow 0$. g est donc continue à l'origine.

ii) Si le dénominateur est radiale, il est aussi facile de passer aux coordonnées polaires :

$$\begin{cases} x_n = r_n \cos \psi_n \\ y_n = r_n \sin \psi_n \end{cases}$$

on a $|g(r \cos \psi, r \sin \psi)| = \left| \frac{r \cos \psi r^2 \sin^2 \psi}{r^2 (\cos^2 \psi + \sin^2 \psi)} \right| = |\cos \psi \sin^2 \psi| \cdot |r| \leq |r| = G(r)$.

L'importance est de trouver une fonction majorante $G(r)$ qui ne dépend pas de l'angle ψ et qui tend vers 0 quand le couple (x, y) –ou le rayon r – tend vers zéro. Donc ici la limite de $g(x, y)$ en $(0, 0)$ existe et vaut 0.

c) Considérer $h(x, y) = \frac{6x^4y^2 - 2x^2y^6}{(x^2 + y^4)^2}$ quand $(x, y) \neq 0$ et zéro sinon. Le dénominateur est de degré hétérogène. Il y a une ruse :

$$2ab \leq a^2 + b^2$$

Avec cela, on simplifie, d'abord : le facteur $2xy^2$ se majore par $x^2 + y^4$ qui simplifie le dénominateur. Puis on utilise "laisser tomber" :

$$\begin{aligned} |h(x, y)| &\leq \left| \frac{(x^2 + y^4)(3x^3 - 2xy^4)}{(x^2 + y^4)^2} \right| = |x| \left| \frac{(3x^2 - 2y^4)}{x^2 + y^4} \right| \\ &\leq |x| \left(\frac{3x^3}{x^2 + 0} + \frac{2y^4}{0 + y^4} \right) \leq |x|(3|x| + 2). \end{aligned}$$

Clairement, le majorant trouvé tend vers zéro lorsque $(x, y) \rightarrow (0, 0)$. Au lieu de "laisser tomber" il y a aussi une autre possibilité "homogénéiser" des coefficients : en effet,

$$\frac{(3x^2 + 2y^4)}{x^2 + y^4} \leq \frac{(3x^2 + 3y^4)}{x^2 + y^4} \leq 3.$$

"Homogénéiser" est une technique qui s'applique aussi dans d'autres situations : par exemple, on a

$$\frac{x^4y^2}{x^2 + y^4} \leq \frac{x^4y^2}{x^4 + y^4}$$

dès que $|x| < 1$, ce qui ramène sur une situation radiale qui se comporte $\leq \frac{c}{r^2}$: En effet, $\frac{1}{c}(x^2 + y^2)^2 \leq (x^4 + y^4) \leq C(x^2 + y^2)^2$ juste par équivalence des normes $\|\cdot\|_2$ et $\|\cdot\|_4$. Pas besoin de connaître ou expliciter $c > 0$ ici !

Théorème. 1.27. Une fonction f est continue en $x = a$ si, et seulement si

$$(\forall \varepsilon > 0, \exists \delta > 0 \text{ tel que } \forall y : \|y - a\| < \delta \Rightarrow \|f(y) - f(a)\| < \varepsilon) \quad (1.1)$$

De plus, les assertions suivantes sont équivalentes :

- f continue en tout point
- $\forall U \subset F$, ouvert, $f^{-1}(U)$ est un ouvert (i.e. l'image réciproque de U par F)
- $\forall A \subset F$ fermé, $f^{-1}(A)$ est fermé.

Démonstration. L'équivalence de la définition par suites et la caractérisation ε/δ : **exercice à faire !**

- a) (a) \Rightarrow (c) (si f est continue alors l'image réciproque par f d'un fermé est fermé) : Soit (x_n) une suite de $f^{-1}(A)$ qui converge dans E vers x . Par continuité $f(x_n) \rightarrow f(x)$ or $f(x_n) \in A \forall n \Rightarrow f(x) \in \overline{A} = A$. Donc $f(x) \in A$ ce qui implique $x \in f^{-1}(A)$.
- b) (c) \Rightarrow (b) (si pour tout fermé de F , l'image réciproque de ce fermé est un fermé alors pour tout ouvert de F , l'image réciproque de cet ouvert par f est un ouvert) : Il suffit de passer au complémentaire : $(f^{-1}(A))^c = f^{-1}(A^c)$
- c) (b) \Rightarrow (a) (si tout ouvert a pour image réciproque par f un ouvert alors f est continue sur cet ouvert) : Soit $\varepsilon > 0$ et $x \in E$. $U = B(f(x), \varepsilon)$ est un ouvert de F . Par conséquent $f^{-1}(U)$ est un ouvert de E et $x \in f^{-1}(U) \Rightarrow \exists \delta > 0$ tel que $B(x, \delta) \subset f^{-1}(U)$. Il suit que $\|x - y\| < \delta \Rightarrow \|f(x) - f(y)\| < \varepsilon$. C'est la caractérisation de la continuité. □

Attention : Le théorème parle bien de l'image réciproque. En revanche, l'image directe d'un ouvert peut être ouvert (p.ex. $f(x) = x$ sur \mathbb{R}) mais aussi fermé, p.ex. $\sin((-42, 42)) = [-1, 1]$ ou ni-ni : p.ex. $\sin((0, 2)) = (0, 1]$. De même, l'image directe d'un fermé peut-être fermé (à nouveau $f(x) = x$) mais aussi ouvert : $F = \{(x, 1/x) \in \mathbb{R}^2 : x > 0\}$ est fermé, mais $\pi_1(F) = (0, \infty)$ est ouvert. Ou ni-ni ! Par exemple $\tilde{F} = F \cap \{(x, y) \in \mathbb{R}^2 : x \leq 42\}$ est fermé mais $\pi_1(\tilde{F}) = (0, 42]$.

Remarque 1.28. Pour $\|\cdot\|_p$, la convergence dans \mathbb{R} équivaut à la convergence coordonnée par coordonnée, ce qui veut dire que la continuité se teste coordonnée par coordonnée. Par conséquent, $f : \Omega \rightarrow \mathbb{R}^m$ continue $\Leftrightarrow f_i : \Omega \rightarrow \mathbb{R}$ continue pour tout $i \in \llbracket 1; m \rrbracket$

Exemple 1.29. Montrons que $A = \{(x, y) \text{ tel que } x < 3 \text{ et } y < 5\}$ est ouvert dans $(\mathbb{R}^2, \|\cdot\|_2)$: nous avons $A = f^{-1}((-\infty, 3)) \cap g^{-1}((5, +\infty))$ où $f(x, y) = x$ et $g(x, y) = y$. Montrons alors que f et g sont continues. Soit $(x_n, y_n)_{n \geq 1}$ une suite dans \mathbb{R}^2 qui converge vers (x, y) . Or $|x_n - x| = ((x_n - x)^2)^{1/2} \leq ((x_n - x)^2 + (y_n - y)^2)^{1/2} = \|(x_n, y_n) - (x, y)\|_2$. Or $\|(x_n, y_n) - (x, y)\|_2 \rightarrow 0$ donc $x_n \rightarrow x$. Symétriquement on obtient $y_n \rightarrow y$. Donc f et g sont continues, d'où A qui est l'intersection (finie !) des images réciproques par f et g d'ouverts, est ouvert.

Lemme 1.30. Soit E, F, G des e.v.n. ; $f : E \rightarrow F$ et $g : F \rightarrow G$ deux fonctions telles que f est continue en x_0 et g continue en $y_0 = f(x_0)$. Alors $g \circ f$ continue en x_0 .

Démonstration. Si $x_n \rightarrow x_0$, par continuité de $f : f(x_n) \rightarrow f(x_0) = y_0$. Par continuité de $g : g(f(x_n)) \rightarrow g(y_0) = g(f(x_0))$ donc $g \circ f(x_n) \rightarrow g \circ f(x_0)$ donc $g \circ f$ est continue en x_0 . □

Exemple 1.31. Toute norme $N : \begin{cases} (E, \|\cdot\|) \rightarrow (\mathbb{R}, |\cdot|) \\ x \rightarrow \|x\| \end{cases}$ est une application continue par

rapport à elle-même : en utilisant l'inégalité triangulaire inférieure on a : $|N(x) - N(y)| \leq N(x - y)$. On veut montrer (1.1). Mais : $N(x - y) < \delta \Rightarrow |N(x) - N(y)| \leq N(x - y) < \delta$ donc si $\delta = \varepsilon$ on a bien l'égalité demandée.

Autre preuve, avec les suites : soit $(x_n) \rightarrow x$. Donc $\forall \varepsilon > 0, \exists \delta > 0$ tel que $\|x_n - x\| < \delta$. $|N(x_n) - N(x)| \leq N(x_n - x) < \delta$. Donc en prenant de même $\varepsilon = \delta$, on a $N(x_n) - N(x) < \varepsilon$ donc N est continue.

Lemme 1.32. (Propriétés de fonctions continues)

- a) Si $x_n \rightarrow x$ et $y_n \rightarrow y$ alors $x_n + y_n \rightarrow x + y$ i.e. L'addition comme opération sur E est continue.
- b) Si $x_n \rightarrow x$ dans E et $\lambda_n \rightarrow \lambda$ dans \mathbb{K} alors $\lambda_n x_n \rightarrow \lambda x$.
- c) Si $x_n \rightarrow x$ et $y_n \rightarrow y$ dans E et si E est normé par une norme provenant d'un produit scalaire alors : $\langle x_n, y_n \rangle \rightarrow \langle x, y \rangle$.

Démonstration. a) $\|(x_n + y_n) - (x + y)\| \leq \|x_n - x\| + \|y_n - y\|$ et chaque norme de droite tend vers 0 d'où le résultat. Cela signifie que $\text{add} : (x, y) \rightarrow x + y$ est continue.

b)

$$\begin{aligned} \|(x_n \lambda_n) - (x \lambda)\| &= \|\lambda_n x_n - \lambda x_n + \lambda x_n - \lambda x\| \\ &\leq \|\lambda_n x_n - \lambda x_n\| + \|\lambda x_n - \lambda x\| \\ &\leq |\lambda_n - \lambda| \cdot \|x_n\| + |\lambda| \cdot \|x_n - x\| \end{aligned}$$

or $\|x_n - x\| \rightarrow 0$, donc $|\lambda| \cdot \|x_n - x\| \rightarrow 0$. D'autre part : $|\lambda_n - \lambda| \rightarrow 0$ et, toute suite convergente étant bornée, $\exists M, \forall n, \|x_n\| \leq M$ donc $|\lambda_n - \lambda| \cdot \|x_n\| \rightarrow 0$. Cela signifie que $m_\lambda : x \mapsto \lambda \cdot x$ est continue.

c)

$$|\langle x_n, y_n \rangle - \langle x, y \rangle| \leq |\langle x_n, y_n \rangle - \langle x, y_n \rangle| + |\langle x, y_n \rangle - \langle x, y \rangle| = |\langle x_n - x, y_n \rangle| + |\langle x, y_n - y \rangle|$$

or d'après Cauchy-Schwarz on a :

$$|\langle x_n - x, y_n \rangle| + |\langle x, y_n - y \rangle| \leq \|x_n - x\| \cdot \|y_n\| + \|x\| \cdot \|y_n - y\|.$$

En résumé, $|\langle x_n, y_n \rangle - \langle x, y \rangle| \leq \|x_n - x\| \cdot \|y_n\| + \|x\| \cdot \|y_n - y\|$ et ceci tend vers zéro car $x_n \rightarrow x$, $\|y_n\|$ est bornée car convergente, et $\|x\|$ est une constante donc $\|x\| \cdot \|y_n - y\|$ tend vers 0. Cela signifie que $\text{prod} : (x, y) \mapsto \langle x | y \rangle$ est continue. \square

Théorème. 1.33. Soit $L : E \rightarrow F$ une application linéaire entre deux e.v.n. Alors sont équivalents :

- a) L est continue
- b) L est bornée, i.e.

$$\|L\|_{\mathcal{L}(E,F)} \stackrel{\text{def.}}{=} \sup \left\{ \frac{\|L(x)\|_F}{\|x\|_E}, x \neq 0 \right\}$$

est fini. La quantité $\|L\|_{\mathcal{L}(E,F)}$ est appelée la norme d'opérateur de L . S'il n'y a pas de risque de confusion sur les normes de E et F , on peut aussi noter $\|L\|$.

Démonstration. \Rightarrow : Si L est bornée on a $\|L(x) - L(y)\| = \|L(x - y)\|$ par linéarité de L . Comme $\|L\| = \sup\{\frac{\|L(x)\|_F}{\|x\|_E}, x \neq 0\}$ on a $\frac{\|L(x-y)\|}{\|x-y\|} \leq \|L\|$ donc $\|L(x - y)\| \leq \|L\| \cdot \|x - y\|$.

Ainsi si ε est fixé, et que $\|x - y\| \leq \delta \stackrel{\text{def.}}{=} \frac{\varepsilon}{\|L\|}$ on a $\|L(x) - L(y)\| < \varepsilon$. Ce qui montre que L est (Lipschitz-) continue.

\Leftarrow : Soit L continue. Supposons que L ne soit pas bornée : cela signifie $\exists x_n \in E$ tel que $\frac{\|L(x_n)\|}{\|x_n\|} > n$ (*). Soit $y_n = \frac{x_n}{\|x_n\|\sqrt{n}}$ alors $\|y_n\| = \frac{1}{\sqrt{n}} \rightarrow 0$. Donc (y_n) est une suite qui tend vers zéro. Comme L est linéaire $L(0_E) = 0_F$. Donc $L(y_n)$ devrait tendre vers $L(0) = 0$. Or : $\|L(y_n)\| = \frac{1}{\|x_n\|\sqrt{n}} \cdot \|L(x_n)\| \geq \frac{n}{\sqrt{n}}$ d'après (*). Donc $L(y_n) \not\rightarrow 0$, ce qui contredit l'hypothèse de L continue. Donc si L est continue elle est nécessairement bornée. \square

Il est fort recommandé de faire l'exercice suivant en détail.

Exercice 1.34. Montrer que $\|\cdot\|_{\mathcal{L}(E,F)}$ est une norme ! Ensuite montrer que

$$\|T\|_{\mathcal{L}(E,F)} = \sup\{\|Tx\|_F; \|x\| \leq 1\} = \inf\{C > 0; \forall x \in E : \|Tx\|_F \leq C\|x\|_E\}$$

L'égalité des deux sup's est relativement simple. Pour l'égalité "inf=sup", montrer une double inégalité. Soit

$$\mathfrak{s} := \sup\{\frac{\|Tx\|_F}{\|x\|_E}; x \neq 0\} \quad \text{et} \quad I = \{c > 0; \forall x \in E; \|Tx\|_F \leq c \cdot \|x\|_E\}$$

Justifier que $\|Tx\| \leq \mathfrak{s} \cdot \|x\|$ pour tout $x \in E$ et donc que $\mathfrak{s} \in I$. En déduire $\|T\| \leq \mathfrak{s}$. Expliquer pourquoi $I \subset \mathbb{R}$ est un intervalle, pour conclure que $(\|T\| + \varepsilon) \in I$. En déduire que $\mathfrak{s} \leq \|T\| + \varepsilon$. Conclure.

Notation. Rappeler les normes $\|\cdot\|_p$ sur \mathbb{R}^d avec $1 \leq p \leq \infty$ de la section 1.2. Dans le cas de $E = (\mathbb{R}^n, \|\cdot\|_p)$ et $F = (\mathbb{R}^m, \|\cdot\|_q)$ on notera pour une matrice A

$$\|A\|_{p \rightarrow q} \stackrel{\text{def.}}{=} \|A\|_{\mathcal{L}(E,F)}$$

pour se rappeler les normes dans l'espace de départ et d'arrivée.

Exercice 1.35. Soit $p = q = \infty$ et $A = (a_{i,j})$. Alors

$$\|A\|_{\infty \rightarrow \infty} = \max_i \sum_j |a_{i,j}|$$

Dans le cas $p = q = 1$ on a

$$\|A\|_{1 \rightarrow 1} = \max_j \sum_i |a_{i,j}|$$

Calculer ces deux normes pour

$$A = \begin{pmatrix} 1 & 3 \\ 2 & 2 \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}$$

La norme $\|A\|_{2 \rightarrow 2}$ est particulièrement intéressante, mais difficile à calculer. Voir Remarque 2.19 en bas.

1.5 Compacité (séquentielle)

Définition 1.36. Soit E un e.v.n. et $K \subset E$. On dit que K est (séquentiellement) compact si toute suite dans K admet une sous-suite extraite qui converge vers une limite dans K :

$$\forall (x_n)_n \in K^{\mathbb{N}}, \exists \varphi : \mathbb{N} \rightarrow \mathbb{N} \text{ strictement croissant t.q. } (x_{\varphi(n)})_n \xrightarrow{n \rightarrow \infty} l \in K$$

Lemme 1.37. Si $K \subset E$ est compact et $f : E \rightarrow F$ est continue alors $f(K)$ est compact. i.e. l'image d'un compact par une fonction continue est un compact.

Démonstration. Soit (y_n) une suite de $f(K)$. Alors il existe (x_n) telle que $y_n = f(x_n)$. Par compacité de K , il existe une sous-suite de (x_n) , appelons la $x_{\varphi(n)} \rightarrow l \in K$. Par continuité de f , $f(x_{\varphi(n)}) = y_{\varphi(n)} \rightarrow f(l) \in f(K)$. Donc $f(K)$ est compact. \square

Attention : L'image *directe* (continue) d'un compact est un compact — contrairement à la caractérisation des applications continues par image *reciproque*. Par contre l'image *reciproque* (continue) d'un compact n'a aucune raison d'être compacte : Soit $f(x) = 0$, et $K = \{0\}$: clairement K est compact, alors que $f^{-1}(K) = \mathbb{R}$ n'est pas compact : la suite $x_n = n$ par exemple n'admet pas de sous-suite convergente !

Lemme 1.38. Toute suite réelle admet une sous-suite monotone.

Démonstration. Notation ad hoc : on appelle x_{n_0} un "sommet" de la suite si $\forall n \geq n_0, x_n \leq x_{n_0}$. Soit il existe un nombre infini de sommets soit non.

Premier cas : Il y a une infinité de sommets. La suite de sommets est décroissante par définition du sommet.

Second cas : il existe un nombre fini de sommets, c'est à dire qu'à partir d'un certain rang $N \forall n \geq N, x_n$ n'est plus un sommet. Posons $\varphi(1) = N$. Comme x_N n'est pas un sommet, il existe k_1 tel que $x_{N+k_1} > x_N$. On pose $\varphi(2) = N + k_1$. Puis $x_{\varphi(2)}$ n'est pas un sommet, donc il existe un k_2 tel que $x_{k_2+\varphi(2)} > x_{\varphi(2)}$. On pose $\varphi(3) = k_2 + \varphi(2)$. On construit ainsi une suite strictement croissante $(x_{\varphi(n)})$ avec $\varphi(n) = k_{n-1} + \varphi(n-1)$. \square

Théorème. 1.39. Une partie $K \subset \mathbb{R}$ est compacte $\Leftrightarrow K$ est bornée et fermée

Démonstration. " \Rightarrow " borné : K compact. Supposons qu'il soit non-borné, alors $\exists x_n \in K$ tel que $|x_n| > n$. Ce qui signifie que (x_n) n'admet pas de sous-suite convergente (et donc bornée) donc K ne serait pas compact.

" \Rightarrow " fermé : K compact. Soit $(x_n) \in K$ tel que $x_n \rightarrow x$. Par compacité, il existe une sous-suite $x_{\varphi(n)}$ qui converge vers une limite $k \in K$. Mais par unicité de la limite $x_{\varphi(n)} \rightarrow x$ donc $x = k$ et $x \in K$. K est fermé.

" \Leftarrow " K est fermé et borné. Soit $(x_n) \subset K^{\mathbb{N}}$ une suite. Par le lemme précédent, (x_n) admet une sous-suite monotone, $x_{\varphi(n)}$. Elle est monotone par définition et bornée car dans K . Donc elle converge vers une limite l . Or K est fermé donc $l \in K$. Donc K est compact. \square

Théorème. 1.40. Une partie $K \subset \mathbb{R}^d$ est compacte $\Leftrightarrow K$ est bornée et fermée dans \mathbb{R}^d .

Démonstration. \Rightarrow : K compact, et $f(x) = \|x\|$ continue. Or $f(K)$ compact, il existe C tel que pour $\forall x \in K, \|x\|_2 \leq C$, par le théorème précédent. Donc K est borné. Si (x_n) est une suite de K qui converge dans \mathbb{R}^d vers x , il existe une sous-suite qui converge vers $k \in K$ et par unicité de la limite $k = x \in K$. Donc K est fermé.

\Leftarrow : K est fermé et borné. Soit $x = (x_1, \dots, x_n)$ et

$$\pi_k : \begin{cases} \mathbb{R}^d \rightarrow \mathbb{R} \\ (x_1, \dots, x_n) \rightarrow x_k \end{cases}$$

C'est à dire π_k est la projection du vecteur x sur sa k -ième composante. π_k est linéaire, continue et $|x_k| \leq \|x\|_2$. Il suit que $\pi_k(K)$ est un borné de \mathbb{R} . Donc pour toute suite $(x_n)_n$ dans K , $(\pi_k(x_n))_n$ est une suite bornée de \mathbb{R} . On construit successivement des sous-suites :

- Or $(\pi_1(x_n))_n$ est bornée dans \mathbb{R} , il existe φ_1 telle que $(\pi_1(x_{\varphi_1(n)}))_n$ converge dans \mathbb{R} vers l_1 .
- Or $(\pi_2(x_{\varphi_1(n)}))_n$ est bornée, donc il existe φ_2 telle que $(\pi_2(x_{\varphi_1 \circ \varphi_2(n)}))_n$ converge vers l_2 . De plus (argument clé!) $(\pi_1(x_{\varphi_1 \circ \varphi_2(n)}))_n$ converge vers l_1 car c'est une sous-suite extraite de $(\pi_1(x_n))_n$.
- Par récurrence finie, après $\dim(\mathbb{R}^d)$ étapes on obtient $\varphi = \varphi_1 \circ \varphi_2 \circ \dots \circ \varphi_d$ et $(\pi_k(x_{\varphi(n)}))_n \rightarrow l_k \forall k \in 1, \dots, n$.

On a vu que la convergence coordonnée par coordonnée et la convergence sont équivalentes dans $(\mathbb{R}, \|\cdot\|_2)$, d'où :

$$x_{\varphi(n)} \rightarrow (l_1, \dots, l_n).$$

Or K est fermé donc $(l_1, \dots, l_n) \in K$, et K est compact. □

1.6 Équivalence de normes

Définition 1.41. Soient $\|\cdot\|$ et $\|\!\|\!\|$ deux normes sur E . On dit que $\|\cdot\|$ est plus faible que $\|\!\|\!\|$ s'il existe une constante $C > 0$ telle que

$$\forall x \in E, \quad \|x\| \leq C \cdot \|\!\|\!\|x\!\!\!\|.$$

Si $\|\cdot\|$ est plus forte que $\|\!\|\!\|$ et si $\|\!\|\!\|$ est plus forte que $\|\cdot\|$, i.e. si on a une double inégalité, on dit que les normes sont équivalentes.

Remarque 1.42. (Illustration des boules associées) Soit $A = B_{\|\cdot\|}(0; r)$ avec $r = 1/C$, où C est la constante dans la définition ci-dessus. Soit $B = B_{\|\!\|\!\|}(0, 1)$. Alors $B \subset A$. La norme plus forte a donc (à changement d'échelle près) des boules plus petites.

Pour les normes équivalentes, la "notion de converger" coïncide : On a $x_n \xrightarrow{\|\cdot\|} x \Leftrightarrow \|x_n - x\| \rightarrow 0$ et par équivalence cela implique $\|\!\|\!\|x_n - x\!\!\!\| \rightarrow 0$ et donc $x_n \xrightarrow{\|\!\|\!\|} x$. Idem dans l'autre sens. Mais alors, les fermés (qui sont les parties stables sous prise de limite)

sont les mêmes pour les deux normes, donc les ouverts aussi. Finalement, les parties séquentiellement compactes sont les mêmes! Ceci explique l'intérêt pour des normes équivalentes.

Exemple 1.43. Dans $E = \mathbb{R}^d$ muni des normes $\|x\|_1 = \sum_{i=1}^d |x_i|$ et $\|x\|_2 = (\sum_{i=1}^d |x_i|^2)^{1/2}$.

- a) On utilise Cauchy-Schwarz avec $\langle \text{sign}(x), x \rangle = \sum_{i=1}^d \text{signe}(x_i) \cdot x_i = \|x\|_1$, où on peut poser le signe de 0 à être +1 :

$$\|x\|_1 = \sum_{i=1}^d |x_i| \leq \left(\sum_{i=1}^d \text{signe}(x_i)^2 \right)^{1/2} \left(\sum_{i=1}^d |x_i|^2 \right)^{1/2} = \sqrt{d} \|x\|_2$$

Ce qui signifie que la convergence en norme 2 implique la convergence en norme 1.

- b) Nous avons

$$\|x\|_2 = \left(\sum_{i=1}^d |x_i|^2 \right)^{1/2} \leq \left(\sum_{i=1}^d \left(\sum_{k=1}^d |x_k| \right)^2 \right)^{1/2} = \|x\|_1 \cdot \sqrt{d}$$

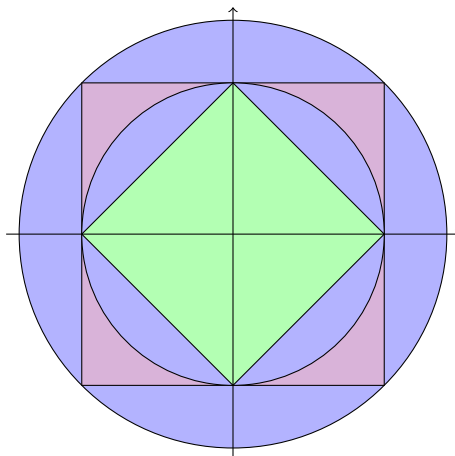
(au lieu de prendre chaque x_i , on prend la somme des coefficients à chaque fois, qui est plus grande que n'importe quel coefficient vu qu'on est en valeur absolue).

Ce qui montre que la convergence en norme 1 implique la convergence en norme 2.

En résumé, on a dans \mathbb{R}^d

$$\|x\|_1 \leq \sqrt{d} \|x\|_2 \leq d \|x\|_1,$$

Ces deux normes sont donc équivalentes. La norme 1 et la norme sup sont également équivalentes (exercice). L'équivalence des normes est visualisable :



On voit bien qu'on peut toujours entourer une boule par une autre. Le changement de rayon nécessaire est lié aux constantes de l'équivalence, voir 1.42.

Théorème. 1.44. Soit $\|\cdot\|, \|\cdot\|$ deux normes sur un e.v. E . On note $B_{\|\cdot\|}$ et $B_{\|\cdot\|}$ leurs boules unités fermées respectives. Alors les assertions suivantes sont équivalentes :

- a) $\|\cdot\|$ est plus faible que $\|\cdot\|$
- b) si $x_n \xrightarrow{\|\cdot\|} x$ alors $x_n \xrightarrow{\|\cdot\|} x$
- c) l'application identité $id : (E, \|\cdot\|) \rightarrow (E, \|\cdot\|)$ est continue
- d) $\exists \alpha > 0 B_{\|\cdot\|} \subseteq \alpha B_{\|\cdot\|}$
- e) $\inf\{\|x\| \text{ tel que } \|x\| = 1\} > 0$

Démonstration. a) \Rightarrow b) : vu juste avant.

b) \Leftrightarrow c) : définition séquentielle de la continuité.

c) \Rightarrow a) : Id est linéaire et continue donc bornée : $\sup_{x \neq 0} \frac{\|\text{Id}(x)\|}{\|x\|} \stackrel{\text{def.}}{=} C < \infty$ (le sup est fini). Ainsi $\|x\| \leq C \cdot \|x\|$ donc $\|\cdot\|$ est plus faible que $\|\cdot\|$

a) \Rightarrow d) : Soit $\|x\| \leq 1 \Rightarrow \|x\| \leq C \cdot \|x\| \leq C \Rightarrow x \in C \cdot B_{\|\cdot\|} = B_{\|\cdot\|}[0, C]$

d) \Rightarrow e) : Soit $\|x\| = 1 \Rightarrow x \neq 0$ et $\|x\| \neq 0$ i.e. $\|x\| > 0$ On observe que si $\lambda = \frac{1}{\|x\|}$, $\lambda \cdot x \in B_{\|\cdot\|} \subseteq \alpha \cdot B_{\|\cdot\|}$ par hypothèse. Cela signifie : $\|\lambda \cdot x\| \leq \alpha$ donc $\frac{\|x\|}{\|x\|} \leq \alpha$ et : $\|x\| \leq \alpha \cdot \|x\|$ or $\|x\| = 1$ donc $\|x\| \geq \frac{1}{\alpha}$ donc $\inf\{\|x\| \text{ tel que } \|x\| = 1\} > \frac{1}{\alpha} > 0$

e) \Rightarrow a) : Soit $\delta = \inf\{\|x\| \text{ tel que } \|x\| = 1\}$. Et soit $x \neq 0$ et $\lambda = \frac{1}{\|x\|}$. $\|\lambda \cdot x\| = |\lambda| \cdot \|x\| = 1$ donc $\|\lambda \cdot x\| \geq \delta$. Ce qui donne $|\lambda| \cdot \|x\| \geq \delta$ et $\frac{\|x\|}{\|x\|} \geq \delta \Rightarrow \|x\| \geq \delta \|x\|$ pour tout x différent de zéro. Pour $x = 0$ c'est trivialement vrai, donc c'est vrai pour tout x . \square

Théorème. 1.45. Soit E un e.v. de dimension finie. Alors toutes les normes sur E sont équivalentes. Cela signifie : Si N_1 et N_2 sont des normes sur E , il existe c et C tels que la double inégalité suivante est vraie :

$$c \cdot N_1(x) \leq N_2(x) \leq C \cdot N_1(x) \quad \forall x \in E$$

Démonstration. Soit $\{b_1, \dots, b_d\}$ une base de E et soit $N(\cdot)$ une norme. Tout $x \in E$ s'écrit de façon unique sous la forme : $x = \sum_{i=1}^d x_i b_i$ où les $x_i \in \mathbb{R}$. On pose $\|x\|_2 \stackrel{\text{def}}{=} (\sum_{i=1}^d |x_i|^2)^{1/2}$ où il est facile de vérifier que $\|\cdot\|_2$ est bien une norme. Observons que

$$|N(x) - N(y)| \leq N(x - y) = N\left(\sum_{i=1}^d (x_i - y_i) b_i\right) \leq \sum_{i=1}^d |x_i - y_i| N(b_i)$$

car $N(\cdot)$ est une norme, donc on peut appliquer l'inégalité triangulaire. D'après Cauchy-Schwarz on a l'inégalité :

$$\sum_{i=1}^d (|x_i - y_i| N(b_i)) \leq \|x - y\|_2 \underbrace{\left(\sum_{i=1}^d N(b_i)^2\right)^{1/2}}_{\stackrel{\text{def.}}{=} \alpha}$$

Donc $N(\cdot)$ est plus faible que $\|\cdot\|_2$ (prendre $y = 0$, par exemple!). Ainsi N est continue par rapport à $\|\cdot\|_2$.

Soit $K = \{x \in E : \|x\|_2 = 1\}$. K est borné et fermé pour la norme $\|\cdot\|_2$ (car $K = f^{-1}(\{1\})$) avec $f(x) = \|x\|_2$. Donc K est compact pour la norme $\|\cdot\|_2$. Or $N : K \rightarrow \mathbb{R}$ est continue

$\Rightarrow N(K) \subseteq \mathbb{R}$ est compact. Donc fermé et borné, il admet un minimum (une fonction continue sur un compact atteint toujours ses bornes) :

$$\exists x^* \in K, \|x^*\|_2 = \inf\{N(x), x \in K\} \quad \|x^*\| = 1 \Rightarrow x^* \neq 0 \Rightarrow \inf\{N(x), x \in K\} > 0.$$

(Tout ça pour dire que $N(\cdot)$ est strictement positive sur K , donc on va pouvoir la multiplier par une constante). Par le théorème 1.44 on en déduit qu'il existe une constante C telle que $\|x\|_2 \leq C.N(x)$ c'est à dire que $\|\cdot\|_2$ est plus faible que $N(\cdot)$. Par ailleurs, en montrant la continuité de $N(\cdot)$ on a trouvé α tel que $N(x) \leq \alpha.\|x\|_2$. On en conclut que $N(\cdot)$ et $\|\cdot\|_2$ sont des normes équivalentes.

On a donc vu que toute norme était équivalente à la norme $\|\cdot\|_2$ donc, si $N_i(\cdot)$, $i = 1, 2$ sont des normes on a : $c_i\|x\|_2 \leq N_i(x) \leq C_i\|x\|_2$ pour certains $c_i, C_i > 0$. Ceci donne, en "transitant" par la norme $\|\cdot\|_2$:

$$N_1(x) \leq C_1\|x\|_2 \leq \frac{C_1}{c_2} N_2(x) \leq \frac{C_1 C_2}{c_2} \|x\|_2 \leq \frac{C_1 C_2}{c_1 c_2} N_1(x)$$

Ainsi, dans un e.v. de dimension finie, toutes les normes sont équivalentes. □

Corollaire 1.46. Dans \mathbb{R}^d (ou tout e.v. de dimension finie), les notions "ouvert", "fermé" et "compact" sont indépendantes de la norme choisie.

Démonstration. (a) Si A est fermé $\Leftrightarrow A = \overline{A} = L = \{\lim(a_n) : (a_n)_n \in \mathbb{N}\}$. Cette caractérisation est indépendante de la norme. Par passage au complémentaire, on en déduit que "ouvert" ne dépend pas non plus de la norme.

(b) Autre méthode : Soit N_1 et N_2 deux normes et O N_1 -ouvert. $\forall x \in O, \exists r_1(x) > 0 : B_{N_1}(x, r_1(x)) \subset O \Rightarrow O = \bigcup_{x \in O} B_{N_1}(x, r_1(x))$. Cela signifie qu'un ouvert quelconque peut toujours s'écrire comme union de boules avec des rayons adaptés à chaque élément de O . On le montre par double inclusion :

\Rightarrow : si $x \in O$, alors il est dans $B_{N_1}(x, r_1(x))$ donc dans $\bigcup_{x \in O} B_{N_1}(x, r_1(x))$

\Leftarrow : si $x \in \bigcup_{x \in O} B_{N_1}(x, r_1(x))$ alors $x \in B_{N_1}(x, r_1(x)) \subset O$.

Mais $B_{N_1}(x, r_1(x)) \supset B_{N_2}(x, \alpha r_1(x))$ par équivalence des normes. On a donc

$$O = \bigcup_{x \in O} B_{N_1}(x, r_1(x)) \supseteq \bigcup_{x \in O} B_{N_2}(x, \alpha r_1(x)) \supseteq O$$

donc $O = \bigcup_{x \in O} B_{N_2}(x, r_1(x)) \Rightarrow O$ est N_2 -ouvert. Ce qui donne par passage au complémentaire A N_1 -fermé $\Leftrightarrow A$ N_2 -fermé.

(c) "compact" :

$$N_1\text{-compact} \Leftrightarrow \left\{ \begin{array}{ll} N_1\text{-fermé (ne dépend pas de la norme, donc)} & \Leftrightarrow N_2\text{-fermé} \\ N_1\text{-borné (ne dépend pas de la norme donc)} & \Leftrightarrow N_2\text{-borné} \end{array} \right\} \\ \Leftrightarrow N_2\text{-compact}$$

□

Corollaire 1.47. Soit E un e.v.n. de dimension finie.

- a) Toute suite de Cauchy converge
- b) Si X est un e.v.n. et $T : E \rightarrow X$ linéaire, alors T est continue
- c) Tout s.e.v. $F \subseteq E$ est fermé.

Démonstration. a) Toute suite de Cauchy converge :

Soit (x_n) de Cauchy pour $\|\cdot\| \Rightarrow (x_n)$ de Cauchy pour $\|\cdot\|_2$. Appelons $\pi_k(x_n), k \in [1, d]$ les coefficients de x_n dans la base (b_1, \dots, b_d) . Ainsi $(\pi_k(x_n))_{n \geq 1}$ est de Cauchy dans \mathbb{R} , ce qui signifie qu'elle converge vers une limite l_k . Donc $x_n \xrightarrow{\|\cdot\|_2} \sum_{k=1}^d l_k b_k \stackrel{\text{def.}}{=} x$. Par le théorème sur l'équivalence des normes on a : $x_n \xrightarrow{\|\cdot\|} x$

- b) Si x est un e.v.n. et $T : E \rightarrow X$ linéaire, alors T est continue :

Soit $T : E \rightarrow X$ linéaire. Et (b_1, \dots, b_d) une base de E . On a $T(x) = T(\sum_{k=1}^d x_k b_k)$ et donc

$$\|T(x)\|_X \leq \sum_{k=1}^d |x_k| \cdot \|b_k\|_X \stackrel{C.S.}{\leq} \|x\|_2 \left(\sum_{k=1}^d \|T(b_k)\|^2 \right)^{1/2} \leq \alpha \cdot C \cdot \|x\|_E$$

par équivalence des normes, avec $C = (\sum_{k=1}^d \|T(b_k)\|^2)^{1/2}$. Conclusion : $\|T(x)\|_X \leq \alpha \cdot C \cdot \|x\|_E \Rightarrow \|T(x_n - x)\|_X \leq \alpha \cdot C \cdot \|x_n - x\|_E$ i.e. la convergence de x_n vers x entraîne la convergence de $T(x_n)$ vers $T(x)$ donc T est continue.

- c) Tout s.e.v. $F \subseteq E$ est fermé :

Soit F un s.e.v. de dimension m de E . $\{f_1, \dots, f_m\}$ une base de F . On peut la compléter pour avoir une base de E : $\{f_1, \dots, f_m, f_{m+1}, \dots, f_n\}$. Alors on définit π :

$$\pi : \begin{cases} E \rightarrow E \\ \sum_{k=1}^d x_k f_k \mapsto \sum_{k=m+1}^d x_k f_k \end{cases}$$

(Ainsi π envoie un élément x de E sur ses coordonnées sans les coordonnées associées à la base de F , ce qui fait que $\pi(x) = 0$ si, et seulement si $x \in F$) Or par b) on sait que π est continue car elle est linéaire. On a donc : $F = \ker(\pi) = \pi^{-1}(\{0\})$. Or l'image réciproque d'un fermé par une application continue est un fermé, donc F est fermé. \square

1.7 Courbes paramétrées

Définition 1.48. Soit I un intervalle réel. Une fonction continue $f : I \rightarrow \mathbb{R}^d$ est appelée courbe paramétrée.

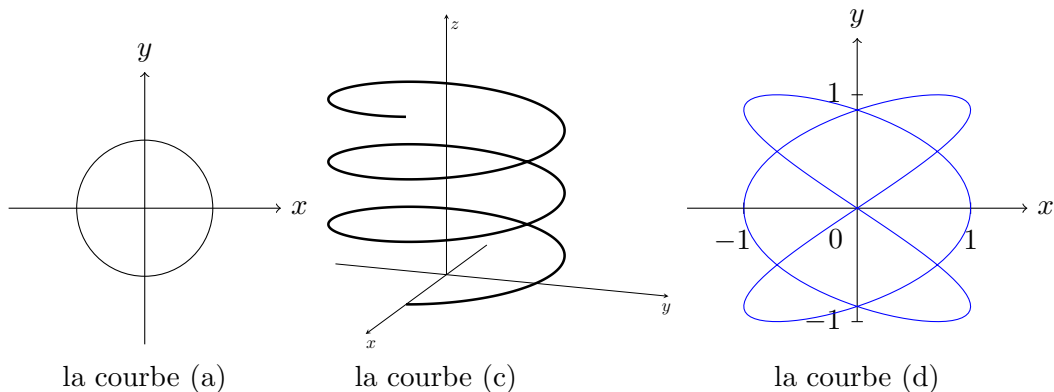
Rappel : $f : I \rightarrow \mathbb{R}^d$ est continue si et seulement si ses fonctions coordonnées le sont. Si $f(t) = (f_1(t), \dots, f_n(t))$ on teste la continuité de chaque f_i .

Exemple 1.49.

- a) $f(t) = (r \cos(t), r \sin(t)), t \in I = \mathbb{R}$. Cette courbe donne un cercle, évidemment.

- b) $f(t) = a + t.v; a, v \in \mathbb{R}^d$: à partir d'un point a on part dans une direction v en ligne droite (dans les deux sens : c'est une droite affine).
- c) $f(t) = (r \cos(t), r \sin(t), t)$, dans \mathbb{R}^3 Spirale avec courbure linéaire, comme une vis
- d) $f(t) = (\cos(3t), \sin(2t))$. Une courbe rigolotte.

Les voici :



Vitesse, vitesse absolue

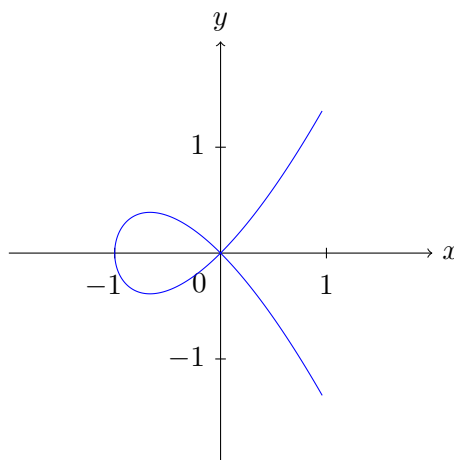
On définit $\frac{f(t+h)-f(t)}{h} = \frac{\text{déplacement}}{\text{temps}}$ comme approximation de la vitesse de la courbe (L'addition se fait coordonnée par coordonnée). La limite, quand $h \rightarrow 0$ est égal à $(f'_1(t), \dots, f'_n(t))$ si toutes les f_i sont dérivables.

On définit la vitesse absolue comme étant :

$$\|f'(t)\|_2 = \sqrt{f'_1(t)^2 + \dots + f'_n(t)^2}.$$

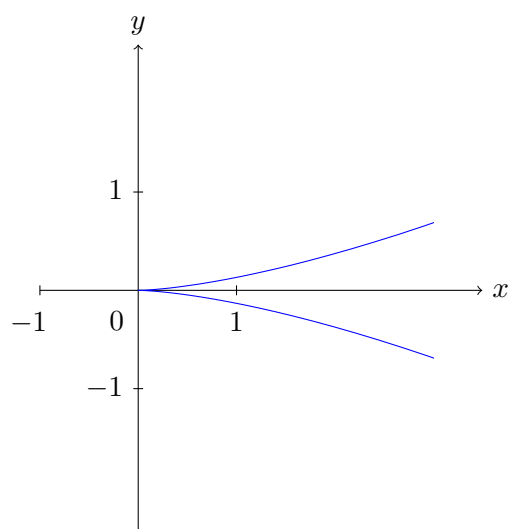
Ainsi la vitesse est donnée par la direction (dérivées coordonnée par coordonnée) et la longueur (la norme).

Rappel : Une courbe paramétrée n'est pas forcément injective : par exemple $f(t) = (t^2 - 1, t^3 - t)$ où $f(1) = f(-1)$:



Courbes régulières

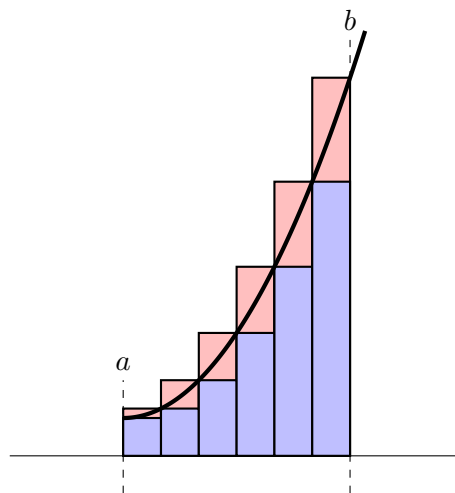
f est régulière (c'est à dire non-singulière) si : $\forall t \in I, f'(t) \neq 0$. Contre-exemple : $f(t) = (t^2, t^3)$ a sa dérivée nulle en 0 : $f'(0) = 0$. A cet endroit sa vitesse est nulle : pas direction, pas de longueur.



Courbes rectifiables

Définition 1.50. (Rappel : intégrale sur un segment)

Soit f définie sur un segment $[a, b]$, on souhaite connaître l'aire sous la courbe de f entre les segments a et b . On délimite de petits rectangles sous la courbe de f et on somme leurs aires, de même pour des rectangles au dessus de la courbe. On obtient une estimation supérieure et inférieure de l'aire. Si pour tout ε il existe un maillage telle que la différence entre ces deux estimations est plus petite que ε , on dit que f est Riemann intégrable. Ceci est le cas pour deux grandes classes de fonctions, les fonctions continues et les fonctions monotones.



Dans le cas d'une fonction Riemann intégrable, $\varepsilon > 0$ et un maillage convenable,

$$\sum_{j=1}^J f(t_j)(t_{j+1} - t_j) \approx \int_a^b f(t) dt$$

Ceci n'est rien que sommer les "blocs" en "hauteur \times largeur". On appliquera ce principe souvent par la suite.

Revenons aux courbes : Soit $f : [a, b] \rightarrow \mathbb{R}^d$ une courbe paramétrée. Une subdivision $\tau = (t_1, \dots, t_n)$ est un n -uplet de réels tels que $a = t_0 < t_1 < \dots < t_n = b$. Pour une subdivision τ on note $P_\tau(f) = \sum_{j=0}^{n-1} \|f(t_{j+1}) - f(t_j)\|_2$ - on remplace donc les sections de courbes par des bouts de droites affines.

Définition 1.51. Une courbe est rectifiable de longueur L si, et seulement si $\forall \varepsilon > 0, \exists \delta > 0$ tel que, pour tout τ satisfaisant $|t_{j+1} - t_j| < \delta$ pour tout j , on ait $|P_\tau(f) - L| < \varepsilon$.

Cela signifie que pour une erreur fixée, on peut toujours trouver une longueur de pas telle que la partition de f soit "presque égale" à L . C'est à dire que la courbe ne zigzague pas entre deux points aussi proches que l'on veut.

Théorème. 1.52. Si $f : [a, b] \rightarrow \mathbb{R}^d$ est de classe \mathcal{C}^1 alors elle est rectifiable et

$$L = \int_a^b \|f'(t)\|_2 dt$$

Explication intuitive : Vu que

$$\begin{aligned} P_\tau(f) &= \sum_{j=0}^{n-1} \|f(t_{j+1}) - f(t_j)\|_2 = \sum_{j=0}^{n-1} (t_{j+1} - t_j) \frac{\|f(t_{j+1}) - f(t_j)\|_2}{(t_{j+1} - t_j)} \\ &\approx \sum_{j=0}^{n-1} (t_{j+1} - t_j) \|f'(t_j)\|_2 \approx \int_a^b \|f'(t)\|_2 dt \end{aligned}$$

on s'attend à cet énoncé. Pour la preuve on note un

Lemme 1.53. Soit $f \in \mathcal{C}^1([a, b]; \mathbb{R}^d)$ et $\varepsilon > 0$. Alors $\exists \delta > 0$ tel que $\forall s, t \in [a, b] : |s - t| < \delta \Rightarrow \left\| \frac{f(s) - f(t)}{s - t} - f'(t) \right\| < \varepsilon$

Démonstration du lemme. a) cas en dimension 1

$f \in \mathcal{C}^1$ donc $f' \in \mathcal{C}^0$ i.e. f' continue sur $[a, b]$. Elle est donc uniformément continue § c'est à dire :

$$\forall \varepsilon > 0, \exists \delta > 0, \forall s, t \in [a, b] : (|s - t| < \delta \Rightarrow |f'(s) - f'(t)| < \varepsilon)$$

(δ dépend de ε mais il ne dépend ni de s ne de t). En combinant cela avec $\frac{f(s) - f(t)}{s - t} = f'(\xi)$ pour un certain ξ entre s et t , on obtient :

$$|\xi - t| \leq |s - t| < \delta \Rightarrow |f'(\xi) - f'(t)| = \left| \frac{f(s) - f(t)}{s - t} - f'(t) \right| < \varepsilon.$$

§. Rappel : Théorème : Soit g est une fonction continue sur $[a, b]$, alors elle est uniformément continue. Démonstration : Supposons par l'absurde que non. Cela signifie : $\exists \varepsilon > 0, \forall \delta > 0, \exists s, t$ tels que : $|s - t| < \delta$ ET $|g(t) - g(s)| > \varepsilon$. Prenons $\delta = \frac{1}{n}$, on obtient s_n et t_n tels que $|s_n - t_n| < \delta$ et $|g(t_n) - g(s_n)| > \varepsilon$ Or (s_n) et (t_n) sont deux suites dans $[a, b]$ qui est compact. Elles admettent donc des sous-suites convergentes. Soit φ une application strictement croissante de \mathbb{N} dans \mathbb{N} , telle que $(s_{\varphi(n)}) \rightarrow s^*$. On a aussi $(t_{\varphi(n)})$ qui est une suite de $[a, b]$, donc elle admet une sous suite convergente $(t_{\psi(\varphi(n))}) \rightarrow t^*$ Alors $(s_{\psi(\varphi(n))})$ converge aussi vers s^* . Par hypothèse $|s_{\psi(\varphi(n))} - t_{\psi(\varphi(n))}| < \frac{1}{n}$. Cela implique $s^* = t^*$. Or on a aussi par hypothèse $\varepsilon < |g(s_n) - g(t_n)| = |g(s_n) - g(s^*) + g(s^*) - g(t_n)| \leq |g(s_n) - g(s^*)| + |g(t_n) - g(t^*)| \Rightarrow g$ est discontinue en $s^* = t^*$. C'est absurde.

b) le cas général : $f = (f_1, \dots, f_n)$. On a

$$\left\| \frac{f(s) - f(t)}{s - t} - f'(t) \right\|_2 \leq \sqrt{n} \cdot \max\left\{ \left| \frac{f_i(s) - f_i(t)}{s - t} - f'_i(t) \right|, i \in \{1, \dots, n\} \right\}$$

Si $\delta = \min\{\delta_1, \dots, \delta_n\}$, ou δ est choisi comme dans le cas numéro 1, on a donc

$$\left\| \frac{f(s) - f(t)}{s - t} - f'(t) \right\|_2 \leq \sqrt{n} \cdot \varepsilon.$$

□

Démonstration du théorème. Soit $\varepsilon > 0$. Alors il existe $\delta_1 > 0$ tel que $\forall j \in [1, n]$:

$$|t_{j+1} - t_j| < \delta_1 \Rightarrow \left| \int_a^b \|f'(t)\|_2 dt - \sum_{j=1}^n \|f'(t_j)\|_2 (t_{j+1} - t_j) \right| < \frac{\varepsilon}{2}$$

(La courbe est Riemann-intégrable : on peut approcher l'intégrale de la courbe par la somme des rectangles $(t_{j+1} - t_j) * f'(t_j)$). Par le lemme : $\exists 0 < \delta_2 < \delta_1$ tel que :

$$|s - t| < \delta_2 \Rightarrow \left\| \frac{f(s) - f(t)}{s - t} - f'(t) \right\|_2 < \frac{\varepsilon}{2(b - a)}$$

En prenant $s = t_{j+1}$ et $t = t_j$ on obtient

$$\begin{aligned} \left| \left\| f(t_{j+1}) - f(t_j) \right\|_2 - \left\| (t_{j+1} - t_j) f'(t_j) \right\|_2 \right| &= (t_{j+1} - t_j) \left| \left\| \frac{f(t_{j+1}) - f(t_j)}{(t_{j+1} - t_j)} \right\|_2 - \|f'(t_j)\|_2 \right| \\ (\text{inégal. triang.}) &\leq (t_{j+1} - t_j) \left\| \frac{f(t_{j+1}) - f(t_j)}{(t_{j+1} - t_j)} - f'(t_j) \right\|_2 \\ &\leq \frac{(t_{j+1} - t_j)}{b - a} \cdot \frac{\varepsilon}{2} \end{aligned}$$

Maintenant sommions sur les j . Il suit que :

$$\begin{aligned} &\left| \sum_{j=0}^{n-1} \|f(t_{j+1}) - f(t_j)\|_2 - \int_a^b \|f'(s)\|_2 ds \right| \\ &\leq \left| \sum_{j=0}^{n-1} \|f(t_{j+1}) - f(t_j)\|_2 - \|(t_{j+1} - t_j) f'(t_j)\|_2 \right| + \left| \sum_{j=0}^{n-1} \|(t_{j+1} - t_j) f'(t_j)\|_2 - \int_a^b \|f'(s) ds\|_2 \right| \\ &\leq \sum_{j=0}^{n-1} \left(\frac{t_{j+1} - t_j}{b - a} \frac{\varepsilon}{2} \right) + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

Ceci montre que l'estimation de la longueur ne dépend pas de L , et que la longueur est explicite et égale à : $L = \int_a^b \|f'(x)\|_2 dx$ □

Exemple 1.54. $f(t) = (r \cos(t), r \sin(t))$. Or $f'(t) = (-r \sin(t), r \cos(t))$ et

$$L_\varphi = \int_0^\varphi \sqrt{r^2 \cos^2(t) + r^2 \sin^2(t)} dt = \int_0^\varphi r \cdot dt = r \cdot \varphi.$$

Donc : si $\varphi = 2\pi$ on trouve $L = 2\pi \cdot r$!! :)

1.8 Reparamétrisation

Soit $f : [a, b] \rightarrow \mathbb{R}^d$ une courbe paramétrée, et Φ une application de $[\alpha, \beta] \rightarrow [a, b]$ continue et bijective. $g \stackrel{\text{def.}}{=} f \circ \Phi$ est une courbe paramétrée de $[\alpha, \beta] \rightarrow \mathbb{R}^d$.

Définition 1.55. Si Φ et Φ^{-1} sont \mathcal{C}^1 , on appelle le passage de f à g une transformation de paramètres.

Observations :

- a) Il n'y a que deux cas possibles (car Φ est bijective) :
 - a) Soit Φ est strictement croissante, on dit alors que Φ préserve l'orientation,
 - b) Soit Φ est strictement décroissante, on dit alors que Φ inverse l'orientation.
- c) De plus : $\Phi^{-1} \circ \Phi = Id$. En dérivant : $(\Phi^{-1})'(\Phi) \cdot \Phi' = 1$. Donc Φ' et Φ'^{-1} sont de même signe : si Φ préserve l'orientation alors Φ^{-1} aussi, et de la même façon, si Φ inverse l'orientation alors Φ^{-1} aussi.
- d) Longueur de courbe et transformation de paramètres $g = f \circ \Phi = f(\Phi(t))$; $g'(t) = f'(\Phi(t))\Phi'(t)$ avec $\Phi'(t) \in \mathbb{R}$.
 $L(g) = \int_{\alpha}^{\beta} \|g'(s)\|_2 ds = \int_{\alpha}^{\beta} \|f'(\Phi(t))\Phi'(t)\|_2 dt = \varepsilon \int_{\alpha}^{\beta} \|f'(\Phi(t))\|_2 |\Phi'(t)| dt$ avec $\varepsilon = 1$ si Φ préserve et $\varepsilon = -1$ si Φ inverse.
Par changement de variable $\Phi(t) = x$ on obtient (qu'il y ait inversion ou pas) :

$$L(g) = \int_a^b \|f'(x)\|_2 dx$$

La longueur d'une courbe paramétrée (de classe \mathcal{C}^1), ne dépend pas du changement de paramètre.

2 Fonctions de plusieurs variables

2.1 Continuité dans \mathbb{R}^m

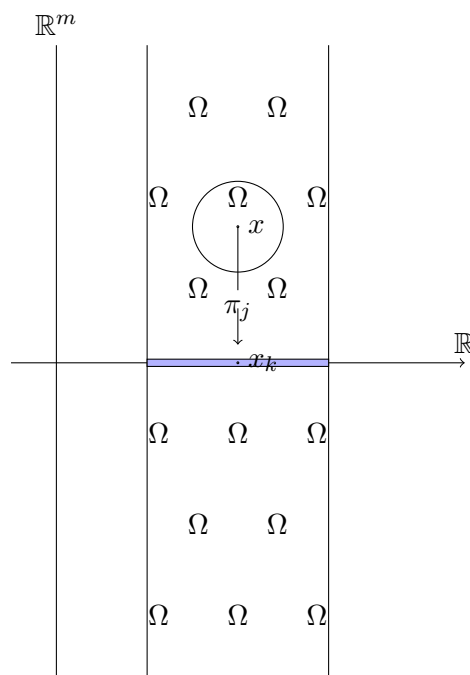
Rappel : Soit $\Omega \subset \mathbb{R}^m$, Ω ouvert, $f : \Omega \rightarrow \mathbb{R}^m$. Alors f est continue en x si : $\forall (x_n) \rightarrow x \in \Omega$ on a $f(x_n) \rightarrow f(x)$, (sachant que $f(x_n)$ est défini à partir d'un certain rang N , peu importe la suite qui converge vers x , car Ω ouvert).

Attention : $\forall x_n$ signifie "toute suite", il est donc absurde d'en prendre une particulière pour démontrer la véracité d'une assertion. On ne peut utiliser une suite "choisie" que pour infirmer une assertion.

Exemple 2.1. Les projections coordonnées sont les fonctions continues les plus importantes.

$$\pi_k \begin{cases} \mathbb{R}^m & \rightarrow \mathbb{R} \\ (x_1, \dots, x_m) & \mapsto x_k \end{cases}$$

Évidemment $|\pi_k(x)| \leq \|x\|_2$. Ainsi $x_n \rightarrow x$ implique $|\pi_k(x_n - x)| \rightarrow 0$, donc $\pi_k(x_n) \rightarrow \pi_k(x)$. Voilà une visualisation par image réciproque : Soit O un ouvert de \mathbb{R} . On note $\Omega = \pi_k^{-1}(O)$. Pour $x \in \Omega$, $\pi_k(x) \in O$ donc $(x-r, x+r) \subset O$. Il suit que $B(x, r) \subset \Omega$!



Remarques, rappels et astuces

- a) $f \rightarrow \mathbb{R}^d$; $f(x) = (f_1(x), \dots, f_m(x))$ où les f_k sont les fonctions coordonnées. Formellement : $f_k = \pi \circ f$. **Alors f est continue si, et seulement si f_k est continue** $\forall k \in \{1, \dots, m\}$. Preuve : " \Rightarrow " : soit $x_n \rightarrow x \in \Omega$. Alors si $f(x_n) \rightarrow f(x)$, $\pi_k \circ f(x_n) \rightarrow \pi_k \circ f(x)$, or $\pi_k \circ f(x_n) = f_k(x_n)$ et $\pi_k \circ f(x) = f_k(x)$. " \Leftarrow " : si $\pi_k \circ f(x_n) \rightarrow \pi_k \circ f(x)$, $\forall k \in \{1, \dots, m\}$, alors on $(f(x_n))_{n \geq 1}$ converge coordonnée par coordonnée, ce qui implique la convergence dans \mathbb{R}^m . Dans la pratique, cela signifie que l'on peut analyser chaque f_k et supposer $m = 1$.

- b) Rappel : Dans \mathbb{R} on sait que les sommes, différences, produits, quotients (si le dénominateur ne s'annule pas) et composés de fonctions continues sont continues.
 c) Exemple 1 : $x^2 + \arctan(\exp(y + \ln(1 + x^2)))$ est continue comme produit, somme et composés de fonctions continues.

- d) Exemple 2 : $f(x, y) = \begin{cases} \frac{x^2 y}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{si } (x, y) = (0, 0) \end{cases}$. Astuce : vu que le dénominateur est radial, passer en coordonnées polaires. Pour rappel, "ça fonctionne bien si les degrés "cumulés du numérateur sont supérieurs à ceux du dénominateurs".
 $g(r, \theta) = f(r \cos \theta, r \sin \theta) = \frac{r^3 \cos^2 \theta \sin \theta}{r^2} = r \cos^2 \theta \sin \theta$. On a $\cos^2 \theta \sin \theta$ borné par 1, donc $|g(r, \theta)| \leq G(r) = r$. et G tend vers zéro (quand r tend vers zéro...) donc $g(r, \theta) \rightarrow 0$.

On peut également passer en coordonnées sphériques dans \mathbb{R}^3 : $\begin{cases} x = r \cos \varphi \cos \theta \\ y = r \sin \varphi \cos \theta \\ z = r \sin \varphi \end{cases}$.

- e) Exemple 3 : $f(x, y) = \begin{cases} \frac{xy^2}{x^2 + y^4} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{si } (x, y) = (0, 0) \end{cases}$

Une des erreurs (logiques!) les plus fréquentes dans les examens est de raisonner la "lignes droites" : remplacer par exemple $x = ay$ (avec $a \neq 0$), ou, plus clairement, écrire $(x, y) = (at, t)$, $t \in \mathbb{R}$, est une façon d'approcher l'origine $(0, 0)$ le long d'une droite. On observe que

$$f(at, t) = \frac{a^2 t^3}{a^2 t^2 + t^4} = \frac{a^2 t}{a^2 + t^2} \rightarrow 0$$

quand $t \rightarrow 0$. Mais **cela n'implique pas que f est continue!** En effet, $f(t^2, t) = 1/2$ qui converge vers $1/2$ lorsque $t \rightarrow 0$. L'erreur est donc le même que de tester "une suite", où il fallait tester "toute suite" – remarque déjà faite en haut!

- f) Revoir aussi les techniques proposées dans l'Exemple 1.26.

2.2 Dérivées partielles

Définition 2.2. Soit $f : \Omega \subseteq \mathbb{R}^m \rightarrow \mathbb{R}$ est **partiellement dérivable** au point a en direction de e_k (vecteur de la base standard de \mathbb{R}^m i.e. $(0, \dots, 0, 1, 0, \dots, 0)$) si :

$$\lim_{h \rightarrow 0} \frac{f(a + h \cdot e_k) - f(a)}{h} = g'(a_k)$$

où a_k est la coordonnée de a en direction de e_k , et $g(t) = f(a_1, \dots, a_{k-1}, t, a_{k+1}, \dots, a_m)$.

En clair, on "gèle" toutes les coordonnées sauf a_k , et on dérive dans \mathbb{R} . On note $\frac{\partial f}{\partial x_k}(a)$ la dérivée partielle de f par rapport à x_k . Ou, encore plus court, disons pour $f(x, y)$ on écrit simplement $f_x(a)$ et $f_y(a)$ les deux dérivées partielles par rapport à x et y .

Exemples

- a) $f(x, y) = x \cdot \sin(y)$ On souhaite connaître sa dérivée par rapport à x en $(1, 0)$.
Formellement :

$$\frac{f(1+h, 0) - f(1, 0)}{h} = \frac{(1+h) \sin 0 - 1 \cdot \sin 0}{h} \xrightarrow{h \rightarrow 0} 0$$

Par le calcul habituel on a : $\frac{\partial f}{\partial x}(1, 0) = \sin(y)|_{x=1, y=0} = 0$.

Dérivée partielle par rapport à y : $\frac{\partial f}{\partial y}(1, 0) = x \cos(y)|_{x=1, y=0} = 1$. Et

$$\lim_{h \rightarrow 0} \frac{f(1, 0+h) - f(1, 0)}{h} = \frac{(1) \sin(h) - 1 \cdot \sin 0}{h} = \frac{\sin(h)}{h} \xrightarrow{h \rightarrow 0} 1$$

- b) $f(x, y) = e^{-x^2-y^2}$ On veut évaluer les dérivées en $(0, 0)$
On a $\frac{f(0+h, 0) - f(0, 0)}{h} = \frac{e^{-h^2} - 1}{h} \xrightarrow{h \rightarrow 0} (e^{-t^2})'(0) = 0$. On a reconnu la définition de la dérivée de e^{-t^2} en 0. On peut aussi utiliser le D.L. d'exp pour calculer la limite.

- c) $r(x) = \|x\|_2 = \sqrt{x_1^2 + \dots + x_m^2}$ On va calculer sa dérivée par rapport à la k -ième coordonnée. On suppose $x \neq 0$ Soit $g(t) = \sqrt{x_1^2 + \dots + x_{k-1}^2 + t^2 + x_{k+1}^2 + \dots + x_m^2}$.
Alors

$$g'(t) = \frac{2t}{2\sqrt{x_1^2 + \dots + x_{k-1}^2 + t^2 + x_{k+1}^2 + \dots + x_m^2}}$$

Ainsi : $\frac{\partial r}{\partial x_k} = g'(x_k) = \frac{x_k}{r(x)}$. Ceci est valable pour tout k : on a calculé toutes les dérivées partielles (en tout point non nul).

On peut noter que l'exemple précédent se traite de la même façon en remarquant :
 $e^{-x^2-y^2} = \varphi(r(x))$, avec $\varphi(t) = e^{-t^2}$

- d) $g(x_1, \dots, x_d) = \begin{cases} \frac{x_1 \cdot x_2 \cdot \dots \cdot x_d}{r(x)^d} & \text{si } x \neq 0 \\ 0 & \text{sinon} \end{cases}$ Regardons ce qui se passe près de $0_{\mathbb{R}^m}$:

$\frac{\partial g}{\partial x_k}(0, \dots, 0) = \lim_{h \rightarrow 0} \frac{g(0+h \cdot e_k) - g(0)}{h} = 0$ (Comme le numérateur de g est le produit de toutes les coordonnées et que là, elles sont toutes nulles sauf la k -ième, le produit est nul). Donc les dérivées partielles en zéro existent pour tout k . Cela dit : $g(\frac{1}{k}, \frac{1}{k}, \dots, \frac{1}{k}) = \frac{k^{-d}}{(\sqrt{d} \frac{1}{k})^d} = \frac{1}{d^{\frac{d}{2}}}$. Or d est fixe (la dimension) donc ceci est une constante strictement positive, elle ne tend donc pas vers 0, alors que : $g(0, \dots, 0) = 0$. Ce qui signifie que g n'est pas continue en zéro, bien qu'elle y admettent toutes ses dérivées partielles !

Si f est partiellement dérivable en tout point, on dit que f est partiellement dérivable. On appelle alors dérivée partielle la fonction qui associe $a \in \Omega \subset \mathbb{R}^m$ à $\frac{\partial f}{\partial x_k}(a) \in \mathbb{R}$.

Définition 2.3. Pour une fonction $f : \Omega \rightarrow \mathbb{R}$ on appelle **gradient** de f , parfois noté ∇f le vecteur des dérivées partielles de f :

$$\text{grad}(f) = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_d} \right)$$

Exemple : $\text{grad}(r(x)) = \frac{1}{r(x)} \cdot x = \frac{x}{r(x)}$ où $x = (x_1, \dots, x_d)$.

Plus généralement, soit $f : \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{pmatrix}$ Alors on écrit la **matrice Jacobienne** * comme suit :

$$\mathcal{J}_f(a) = \begin{pmatrix} \partial_{x_1} f_1 & \cdots & \partial_{x_n} f_1 \\ \vdots & & \vdots \\ \partial_{x_1} f_m & \cdots & \partial_{x_n} f_m \end{pmatrix}$$

2.3 Dérivées partielles supérieures

On peut définir de proche en proche les **dérivées supérieures** comme suit :

$$\frac{\partial}{\partial x_k} \left(\frac{\partial}{\partial x_j} f \right)$$

Exemple 2.4. Soit $f(x, y) = x^2 y^3$. Alors $\frac{\partial f}{\partial x} = 2xy^3$ et $\frac{\partial f}{\partial y} = 3x^2 y^2$. Puis $\frac{\partial}{\partial x} \frac{\partial f}{\partial y} = 6xy^2$ et $\frac{\partial}{\partial y} \frac{\partial f}{\partial x} = 6xy^2$. On peut aussi calculer : $\frac{\partial}{\partial x} \frac{\partial f}{\partial x} = 2y^3$ et $\frac{\partial}{\partial y} \frac{\partial f}{\partial y} = 6x^2 y$. On remarque que : $\frac{\partial}{\partial y} \frac{\partial f}{\partial x} = \frac{\partial}{\partial x} \frac{\partial f}{\partial y} = 6xy^2$. Ce n'est pas un hasard !

Théorème. 2.5. (Théorème de Schwarz †) Soit $\Omega \subseteq \mathbb{R}^m$ et $f : \Omega \rightarrow \mathbb{R}$, deux fois partiellement dérivable, et un point $a \in \Omega$ où $\frac{\partial}{\partial x_k} \frac{\partial f}{\partial x_j}$ et $\frac{\partial}{\partial x_j} \frac{\partial f}{\partial x_k}$ sont continues. Alors

$$\frac{\partial}{\partial x_k} \frac{\partial f}{\partial x_j}(a) = \frac{\partial}{\partial x_j} \frac{\partial f}{\partial x_k}(a) \stackrel{\substack{\text{notation} \\ \text{def.}}}{=} \frac{\partial^2}{\partial x_k \partial x_j} f(a)$$

La notation se justifie du fait que l'ordre de dérivation n'importe pas.

Démonstration. On peut supposer que $n=2$ et que $a = 0_{\mathbb{R}^2}$. En effet, il suffit de prendre $g(s, t) = f(a_1, \dots, a_{j-1}, a_j + s, a_{j+1}, \dots, a_{k-1}, a_k + t, a_{k+1}, \dots, a_n)$. Par l'hypothèse Ω ouvert (non vide!) on a : $\exists \delta_1 > 0$ tel que $[-\delta_1, \delta_1] \times [-\delta_1, \delta_1] \subset \Omega$. Par la continuité des dérivées mixtes en $(0, 0)$ on a : Pour un $\varepsilon > 0$, $\exists \delta > 0$, $\delta < \delta_1$, tel que

$$\begin{aligned} \left| \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right)(x, y) - \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right)(0, 0) \right| &\leq \varepsilon \\ \left| \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right)(x, y) - \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right)(0, 0) \right| &\leq \varepsilon \end{aligned}$$

*. d'après Carl Gustav Jacob Jacobi (né le 10 décembre 1804 à Potsdam, mort le 18 février 1851 à Berlin), est un mathématicien allemand. Il était professeur à Königsberg (aujourd'hui : Kaliningrad)

†. d'après Hermann Amandus Schwarz, le même que vous connaissez déjà de l'inégalité de Cauchy-Schwarz

dès que $|x|, |y| < \delta$. Regardons $h(x, y) = f(x, y) - f(x, 0) - (f(0, y) - f(0, 0))$: par accroissements finis, on a

$$\begin{aligned} h(x, y) &= x \frac{\partial}{\partial x} (f(x, y) - f(x, 0))(\xi_1) && \text{avec } \xi_1 \text{ entre } 0 \text{ et } x \\ &= xy \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y}(x, \eta_1) \right)(\xi) && 0 < \eta_1 < y \end{aligned}$$

De la même façon on obtient :

$$\begin{aligned} h(x, y) &= y \frac{\partial}{\partial y} (f(x, y) - f(0, y))(\xi_2) && \text{avec } \xi_2 \text{ entre } 0 \text{ et } y \\ &= xy \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x}(\eta_2, y) \right)(\xi_2) && 0 < \eta_2 < x \end{aligned}$$

On en déduit que $\frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y}(\eta_1, \xi_1) \right) = \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x}(\eta_2, \xi_2) \right)$. C'est à dire qu'on a trouvé deux points dans la boule autour de (x, y) tels que leurs dérivées partielles mixtes soient égales. Ceci pour tout $(x, y) \neq (0, 0)$. On en déduit par l'hypothèse de continuité que :

$$\begin{aligned} & \left| \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right)(0, 0) - \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right)(0, 0) \right| \\ & \leq \left| \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right)(0, 0) - \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right)(\eta_1, \xi_1) \right| + \left| \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right)(\eta_2, \xi_2) - \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right)(0, 0) \right| \\ & \leq 2\varepsilon \end{aligned}$$

□

Exercice 2.6. Soit

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

Montrer que les dérivées partielles secondes en $(0, 0)$ sont différentes. Conclure.

Corollaire 2.7. Soit $f : \Omega \subseteq \mathbb{R}^m \rightarrow \mathbb{R}$, k -fois partiellement dérivable, avec les dérivées partielles continues sur Ω . Alors, pour tout j_1, \dots, j_k et toute permutation σ :

$$\frac{\partial}{\partial x_{j_1}} \dots \frac{\partial}{\partial x_{j_k}} f = \frac{\partial}{\partial x_{\sigma(j_1)}} \dots \frac{\partial}{\partial x_{\sigma(j_k)}} f$$

Les dérivées partielles commutent donc.

Notations

a) On note $\partial_{x_k} = \frac{\partial}{\partial x_k} f = f_{k_k}(x)$ où x est un vecteur de \mathbb{R}^m

- b) On appelle $\alpha \in \mathbb{N}^d$ un multi-indice, tel que $|\alpha| = \alpha_1 + \dots + \alpha_d$ (ici, $|\alpha|$ est l'ordre total de la dérivée)
 $\partial^\alpha f = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \dots \frac{\partial^{\alpha_d}}{\partial x_d^{\alpha_d}}(f)$ avec f $|\alpha|$ fois partiellement dérivable. Cela signifie que l'on dérive α_i fois en direction de $x_i^{\alpha_i}$. Si $h = (h_1, \dots, h_d) \in \mathbb{R}^d$ on note aussi $h^{(\alpha)} = h_1^{\alpha_1} + h_2^{\alpha_2} + \dots + h_d^{\alpha_d}$
- c) Si f est $(|\alpha| + |\beta|)$ fois partiellement dérivable, avec les dérivées partielles continues alors on écrit : $\partial^{\alpha+\beta} = \partial^\alpha \partial^\beta f = \partial^\beta \partial^\alpha f$

2.4 La différentielle

Rappel du concept de la dérivée :

$$f'(a) = \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}$$

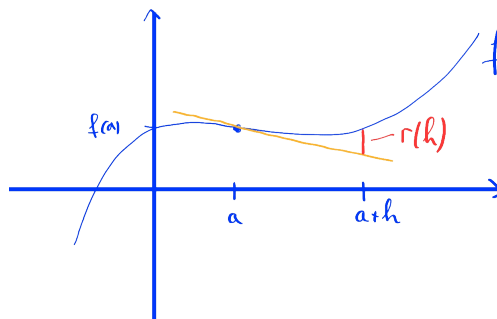
Pourquoi cette définition ? On veut parler d'une tangente bien définie. Soit

$$r(h) = f(a+h) - (f(a) + h \cdot f'(a))$$

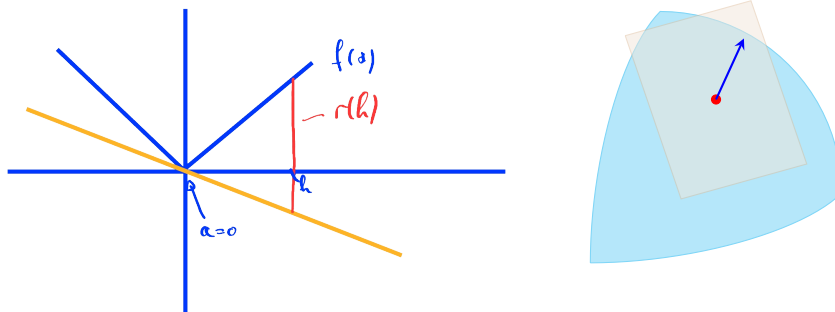
Ici, $r(h)$ est l'erreur de la "linéarisation" de f , c'est à dire, la différence entre $f(a+h)$ et la valeur "estimé" par une application affine, ici $f(a) + h \cdot m$ avec $m = f'(a)$. Dire que la limite existe est équivalent à dire que $\frac{r(h)}{h} \xrightarrow{h \rightarrow 0} 0$, c'est à dire $r(h) = o(h)$.

Si on relâchait cette condition en demandant seulement $r(h) = O(h)$, i.e. $\exists C > 0 : r(h) \leq Ch$, la tangente n'a plus de "pente" unique : on pourrait mettre "une partie de l'erreur" dans la "partie linéaire", ce qui fait que la partie linéaire n'est plus bien définie.

Regardons cela avec un deuxième exemple, $f(x) = |x|$ en $a = 0$: Ici, pour n'importe quelle pente m , i.e. n'importe quelle droite passant par $(0, 0)$, $r(h) = \alpha \cdot h$!!! Cette situation



n'est pas voulue.



Dérivabilité en a signifie qu'il existe *une* tangente bien-définie à la courbe au point $(a, f(a))$. Elle est donnée par

$$\{f(a) + f'(a)h : h \in \mathbb{R}\}$$

Ce qui nous importe est la structure *linéaire* de l'application $h \mapsto f'(a)h$, le reste dans l'expression ci-dessus est une translation. C'est pourquoi on parle d'une linéarisation. L'équivalent en plusieurs dimensions d'une droite tangente passant par $f(a)$ est un sous-espace affine passant par $f(a)$. Voir image en haut à droite. Un tel sous espace est nécessairement de la forme

$$\{f(a) + L(h) : h \in \mathbb{R}^d\}$$

où $L : \mathbb{R}^d \rightarrow \mathbb{R}^d$ est linéaire.

Définition 2.8. Soit $\Omega \subset \mathbb{R}^d$ ouvert, $a \in \Omega$, $f : \Omega \rightarrow \mathbb{R}^m$. On dit que f est **différentiable** en a , et on le note $D_f(a)$, s'il existe $L \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^m)$ telle que :

$$f(a+h) = f(a) + L(h) + r(h) \quad \text{où} \quad \frac{\|r(h)\|}{\|h\|} \xrightarrow{h \rightarrow 0} 0.$$

Dans ce cas on pose $D_f(a) \stackrel{\text{def.}}{=} L$.

▮ A noter qu'on peut utiliser n'importe quelle norme pour tester $r(h)$ puisqu'on est en dimension finie.

Lemme 2.9 (Unicité de la différentielle). *Soit f différentiable en a et L une application linéaire telle que $f(a+h) = f(a) + L(h) + r(h)$. Alors $\mathbf{L}(v) = \lim_{t \rightarrow 0} \frac{f(a+tv) - f(a)}{t}$. Ainsi, si L existe, elle est unique!*

Démonstration. Pour tout scalaire t , $f(a+tv) = f(a) + L(tv) + r(tv)$. Or L est linéaire on a donc $f(a+tv) = f(a) + tL(v) + r(tv)$. Ainsi, $\frac{f(a+tv) - f(a)}{t} = L(v) + \frac{r(tv)}{t}$. D'autre part $\frac{\|r(tv)\|}{\|t\|} = \frac{\|r(tv)\|}{\|tv\|} \|v\|$. Où $\|v\|$ est une constante, et $\frac{\|r(tv)\|}{\|tv\|} \rightarrow 0$ quand $t \rightarrow 0$ par l'hypothèse que f différentiable en $x = a$. Donc $\frac{r(tv)}{t} \xrightarrow{t \rightarrow 0} 0$. La limite existe alors. \square

Par exemple si on prend $v = e_k$, on obtient $L(e_k) = \partial_{x_k} f(a)$, c'est à dire la dérivée partielle de f sur x_k . Ainsi $L(v) = L(\sum v_k e_k) = \sum (v_k \partial_{x_k} f(a))$. Si L existe, c'est une combinaison linéaire des dérivées partielles. On a alors :

$$L(v) = \mathcal{J}_f(a) \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix}$$

C'est à dire : **Le seul "candidat" pour la différentielle est l'application**

$$L : h \mapsto \mathcal{J}_f(a).h$$

On dit "candidat" car l'existence de la Jacobienne ne garantit pas la "bonne" décroissance du reste : pour différentiabilité il faut que $\|r(h)\| = o(\|h\|)$. Voir l'exemple dans la Remarque 2.15 en bas. Un cas particulier à mentionner est $d = m = 1$: ici $L(v) = f'(a).v$, car la Jacobienne dans une dimension est juste la dérivée.

Exemple 2.10. a) La fonction constante : $f(x) = C \in \mathbb{R}^m$: Ici $f(a+h) = f(a) + 0 + 0$ où le premier 0 = $L(h)$ et le second 0 = $r(h)$. Visiblement $D_f(a) = L(h) = 0$ (en effet $\frac{\|r(h)\|}{\|h\|} = 0 \rightarrow 0$ quand $h \rightarrow 0$)

b) L'identité $f(x) = x : f(x+h) = x+h = f(x) + Id(h) + 0$. Donc $D_f(x) = Id$

c) Plus généralement, une application linéaire : $f(x) = A.x$, où A est une matrice. On a $f(x+h) = Ax + Ah = f(x) + L(h) + 0$ avec $L(h) = A.h$ Donc $D_f(a) = L : h \mapsto A.h$. L'erreur $r(h)$ est nulle, donc négligeable devant h .

d) Un cas particulièrement intéressant est, à nouveau, $\pi_k : \mathbb{R}^d \rightarrow \mathbb{R}$, $\pi_k(x_1, \dots, x_d) = x_k$. Cette application est linéaire, donc différentiable.

e) On pose $\|(h, k)\|_2 = \sqrt{\|h\|^2 + \|k\|^2}$ sur $\mathbb{R}^k \times \mathbb{R}^k$. Soit $f : \begin{cases} \mathbb{R}^k \times \mathbb{R}^k \\ (x, y) \longrightarrow \langle x, y \rangle \end{cases}$ (ou plus généralement, soit f une application bilinéaire). Alors

$$\begin{aligned} f(x+h, y+k) &= \langle x+h | y+k \rangle \\ &= \langle x | y \rangle + \langle x | k \rangle + \langle h | y \rangle + \langle h | k \rangle \\ &= f(x, y) + \underbrace{f(x, k) + f(h, y)}_{\text{partie linéaire en } (h, k)} + r(h, k) \end{aligned}$$

où $|r(h, k)| \leq \|h\| \cdot \|k\|$ (Cauchy-Schwarz). Rappelons qu'on a toujours $ab \leq \frac{a^2+b^2}{2}$. On l'applique deux fois :

$$\frac{|r(h, k)|}{\|(h, k)\|_2} \leq \frac{\|h\| \|k\|}{\sqrt{\|h\|^2 + \|k\|^2}} \leq \frac{1}{2} \sqrt{\|h\| \cdot \|k\|} \leq \frac{1}{\sqrt{8}} \|(h, k)\|_2$$

Cette fraction tend donc vers 0 si $(h, k) \rightarrow (0, 0)$. Ce qui donne $[D_f(x, y)](h, k) = \langle x | k \rangle + \langle h | y \rangle$.

2.5 Propriétés de la différentielle

Proposition 2.11. Soit $\Omega \subseteq \mathbb{R}^d$ ouvert, $f, g : \Omega \rightarrow \mathbb{R}^m$.

- a) Si f est différentiable en $x = a$, elle y est continue
 b) f, g différentiable en $a \Rightarrow (f + g)$ aussi et :

$$D_{f+g}(a) = D_f(a) + D_g(a)$$

- c) si f différentiable en $x = a$ et $L : \mathbb{R}^m \rightarrow \mathbb{R}^k$ linéaire, alors $L \circ f : \Omega \rightarrow \mathbb{R}^k$ différentiable et :

$$D_{L \circ f}(a) = L \circ D_f(a)$$

- d) f différentiable en $x = a$ si et seulement si $\forall j \in \{1, \dots, m\} : (\pi_j \circ f)$ différentiable.,
 où, comme toujours,

$$\pi_j : \begin{cases} \mathbb{R}^m & \rightarrow \mathbb{R} \\ (x_1, \dots, x_m) & \mapsto x_j \end{cases}$$

est la projection coordonnée – on écrit donc simplement $f = (f_1 \dots f_m)$.

Démonstration. Rappel : f différentiable en a signifie

$$f(a+h) = f(a) + L(h) + r(h) \quad \text{avec} \quad \frac{\|r(h)\|}{\|h\|} \xrightarrow{h \rightarrow 0} 0$$

- a) (Différentiable implique continue) Or $\|L\| \leq \|L\| \cdot \|h\|$ on a $L(h) \xrightarrow{h \rightarrow 0} 0$. Ainsi, $f(a+h) \rightarrow f(a)$ quand $h \rightarrow 0$, donc f est continue en a .
 b) (Sommes de fonctions différentiables)

$$\begin{aligned} f(a+h) + g(a+h) &= (f(a) + L_f(h) + r_f(h)) + (g(a) + L_g(h) + r_g(h)) \\ &= (f+g)(a) + (L_f + L_g)(h) + (r_f + r_g)(h) \end{aligned}$$

où $(L_f + L_g)$ est linéaire et $(r_f + r_g)$ tend vers 0 quand h tend vers 0. Donc :
 $D_{f+g}(a) = L_f + L_g = D_f(a) + D_g(a)$.

- c) (Composition avec une application linéaire).

$$(L \circ f)(a+h) = L(f(a+h)) = L(f(a) + D_f(a) \cdot h + r(h)) = L \circ f(a) + L \circ D_f(a) \cdot h + L \circ r(h)$$

On a d'une part $L \circ D_f(a)$ qui est linéaire, et d'autre part $\frac{\|L(r(h))\|}{\|h\|} \leq \frac{\|L\| \cdot \|r(h)\|}{\|h\|}$
 or $\frac{\|r(h)\|}{\|h\|}$ tend vers 0 et $\|L\|$ est une constante, donc le produit tend vers 0. D'où :
 $D_{L \circ f}(a) = L \circ D_f(a)$

- d) (Différentiabilité via les "fonctions coordonnées") \Rightarrow Or π_i est linéaire, d'après (c),
 $\pi_i \circ f$ est différentiable en a .

\Leftarrow Soit $l_i \begin{cases} \mathbb{R} \rightarrow \mathbb{R}^m \\ t \rightarrow t \cdot e_i \end{cases}$ avec l_i linéaire et (e_i) vecteur de la base canonique de \mathbb{R}^m .

On a :

$$f(a) = \sum_{i=1}^m f_i(a) \cdot e_i = \sum_{i=1}^m l_i(f_i(a)) = \sum_{i=1}^m l_i \circ f_i(a).$$

Mais comme l_i est linéaire, et f_i différentiable en a d'après (c), $l_i \circ f_i$ différentiable en a . En utilisant (b), leur somme $\sum_{i=1}^m l_i \circ f_i(a)$ différentiable en a . □

Théorème. 2.12. Soit $\Omega \subseteq \mathbb{R}^d$, un ouvert. Les propositions suivantes sont équivalentes :

- a) f a ses dérivées partielles continues
- b) f est différentiable et l'application

$$D_f : \begin{cases} \Omega & \rightarrow \mathcal{L}(\mathbb{R}^d, \mathbb{R}^m) \\ a & \mapsto D_f(a) \end{cases}$$

est continue. Dans ce cas on dit (c'est la définition !) que $f \in \mathcal{C}^1(\Omega, \mathbb{R}^m)$

Dans cet énoncé l'implication (a) \Rightarrow (b) est importante : si les fonctions coordonnées de la Jacobienne sont continues, alors f est différentiable, et la différentielle, donnée par

$$D_f(a)h = \mathcal{J}_f(a).h,$$

est continue. A combiner avec les propriétés de fonctions différentiables, et des compositions (voir en bas).

Démonstration. (b) \Rightarrow (a) : On suppose que la différentielle existe et qu'elle est continue. D_f est représentée par \mathcal{J}_f si elle existe. C'est le cas, et elle est continue ; la Jacobienne est continue coordonnée par coordonnée, c'est à dire que chacune de ses dérivées partielles est continue.

(a) \Rightarrow (b) : On suppose que f a ses dérivées partielles continues. Il suffit de montrer que f est différentiable en tout point. Pour cela on peut supposer $m = 1$, c'est à dire, on va travailler coordonnée par coordonnée. Soit

$$\varphi(h) = \frac{1}{\|h\|_\infty} \left(f(a+h) - f(a) - \sum_{j=1}^n \frac{\partial f}{\partial x_j}(a).h_j \right)$$

où $L(h) = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(a).h_j = \nabla(f).h$ est le candidat naturel de la différentielle. On veut montrer que $\varphi(h)$ tend vers 0 quand h tend vers 0. A cet effet, soit $\varepsilon > 0$. Par hypothèse, $\exists \delta$ tel que :

$$\|h\|_\infty < \delta \Rightarrow \max \left\{ \left| \frac{\partial f}{\partial x_j}(a-h) - \frac{\partial f}{\partial x_j}(a) \right| : j = 1, \dots, n \right\} < \frac{\varepsilon}{n}. \quad (2.1)$$

On définit $x^{(0)} = a$ et $x^{(m)} = a + \sum_{j=1}^m h_j \cdot e_j$; donc

$$\begin{aligned} x^{(0)} &= a \\ x^{(1)} &= a + h_1 e_1 \\ x^{(2)} &= a + h_1 e_1 + h_2 e_2 \\ &\vdots \\ x^{(d)} &= a + h_1 e_1 + h_2 e_2 + \dots + h_d e_d = a + h. \end{aligned}$$

(on avance en crabe en suivant les directions e_i). Comme f est continûment partiellement dérivable, on sait par le TAF entre $x^{(m-1)}$ et $x^{(m)} = x^{(m-1)} + h_m e_m$ qu'il existe $0 \leq \xi_m \leq h_m \leq \delta$ avec

$$f(x^{(m)}) - f(x^{(m-1)}) = h_m \frac{\partial f}{\partial x_m}(x^{(m-1)} + \xi_m e_m)$$

On reconnaît une somme télescopique :

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

Ainsi, en remplaçant dans l'expression de φ et en prenant la valeur absolue on obtient :

$$\begin{aligned} |\varphi(h)| &= \frac{1}{\|h\|_\infty} \left| \sum_{j=1}^n (f(x^{(j)}) - f(x^{(j-1)})) + \sum_{j=1}^n \frac{\partial f}{\partial x_j}(a) \cdot h_j \right| \\ &= \frac{1}{\|h\|_\infty} \left| \sum_{j=1}^n (h_j \frac{\partial f}{\partial x_j}(x^{(j-1)} + \xi_j e_j) - \frac{\partial f}{\partial x_j}(a) \cdot h_j) \right| \\ &= \frac{1}{\|h\|_\infty} \left| \sum_{j=1}^n h_j \left[\frac{\partial f}{\partial x_m}(x^{(m-1)} + \xi_m h_m) - \frac{\partial f}{\partial x_m}(a) \right] \right| \end{aligned}$$

Et enfin, d'après l'hypothèse (2.1), on achève la preuve :

$$|\varphi(h)| \leq \frac{1}{\|h\|_\infty} \sum_{j=1}^n \frac{\varepsilon}{n} |h_j| \leq \frac{1}{\|h\|_\infty} \sum_{j=1}^n \frac{\varepsilon}{n} \|h\|_\infty = \varepsilon \quad ! \quad \square$$

Théorème. 2.13. (Composition et différentielle) Soit $\Omega \subseteq \mathbb{R}^d$ ouvert, $a \in \Omega$; $f : \Omega \rightarrow \mathbb{R}^m$. Soit $\mathcal{U} \subseteq \mathbb{R}^m$ ouvert, $b \stackrel{\text{def.}}{=} f(a) \in \mathcal{U}$ et $g : \mathcal{U} \rightarrow \mathbb{R}^k$.

Si f est différentiable en a et g différentiable en $f(a)$ alors $g \circ f$ est différentiable en a et :

$$D_{g \circ f}(a) = D_g(f(a)) \circ D_f(a)$$

autre écriture : $D_a(f \circ g) = D_{g(a)}f \circ D_a g$

Démonstration. On note $f(a) = b$; $A = D_f(a)$, donc $f(a+h) = f(a) + A.h + r_1(h)$; $B = D_b(g)$ donc $g(b+k) = g(b) + B.k + r_2(k)$

On pose $k = A.h + r_1(h)$. On compose g avec f et développe :

$$\begin{aligned} (g \circ f)(a+h) &= g(f(a+h)) = g(f(a) + A.h + r_1(h)) \\ &= g(b) + B.k + r_2(k) \\ &= g(b) + B(A.h + r_1(h)) + r_2(k) \\ &= g(b) + B.A.h + B.r_1(h) + r_2(k) \end{aligned}$$

La dedans on a $B.A = D_g(f(a)) \circ D_f(a)$ qui est linéaire et un reste : $B.r_1(h) + r_2(k)$, on va montrer qu'il tend vers zéro quand h tend vers zéro.

Premièrement : $\frac{\|B.r_1(h)\|}{\|h\|} \leq \|B\| \frac{\|r_1(h)\|}{\|h\|}$: or $\frac{\|r_1(h)\|}{\|h\|} \rightarrow 0$ et $\|B\|$ est une constante finie, le tout tend vers zéro.

Deuxième partie : $r_2(k) = r_2(A.h + r_1(h))$. Ainsi

$$\frac{\|r_2(k)\|}{\|h\|} = \frac{\|r_2(A.h + r_1(h))\|}{\|A.h + r_1(h)\|} \cdot \frac{\|A.h + r_1(h)\|}{\|h\|}$$

Lorsque $h \rightarrow 0$, $A.h + r_1(h) \rightarrow 0$; de ce fait le premier facteur tend vers 0. De plus,

$$\frac{\|A.h + r_1(h)\|}{\|h\|} \leq \frac{\|A.h\| + \|r_1(h)\|}{\|h\|} \leq \|A\| + \frac{\|r_1(h)\|}{\|h\|}$$

Puisque $\frac{\|r_1(h)\|}{\|h\|}$ est borné (pour $|h| \leq \eta$ où $\eta > 0$ est choisi tel que $B[a, \eta] \subseteq \mathcal{U}$), le deuxième facteur est borné. Le produit tend donc vers 0. \square

Définition 2.14 (Dérivées directionnelles). Soit $\|V\|_2 = 1$, $\Omega \subseteq \mathbb{R}^d$ ouvert, $f : \Omega \rightarrow \mathbb{R}^m$. Alors

$$\frac{\partial f}{\partial V}(a) = \lim_{h \rightarrow 0} \frac{f(a + Vh) - f(a)}{h}$$

où $h \in \mathbb{R}$

Remarque 2.15. a) Cette définition est proche de la définition de la dérivée partielle : si on prend $V = (0 \dots 0, 1, 0 \dots 0) = e_k$ on obtient : $\frac{\partial f}{\partial e_k}(a) = \frac{\partial f}{\partial x_k}(a)$: la dérivée partielle selon x_k

b) En particulier, $\frac{\partial f}{\partial V} = D_f(a)(V)$ si f est différentiable.

c) **Attention** : Il est possible que $\frac{\partial f}{\partial V}(a)$ existe pour tout $V \in \mathbb{R}^d$, $\|V\|_2 = 1$ sans que f soit différentiable en $x = a$:

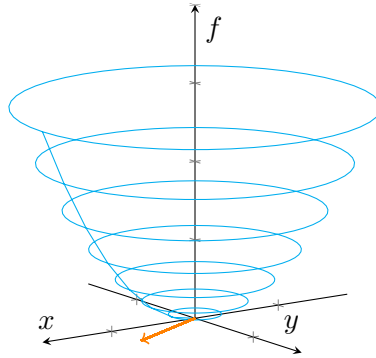
Exemple : $f(x, y) = \frac{x^3}{x^2 + y^2}$ et $a = (0, 0)$. En effet (exo maison) $\frac{\partial f}{\partial x}(0, 0) = 1$ et $\frac{\partial f}{\partial y}(0, 0) = 0$ mais $r(h, k) = f(h, k) - \langle (h, k), (1, 0) \rangle = \frac{hk^2}{h^2 + k^2}$ ne décroît pas assez vite pour être $o(\sqrt{h^2 + k^2})$. Pourtant $\lim_{t \rightarrow 0} \frac{1}{t} f(t \cos \varphi, t \sin \varphi) = \lim_{t \rightarrow 0} \cos^3(\varphi)$ existe pour tout φ .

Remarque 2.16. Le gradient indique la direction de la "pente la plus raide"

Soit $f : \Omega \rightarrow \mathbb{R}$ est différentiable, on a : $\frac{\partial f}{\partial V}(a) = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(a_j) = \langle (\nabla f)(a), V \rangle$. Par conséquent (C.S.) : $|\frac{\partial f}{\partial V}(a)| \leq \|V\|_2 \|(\nabla f)(a)\|_2 = \|(\nabla f)(a)\|_2$, avec égalité si et seulement si $(\nabla f)(a) = \lambda.V$, avec $\lambda \in \mathbb{R}$. Cela signifie que la dérivée directionnelle (la pente le long d'une courbe en direction de V) est maximale en direction du gradient.

Illustration avec $f(x, y) = x^2 + y^2$: $(\nabla f)(x, y) = (2x, 2y)$ On veut un vecteur normalisé : $\vec{V} = \frac{(x, y)}{\sqrt{x^2 + y^2}}$

En un point (x, y) la pente la plus raide sera en direction de \vec{V}



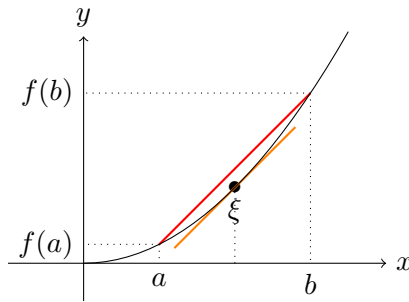
(la tangente la plus pentue est en direction du gradient (en orange))

2.5.1 Accroissements finis

Rappel : en dimension $n = m = 1$ on a

$$\frac{f(a) - f(b)}{a - b} = f'(\xi), \quad \xi \in [a, b] \quad (2.2)$$

La tangente en $(\xi, f(\xi))$ est parallèle à la droite passant par $(a, f(a))$ et $(b, f(b))$.



Peut on étendre ce théorème à plusieurs dimensions ? Par exemple $f(t) = (\cos(t), \sin(t))$. Observons que $f(0) = f(2\pi)$. Mais puisque $f'(\xi) = (-\sin(\xi), \cos(\xi))$ est de norme (euclidienne) 1, $f(2\pi) - f(0) \neq f'(\xi)$ peu importe ξ . La formule (2.2) est donc fausse pour de fonctions vectorielles. On a néanmoins un inégalité :

Lemme 2.17. Soit $f \in \mathcal{C}^1 : \mathbb{R} \rightarrow \mathbb{R}^m$. On suppose $x < y$ Alors il existe $\xi \in [x, y]$ tel que $\|f(x) - f(y)\|_2 \leq (y - x)\|f'(\xi)\|_2$

Démonstration. Soit $g(t) = \langle f(t) | V \rangle$. Par construction c'est une fonction scalaire et donc il existe ξ_V tel que : $g(y) - g(x) = (y - x)g'(\xi_V)$. On fixe $V = \frac{f(y) - f(x)}{\|f(y) - f(x)\|_2}$ pour qu'il soit

normalisé. Alors

$$\begin{aligned}
 (y-x)g'(\xi_V) &= \langle f(y)|V \rangle - \langle f(x)|V \rangle = \langle f(y) - f(x)|V \rangle \\
 &= \langle f(y) - f(x) | \frac{f(y) - f(x)}{\|f(y) - f(x)\|_2} \rangle \\
 &= \langle f(y) - f(x) | f(y) - f(x) \rangle \cdot \frac{1}{\|f(y) - f(x)\|_2} \\
 &= \|f(y) - f(x)\|_2.
 \end{aligned}$$

Par Cauchy-Schwarz,

$$\|f(y) - f(x)\|_2 = |y-x| |g'(\xi_V)| = |y-x| |\langle f'(\xi_V), V \rangle| \leq (y-x) \|f'(\xi_V)\|_2 \|V\|_2.$$

et $\|V\|_2 = 1$. □

Corollaire 2.18. Soit $f : \Omega \subseteq \mathbb{R}^d \rightarrow \mathbb{R}^m$, f différentiable, $x \in \Omega$, $h \in \mathbb{R}^d$, tels que : $[x, x+h] \stackrel{\text{def}}{=} \{x+t.h, t \in [0,1]\} \subseteq \Omega$ (un déplacement en ligne droite de x vers $x+h$, tel que ce "segment" soit entièrement dans Ω). Alors :

$$f(x+h) - f(x) = \int_0^1 D_f(x+sh)(h) ds$$

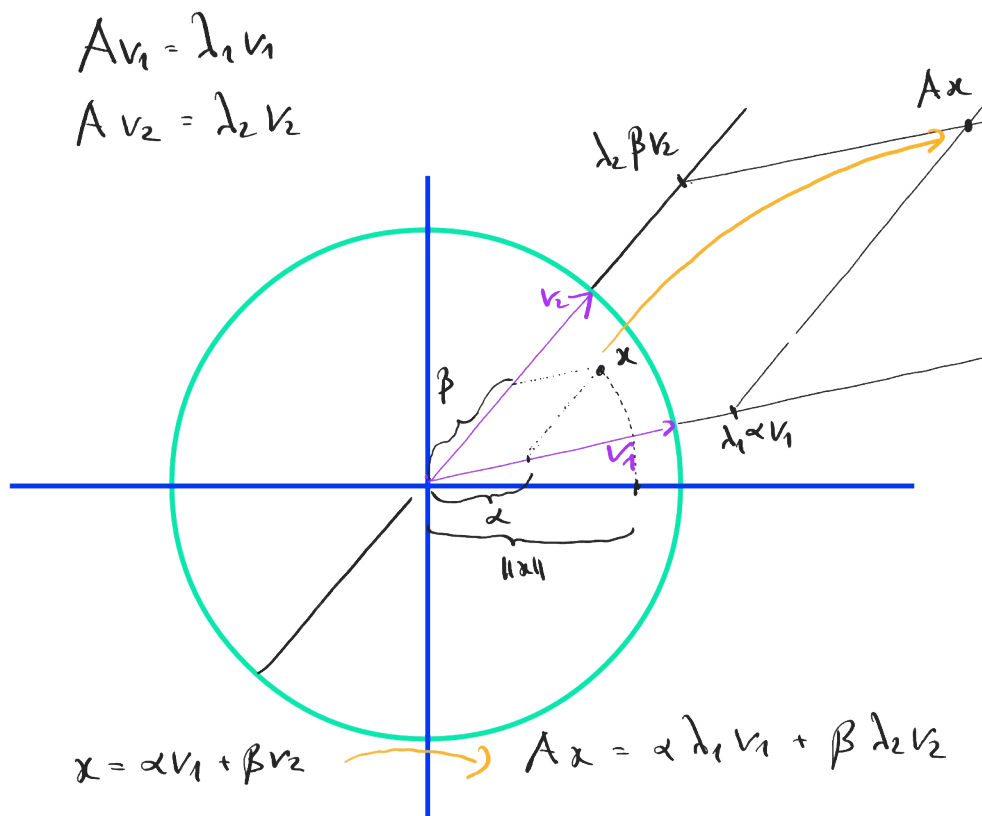
Démonstration. On pose $g(t) = f(x+th)$, $g : [0,1] \rightarrow \mathbb{R}^m$. On applique la formule : $g(1) - g(0) = \int_0^1 g'(s) ds$, car f différentiable donc g est différentiable ; Ce qui donne $f(x+h) - f(x) = \int_0^1 (D_f(x+sh))(h) ds$. □

En particulier :

$$\boxed{\|f(x+h) - f(x)\| \leq \max\{\|D_f(x+sh)\|_{2 \rightarrow 2} : s \in [0,1]\} \cdot \|h\|} \quad (2.3)$$

où $\|D_f(x+sh)\|_{2 \rightarrow 2}$ est la norme d'opérateur : $\|M\|_{p \rightarrow q} = \sup\{\frac{\|Mx\|_q}{\|x\|_p} : x \neq 0\}$, ici pour $p = q = 2$. Ceci est l'**inégalité des accroissements finis**.

Remarque 2.19. La norme d'opérateur d'une matrice M pour la norme Euclidienne n'est pas facile à calculer exactement. En fait c'est la racine de la plus grande valeur propre de $A = {}^t M.M$. Voici l'explication : A est symétrique et réelle, donc d'après Théorème 2.25 diagonalisable : il existent $\lambda_1, \dots, \lambda_n$ réels, positifs, et $v_1, \dots, v_n \in \mathbb{R}^n$ avec $Av_i = \lambda_i v_i$. Voici ce qui se passe, dans le cas de la dimension $n = 2$ (le cas $n > 2$ se traite de la même façon). Un point x a une décomposition dans la base (!) des vecteurs propres, $x = \alpha v_1 + \beta v_2$. Cette décomposition nous donne la valeur de Ax , notamment $Ax = \alpha \lambda_1 v_1 + \beta \lambda_2 v_2$. Voir image :



On voit que "l'amplification de longueur" est au plus celle de la plus grande valeur propre, donc $\|A\|_{2 \rightarrow 2} \leq \lambda_{\max}$, mais on voit aussi égalité car $\frac{\|Av_{\max}\|_2}{\|v_{\max}\|_2} = \lambda_{\max}$. Observons que Cauchy-Schwarz nous donne

$$\begin{aligned} \|Mx\|_2^2 &= \sup\{\langle Mx, My \rangle : \|y\|_2 \leq 1\} \\ &= \sup\{\langle Ax, y \rangle : \|y\|_2 \leq 1\} \\ &= \|Ax\|_2 \end{aligned}$$

ce qui explique qu'il faut prendre la racine de $\lambda_{\max}(A)$ pour trouver $\|M\|_{2 \rightarrow 2}$. Je conviens que ceci n'est pas commode à calculer. Il est cependant facile de voir que

$$\|M\|_{2 \rightarrow 2} \leq \|M\|_{\text{H.S.}} \stackrel{\text{def.}}{=} \left(\sum_{i,j} |m_{i,j}|^2 \right)^{1/2}$$

et souvent cette borne supérieure peut-être utilisée dans l'inégalité des accroissements finis à la place de la bonne norme – vu que ce sont deux normes dans $\mathbb{R}^{n \times n}$ (e.v.n. de dim. finie) on ne 'perd' au plus constante universelle par rapport à la 'vraie' norme $2 \rightarrow 2$, mais on gagne du temps. Je note que le sigle 'H.S.' signifie "Hilbert-Schmidt", et pas "hors service" ;-)

Une application utile des accroissements apparaît dans la discussion différentiabilité de fonctions :

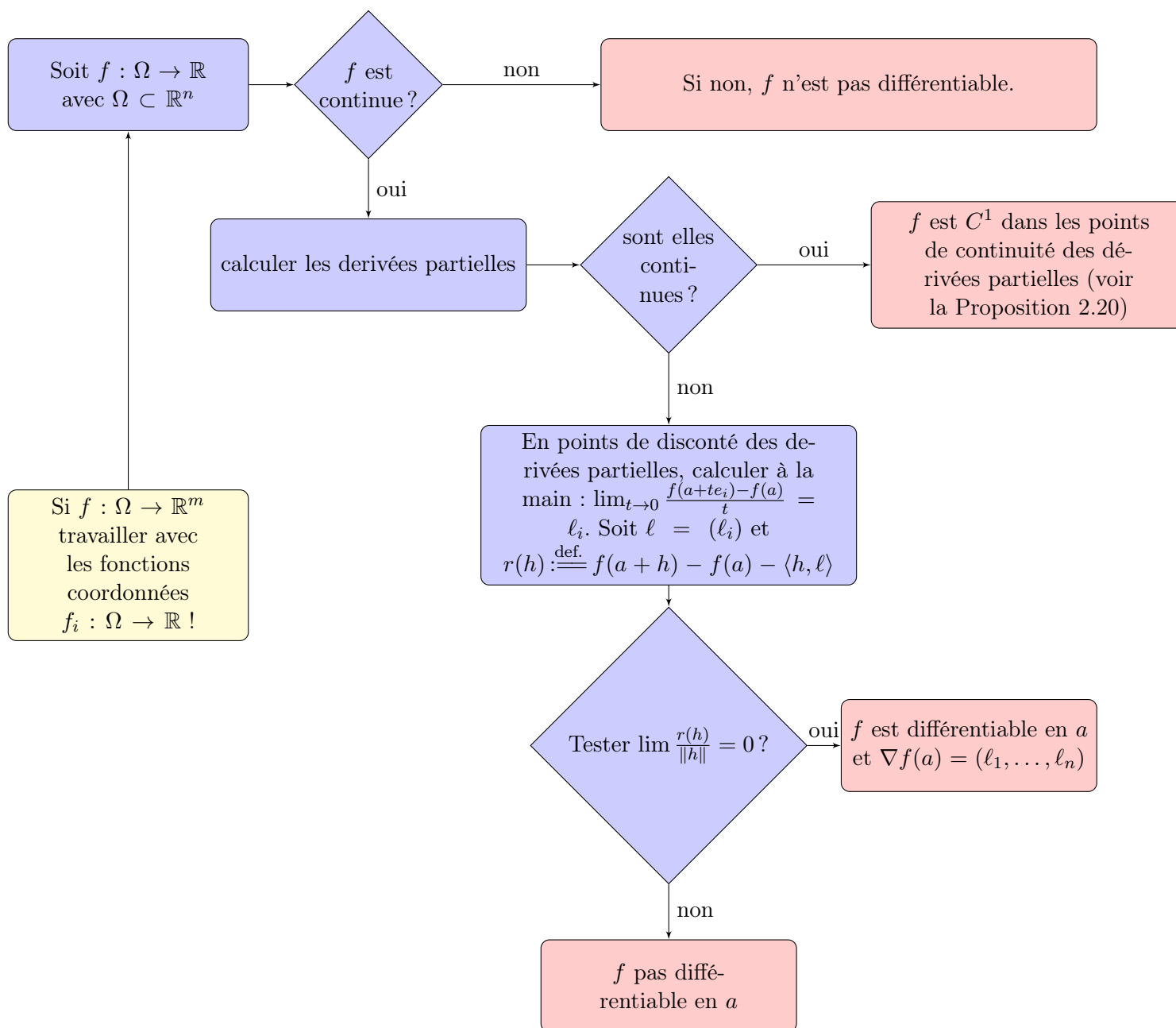
Proposition 2.20. Soit $\Omega \subseteq \mathbb{R}^d$ ouvert et $f : \Omega \rightarrow \mathbb{R}$ une fonction continue qui est $C^1(\Omega \setminus \{a\})$. Si $\lim_{x \rightarrow a} D_f(x) \stackrel{\text{def.}}{=} L$ existe, alors $f \in C^1(\Omega)$ et $D_f(a) = L$.

Démonstration. On pose $r(h) = f(a+h) - f(a) - L(h)$. Alors

$$r(h) = \int_0^1 (D_f(a+th) - L)h dt$$

Soit $\varepsilon > 0$. Il existe $\delta > 0$ tel que $\|k\| < \delta$ implique $\|D_f(a+k) - L\| < \varepsilon$. Il suit pour $\|h\| < \delta$ que $\|r(h)\| \leq \varepsilon\|h\|$. Ainsi $\|r(h)\|/\|h\| \rightarrow 0$ quand $h \rightarrow 0$. \square

Ceci amène à la procédure suivante si $f : \Omega \rightarrow \mathbb{R}$ une fonction continue qui est $C^1(\Omega \setminus \{a\})$:



Si $f : \Omega \rightarrow \mathbb{R}^m$
travailler avec
les fonctions
coordonnées
 $f_i : \Omega \rightarrow \mathbb{R}!$

Flowchart d'analyse de différentabilité d'une fonction de plusieurs variables.

2.6 La formule de Taylor

On s'intéresse à étendre les formules de Taylor[‡] au cas multi-dimensionnel. Revenons alors d'abord au cas 1D :

2.6.1 Le cas 1D

Commençons avec un lemme auxiliaire :

Lemme 2.21. Soit $f, g : [a, b] \rightarrow \mathbb{R}$ continues et de classe C^1 dans (a, b) . Supposons $g'(x) \neq 0$ pour tout $x \in (a, b)$. Alors $g(a) \neq g(b)$ et il existe un $\xi \in (a, b)$ tel que

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(\xi)}{g'(\xi)}$$

Démonstration. Si $g(a) = g(b)$ le lemme de Rolle[§] montre l'existence d'un x avec $g'(x) = 0$. Ceci n'est pas le cas : ainsi, les hypothèses assurent $g(a) \neq g(b)$. Posons

$$F(x) = f(x) - \frac{f(b) - f(a)}{g(b) - g(a)} (g(x) - g(a)).$$

Puisque $F(a) = F(b) = f(a)$, le lemme de Rolle montre qu'il existe un $\xi \in (a, b)$ avec $F'(\xi) = 0$, c'est à dire

$$f'(\xi) = \frac{f(b) - f(a)}{g(b) - g(a)} (g'(\xi)). \quad \square$$

Théorème. 2.22 (Formule de Taylor avec reste de Lagrange[¶]). Soit I un intervalle ouvert de \mathbb{R} et $f : I \rightarrow \mathbb{R}$ de classe C^k . Alors, pour tout $x \neq a \in I$ il existe un ξ entre a et x tel que

$$f(x) = \sum_{j=0}^{k-1} \frac{f^{(j)}(a)}{j!} (x-a)^j + \frac{f^{(k)}(\xi)}{k!} (x-a)^k$$

Démonstration. On peut supposer $a < x$ pour simplifier la notation. Soit $F(x) = f(x) - \sum_{j=0}^{k-1} \frac{f^{(j)}(a)}{j!} (x-a)^j$. et $G(x) = (x-a)^k$. Alors $F(a) = F'(a) = \dots = F^{(k-1)}(a) = 0$ et pour $0 \leq j < k$, la dérivée $G^{(j)}(x) = k(k-1)\dots(k-j+1)(x-a)^{k-j}$ s'annule si et seulement si

‡. Brook Taylor est un homme de science anglais, né à Edmonton, aujourd'hui un quartier de Londres, le 18 août 1685, et mort à Londres le 29 décembre 1731.

§. Michel Rolle, né à Ambert le 21 avril 1652 et mort à Paris le 8 novembre 1719, est un mathématicien français.

¶. Joseph Louis de Lagrange (en italien Giuseppe Luigi Lagrangia ou aussi Giuseppe Ludovico De la Grange Tournier), né à Turin en 1736 de parents français descendants de Descartes et mort à Paris en 1813, est un mathématicien, mécanicien et astronome sarde, avec une *vitae* "européenne" : à l'âge de trente ans, il quitte Turin et va séjourner à Berlin pendant 21 ans. Ensuite, il s'installe pour ses vingt-six dernières années à Paris où il prend la nationalité française en 1802

$x = a$. On peut appliquer le lemme de proche en proche pour obtenir

$$\begin{aligned} \frac{F(x)}{G(x)} &= \frac{F(x) - F(a)}{G(x) - G(a)} = \frac{F'(\xi_1)}{G'(\xi_1)} \\ &= \frac{F'(\xi_1) - F'(a)}{G'(\xi_1) - G'(a)} = \frac{F'(\xi_2)}{G'(\xi_2)} \\ &\vdots \\ &= \frac{F^{(k-1)}(\xi_{k-1}) - F^{(k-1)}(a)}{G^{(k-1)}(\xi_{k-1}) - G^{(k-1)}(a)} = \frac{F^{(k)}(\xi_k)}{G^{(k)}(\xi_k)} = \frac{f^{(k)}(\xi_k)}{k!} \end{aligned}$$

Alors $F(x) = \frac{f^{(k)}(\xi_k)}{k!} G(x)$ ce qui est équivalent à l' énoncé. \square

On appelle $\frac{f^{(k)}(a)}{k!} (\xi - a)^k$ le **reste de Lagrange** dans la formule de Taylor.

2.6.2 Le cas de plusieurs variables

Soit $\Omega \subseteq \mathbb{R}^d$ ouvert, et $f : \Omega \rightarrow \mathbb{R}$ k -fois partiellement et continuellement différentiable. On pose :

$$(P_j f)(x, h) \stackrel{\text{def.}}{=} \sum_{l_1=1}^d \dots \sum_{l_j=1}^d \left(\frac{\partial}{\partial x_{l_1}} \dots \frac{\partial f}{\partial x_{l_j}} \right)(x) \cdot h_{l_1} \dots h_{l_j}$$

Théorème. 2.23. *Alors si $a \in \Omega$ et $h \in \mathbb{R}^d$ avec $[a, a+h] = \{a + t.h, t \in [0, 1]\} \subseteq \Omega$ on a :*

$$\begin{aligned} f(a+h) &= \sum_{j=0}^{k-1} \frac{(P_j f)(a, h)}{j!} + \frac{P_k(a + \theta h, h)}{k!} \\ &= \sum_{j=0}^k \frac{(P_j f)(a, h)}{j!} + r(h) \quad \text{avec } r(h) = o(\|h\|^k), \text{ i.e. } \frac{\|r(h)\|}{\|h\|^k} \xrightarrow{h \rightarrow 0} 0 \end{aligned}$$

Démonstration. Soit $g(t) = f(a + th)$. Alors

$$g'(t) = D_f(a+th)(h) = \langle \nabla f(a+th) | h \rangle = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(a+th) \cdot h_j = \frac{\partial f}{\partial x_1}(a+th) \cdot h_1 + \dots + \frac{\partial f}{\partial x_n}(a+th) \cdot h_n$$

On re-dérive partiellement pour chaque $j \in [1, n]$, on obtient une somme de dérivées partielles. (le premier terme de la dérivée seconde $g''(t)$ est donné par $\sum_{i=1}^n \frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_1}(a+th) \cdot h_1 \right) \cdot h_i$) Ce qui nous donne au rang j : $g^{(j)}(t) = (P_j f)(a + th, h)$. On écrit le théorème de Taylor 2.22 pour g :

$$g(t) = \sum_{j=0}^{k-1} \frac{g^{(j)}(a)}{j!} (t-a)^j + \frac{g^{(k)}(\xi)}{k!} (t-a)^k \Leftrightarrow f(a+h) = \sum_{j=0}^{k-1} \frac{(P_j f)(a, h)}{j!} + \frac{P_k(a + \theta h, h)}{k!}$$

Deuxième égalité de Taylor-Lagrange : Soit

$$\begin{aligned} r(h) &= \frac{1}{k!} ((P_k f)(a + \theta h, h) - (P_j f)(a, h)) \\ &= \frac{1}{k!} \sum_{l_1=1}^n \cdots \sum_{l_j=1}^n \left[\frac{\partial^k f}{\partial x_{l_1} \cdots \partial x_{l_k}}(a + \theta h) - \frac{\partial^k f}{\partial x_{l_1} \cdots \partial x_{l_k}}(a) \right] h_{l_1} \cdots h_{l_k} \\ &\leq \frac{1}{k!} \sum_{l_1=1}^n \cdots \sum_{l_j=1}^n \|h\|_2^k \left[\frac{\partial^k f}{\partial x_{l_1} \cdots \partial x_{l_k}}(a + \theta h) - \frac{\partial^k f}{\partial x_{l_1} \cdots \partial x_{l_k}}(a) \right] \end{aligned}$$

Par continuité des dérivées k-ièmes, pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que : $\|h\|_2 < \delta$ implique

$$\left| \frac{\partial^k f}{\partial x_{l_1} \cdots \partial x_{l_k}}(a + \theta h) - \frac{\partial^k f}{\partial x_{l_1} \cdots \partial x_{l_k}}(a) \right| \leq \varepsilon$$

Ce qui donne $r(h) \leq \frac{\|h\|_2^k}{k!} \cdot \varepsilon$. □

Notation avec le multi-indices : On appelle $\alpha \in \mathbb{N}^d$ un “multi-indice” : On note alors

$$\begin{cases} |\alpha| = \alpha_1 + \dots + \alpha_d \\ \partial^\alpha f = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_d}}{\partial x_d^{\alpha_d}}(f) \\ h^{(\alpha)} = h_1^{\alpha_1} h_2^{\alpha_2} \cdots h_d^{\alpha_d} \end{cases}$$

La formule de Taylor s’écrit alors plus confortablement :

$$f(a + h) = \sum_{j=0}^k \frac{1}{j!} \left(\sum_{|\alpha|=j} (\partial^\alpha f)(a) \cdot h^{(\alpha)} \right) + r(h) \quad \text{avec } \frac{\|r(h)\|}{\|h\|^k} \xrightarrow{h \rightarrow 0} 0$$

2.6.3 Le polynôme de Taylor d’ordre 2 et la matrice Hessienne

Le cas $k = 2$ est particulièrement important. Soit $f : \Omega \rightarrow \mathbb{R}$, deux fois continûment partiellement dérivable. On définit **la matrice Hessienne** :^{||}, la matrice des dérivées partielles secondes comme telle

$$\text{Hess}_f(a) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(a) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(a) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(a) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(a) & \vdots & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2}(a) \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(a) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(a) \end{pmatrix}$$

Par le lemme de Schwarz, cette matrice est symétrique si f est de classe \mathcal{C}^2 ce qui donne :

$$f(a + h) = f(a) + \langle (\nabla f)(a) | h \rangle + \frac{1}{2} \langle \text{Hess}_f(a) \cdot h | h \rangle + r_2(h)$$

où $r_2(h_1, \dots, h_n) = o(h_1^2 + \dots + h_n^2)$.

^{||}. d’après Ludwig Otto Hesse, né le 22 avril 1811 à Königsberg (Prusse), mort le 4 août 1874 à Munich, Allemagne), un mathématicien allemand.

Exemple 2.24. Soit $f(x, y) = x^2 + y \arctan(x)$:

$$f_x(x, y) = 2x + \frac{y}{1+x^2} \quad \text{et} \quad f_y(x, y) = \arctan(x)$$

Puis : $f_{x,y}(x, y) = \frac{1}{1+x^2}$; $f_{x,x}(x, y) = 2 - \frac{2x}{(1+x^2)^2}$; $f_{y,y}(x, y) = 0$. Évaluées en $a = (0, 0)$ cela

donne : $\text{Hess}_f(0, 0) = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$ Avec $h = (h_1, h_2)$:

$$f(a+h) = f(0+h_1, 0+h_2) = f(0, 0) + \langle (\nabla f)(a), h \rangle + \langle \text{Hess}_f(0, 0).h, h \rangle + r(h)$$

• $(\nabla f)(a) = (f_x, f_y)$ donc $(\nabla f)(0, 0) = (0, 0)$. Ensuite $\langle (\nabla f)(a)|h \rangle = 0$

• $\text{Hess}_f(0, 0).h = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} 2h_1 + h_2 \\ h_1 \end{pmatrix}$.

D'où : $\langle \text{Hess}_f(0, 0).h|h \rangle = 2h_1^2 + 2h_1h_2$, puis $f(x, y) = 0 + 0 + (2x^2 + 2xy) + r(x, y)$ (avec $h = (x, y)$).

2.7 Quelques résultats d'algèbre linéaire

Rappel : une matrice est orthogonale si ${}^t A.A = I_n$. Soit $A = (a_1 | \dots | a_n)$ avec vecteurs colonnes a_i . Alors ${}^t A.A = (\langle a_i | a_j \rangle)_{i,j}$, la matrice formée de tous les produits scalaires ! Cela signifie que A est orthogonale si et seulement si $\{a_1, \dots, a_n\}$ est une base orthonormale (on devrait appeler la matrice orthonormale et pas orthogonale, mais l'expression est figée, même si elle manque de la cohérence).

Théorème. 2.25. Si A est une matrice à coef. réels, symétrique, alors elle est orthogonalement diagonalisable : $A = {}^t O D O$ où O est une matrice orthogonale. Les colonnes de O sont des vecteurs propres de A et où D est une matrice diagonale dont les coefficients sont précisément les valeurs propres de A .

Démonstration. Par récurrence : le cas $n = 1$ est clair. Soit A de taille $n \times n$, et supposons que le théorème soit vrai pour des matrices $k \times k$ pour tout $k < n$. On utilise deux ingrédients :

a) (ingrédient 1) Le polynôme caractéristique $p(\lambda) = \det(A - \lambda)$ admet au moins une racine complexe μ ; il suit que $\ker(A - \mu) \neq \{0\}$. Prenons $v \in \mathbb{C}^n$ non-nul avec $Av = \mu v$. Alors, puisque

$$\langle Av, v \rangle = \overline{\langle v, Av \rangle} = \overline{\langle Av, v \rangle}$$

nous avons que $\langle Av, v \rangle = \mu \|v\|_2^2$ est réel. Il suit que μ est réel, et donc $v \in \mathbb{R}^n$.

b) (ingrédient 2). Soit v vecteur propre à $\lambda \in \mathbb{R}$ et w vecteur propre à $\mu \neq \lambda$. Alors

$$\mu \langle v, w \rangle = \langle Av, w \rangle = \langle v, Aw \rangle = \lambda \langle v, w \rangle$$

Il suit que $v \perp w$.

Soit $\lambda_1, \dots, \lambda_k$ les valeurs propres (nécessairement réelles) distinctes.

• Si $k = n$, les vecteurs propres associés (renormalisés) forment une base orthonormale (clair : orthogonal implique indépendant, c'est donc bien une base !), il n'y a rien à faire.

• Sinon, $k < n$. On forme l'espace U en associant à tout λ_k un vecteur propre v_k normalisé. Par construction, $U = \text{vect}(v_1 \dots v_k)$ est invariant sous A : $A(U) \subseteq U$. Choisissons

une base orthogonale $\{w_{k+1}, \dots, w_n\}$ de U^\perp qui complémente les vecteurs propres en une base orthonormale de \mathbb{R}^n . Soit $Q_1 = (v_1 | \dots | v_k | w_{k+1} | \dots | w_n)$. Q_1 est orthogonale, et

$$A_1 \stackrel{\text{def.}}{=} {}^t Q_1 A Q_1 = \begin{pmatrix} \begin{matrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_k \end{matrix} & 0 \\ 0 & \boxed{B} \end{pmatrix}$$

est une matrice par blocs : B est une matrice de taille $(n-k) \times (n-k)$. Puisque A est symétrique, A_1 et donc B sont symétriques. La structure du bloc diagonal supérieur se justifie par le fait que pour $1 \leq i \leq k$, on a $Q_1 e_i = v_i$, $Av_i = \lambda_i v_i$ et ${}^t Q_1 v_i = Q_1^{-1} v_i = e_i$.

C'est l'heure de l'hypothèse de récurrence ! Il existe une matrice orthogonale P de taille $(n-k) \times (n-k)$ avec PBP^t diagonale. Si on utilise la matrice par blocs

$$Q_2 = \begin{pmatrix} \boxed{\text{Id}_k} & 0 \\ 0 & \boxed{P} \end{pmatrix}$$

on obtient que ${}^t Q_2 {}^t Q_1 A Q_1 Q_2$ est diagonale. Finalement, $Q = Q_1 Q_2$ est orthogonale. \square

Définition 2.26. Soit $A \in \mathcal{M}_{d \times d}(\mathbb{R})$ une matrice carrée de taille d .

- On dit que A est **positivement définie** si : $\forall h \in \mathbb{R}^d \setminus \{0\}$, $\langle A.h|h \rangle > 0$
- On dit que A est **positivement semi-définie** si : $\forall h \in \mathbb{R}^d$: $\langle A.h|h \rangle \geq 0$
- On dit que A est **négativement définie**, resp. **semi-négativement définie** si $(-A)$ est positivement, resp. semi-positivement, définie.
- On dit que A est **indéfinie** si aucun des trois cas précédents ne s'appliquent.

Corollaire 2.27. A est positivement définie si et seulement si toutes ses valeurs propres sont strictement positives.

Démonstration. Puisque A est à coef. reels et symétrique, il existe une matrice O orthogonale telle que ${}^t O A O = D = \text{diag}(\lambda_1, \dots, \lambda_d)$ où les λ_i sont les valeurs propres de A . Mais $\sum_{i=1}^d \lambda_i h_i^2 = \langle D.h|h \rangle = \langle A O.h | O.h \rangle > 0$ par hypothèse. Il suit que $\lambda_i > 0$ pour tout $i \in [1, d]$. Réciproquement, si les valeurs propres sont str. positives, $\sum_{i=1}^d \lambda_i h_i^2 > 0$. Or $\{O.h : h \in \mathbb{R}^d\} = \mathbb{R}^d$, A est positivement définie. \square

Corollaire 2.28. A est négativement définie si et seulement si toutes ses valeurs propres sont strictement négatives.

Théorème 2.29 (Le critère de Hurwitz **). Soit A symétrique de taille $d \times d$. Pour $k = 1 \dots d$ soit A_k la sous-matrice "principale"

$$A_k = \begin{pmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{kk} \end{pmatrix}.$$

Alors A est **positivement définie** si et seulement si : $\forall 1 \leq k \leq d \quad \det(A_k) > 0$

** . Adolf Hurwitz (né à Hildesheim le 26 mars 1859 - mort à Zurich le 18 novembre 1919) est un mathématicien allemand. Il était une des figures importantes des mathématiques de la seconde moitié du XIXe siècle.

Démonstration. “ \Rightarrow ” Observons que $A_k = P_k A \mathcal{J}_k$ où $\mathcal{J}_k : \mathbb{R}^k \rightarrow \mathbb{R}^d$ est l’inclusion linéaire (on rajoute des zéros), et P_k est la projection sur les k premières coordonnées. On a pour tout $x \in \mathbb{R}^k \setminus \{0\}$

$$\langle A_k x | x \rangle_{\mathbb{R}^k} = \langle P_k A \mathcal{J}_k x | x \rangle_{\mathbb{R}^k} = \langle A \mathcal{J}_k x | \mathcal{J}_k x \rangle_{\mathbb{R}^d} > 0$$

ce qui montre bien que A_k est positivement définie car A l’est. Mais le déterminant d’une matrice est le produit des ses valeurs propres, donc $\det(A_k) > 0$ pour tout k .

“ \Leftarrow ” : On procède par récurrence. Pour $d = 1$ rien n’est à faire. On suppose donc le théorème vrai pour des matrices $(d-1) \times (d-1)$. Par hypothèse, $\det(A_k) > 0$. Soit $B = A_{d-1}$. Alors il existe $\alpha \in \mathbb{R}$ et $a \in \mathbb{R}^{d-1}$ avec

$$A = \begin{pmatrix} \boxed{B} & a \\ a^t & \boxed{\alpha} \end{pmatrix}.$$

B est inversible, car son déterminant est non-nul. Soit $b = B^{-1}a$. Soit P la matrice de passage telle que :

$$P = \begin{pmatrix} \boxed{\text{Id}_{d-1}} & -b \\ 0 & \boxed{1} \end{pmatrix}.$$

Alors

$${}^t P A P = \begin{pmatrix} \boxed{B} & 0 \\ 0 & \boxed{\tilde{\alpha}} \end{pmatrix}.$$

Or $\det({}^t P A P) = \det(P)^2 \det(A) = \det(A) > 0$ par hypothèse. Mais d’autre part : $\det(A) = \tilde{\alpha} \det(B)$ et $\det(B) > 0$ ce qui implique $\tilde{\alpha} > 0$. Il suit que $\tilde{\alpha}$ est valeur propre (avec la dernière colonne de P^{-1} comme vecteur propre). Par hypothèse de récurrence, B a des valeurs propres strictement positives, et on conclut. \square

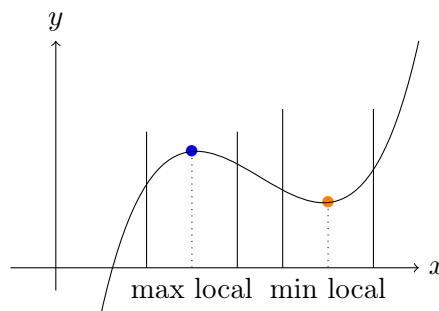
Corollaire 2.30 (Caractérisation de matrices négativement définies). *A est négativement définie si et seulement si $\det(A_1) < 0$; $\det(A_2) > 0$; $\det(A_3) < 0 \dots$. En effet :*

$$-A_k = \begin{pmatrix} -a_{11} & \dots & -a_{1k} \\ \vdots & & \vdots \\ -a_{k1} & \dots & -a_{kk} \end{pmatrix}$$

nous donne $\det(-A_k) = (-1)^k \det(A_k)$ d’où l’alternance du signe.

2.8 Extrema locaux

Définition 2.31. f admet un minimum local en $x = a$ s'il existe un ouvert $\mathcal{U} \subseteq \Omega$ avec $a \in \mathcal{U}$ et $f(a) \leq f(x) \forall x \in \mathcal{U}$. De même : f admet un maximum local en $x = a$ s'il existe un ouvert $\mathcal{U} \subseteq \Omega$ avec $a \in \mathcal{U}$ et $f(a) \geq f(x) \forall x \in \mathcal{U}$.

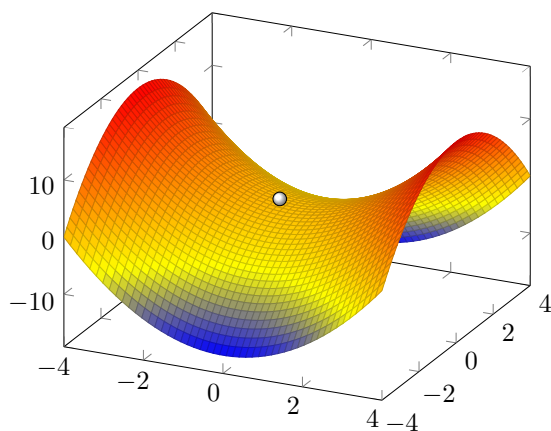
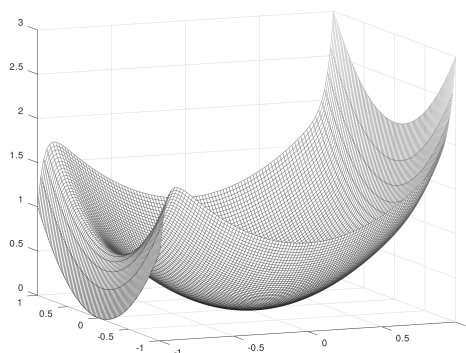


Observation :

Si f admet un minimum local en a et si $\gamma : [-1, 1] \rightarrow \mathbb{R}^d$ est une courbe paramétrée avec (par exemple) $\gamma(0) = a$, alors $f \circ \gamma$ admet un minimum local en 0. En particulier, si f admet un minimum local en a alors nécessairement $\frac{\partial f}{\partial x_k}(a) = 0$ (où $\frac{\partial f}{\partial x_k}(a) = 0 = (f \circ \gamma_k)'(0)$ avec

$$\gamma_k(t) = (a_1, \dots, a_{k-1}, a_k + t, a_{k+1}, \dots, a_d).$$

On déduit alors :



Lemme 2.32 (critère nécessaire).

Si $f \in \mathcal{C}^2(\Omega, \mathbb{R})$ admet un minimum local en $x = a$ alors $(\nabla f)(a) = 0$. Un tel point a est appelé un **point critique**.

Le critère nécessaire n'est pas suffisant. L'exemple $f(x, y) = x^2 - y^2$ à gauche illustre : dans un "point selle" le gradient peut s'annuler sans qu'on ait un extremum local.

Théorème. 2.33 (critère suffisant). Soit $f \in \mathcal{C}^2(\Omega)$, avec Ω un ouvert de \mathbb{R}^d , et $a \in \Omega$ un point critique. Alors :

- Si $\text{Hess}_f(a)$ est **négativement définie**, f admet un **maximum local** en a
- Si $\text{Hess}_f(a)$ est **positivement définie**, f admet un **minimum local** en a
- Si $\text{Hess}_f(a)$ est **indéfinie**, alors f admet un **point de selle** en a .

Démonstration. a) On se place dans la situation (b) : $\text{Hess}_f(a)$ est positivement définie. Soit $\alpha = \min\{\langle Ah|h \rangle : \|h\|_2 = 1\}$. Ce minimum existe car $S = \{h \in \mathbb{R}^d : \|h\|_2 = 1\}$ est compact, et $f : h \mapsto \langle Ah|h \rangle$ est continue. Par hypothèse $\alpha > 0$. On observe que pour

$h \in \mathbb{R}^d \setminus \{0\} : \langle Ah|h \rangle = \|h\|_2 \langle A \frac{h}{\|h\|_2} | \frac{h}{\|h\|_2} \rangle \geq \alpha \|h\|_2^2$ car $\frac{h}{\|h\|_2}$ est de norme 1. On applique le théorème de Taylor :

$$f(a+h) = f(a) + \langle (\nabla f)(a).h|h \rangle + \langle A.h|h \rangle + r(h)$$

où : $\langle (\nabla f)(a).h|h \rangle = 0$ et $\langle A.h|h \rangle \geq \alpha \|h\|_2^2$. Or $\frac{|r(h)|}{\|h\|_2^2} \xrightarrow{h \rightarrow 0} 0$, il existe un rayon $R > 0$ tel que $\frac{|r(h)|}{\|h\|_2^2} \leq \frac{\alpha}{2}$ si $\|h\|_2 < R$. Il suit que $\forall h$ tel que $\|h\|_2 < R$

$$f(a+h) = f(a) + \frac{\alpha}{2} \|h\|_2^2 \geq f(a).$$

Donc a est un minimum local.

- b) Ce qui donne pour le (a) : f admet un maximum local en a si et seulement si $-f$ admet un minimum local en a . Et comme $\text{Hess}_f(a) = -\text{Hess}_{-f}(a)$, a est bien dans ce cas un max local pour f .
- c) Par la formule de Taylor $f(a+h) = f(a) + \langle A.h|h \rangle + r(h)$. Si $\langle A.h|h \rangle$ change de signe, $r(h)$ ne pourra pas “venir au secours” dès que $\|h\|$ est suffisamment petit : En effet, supposons $\eta > 0$ et h_1, h_2 tels que

$$\langle A.h_1|h_1 \rangle > \eta \quad \text{et} \quad \langle A.h_2|h_2 \rangle < -\eta.$$

Soit δ tel que $\|r(h)\|/\|h\|^2 < \frac{\eta}{2 \max(\|h_1\|, \|h_2\|)^2}$ pour tout h de norme plus petit que δ . Alors pour tout $|t| < \delta$ suffisamment petit,

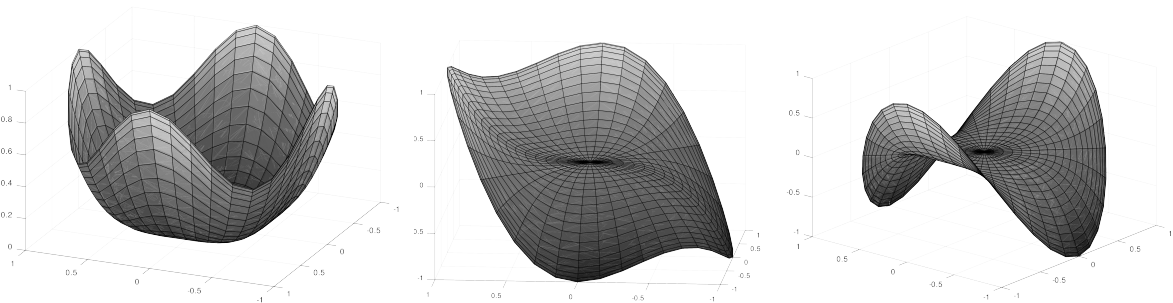
$$f(a+th_1) > f(a) + t^2\eta + t^2 \frac{r(th_1)}{t^2} > f(a) + \frac{1}{2}t^2\eta$$

tandis que

$$f(a+th_2) < f(a) - t^2\eta + t^2 \frac{r(th_2)}{t^2} < f(a) - \frac{1}{2}t^2\eta$$

Ainsi, f n’admet ni minimum ni maximum local : dans tout voisinage de a , f admet des valeurs plus petites ET plus grandes que $f(a)$. \square

Attention, dans le théorème il y a un cas non-traité : si (au moins une) valeur propre est nulle ! Dans ce cas, l’étude de la Hessienne ne permet pas de conclure. Ceci est un phénomène bien connu d’une variable réelle : soit $f(x) = x^3$. Le point critique $x = 0$ satisfait $f''(x) = 0$ et il n’est pas extremal. Mais pour $g(x) = \pm x^4$ nous avons $g'(0) = 0$ et $g''(0) = 0$ avec un minimum/maximum local et global. La même chose se passe en plusieurs dimensions : il suffit de regrader quelques exemples : de gauche à droite, un minimum ordre 4 avec $f(x, y) = x^4 + y^4$, un point “fauteil” avec $g(x, y) = x^3 + y^3$ et un point selle ordre 4 avec $h(x, y) = x^4 - y^4$.



Tous les trois exemples ont le seul point critique $(0, 0)$ avec la matrice Hessienne qui est nulle au point $(0, 0)$. Dans un tel cas, il faut regarder “à la main” ou bien utiliser des dérivées supérieures.

3 Les grands théorèmes du calcul différentiel

3.1 Le théorème du point fixe

Théorème. 3.1. (Théorème du point fixe/ de Banach) Soit F une partie fermée d'un e.v.n. (ou d'un espace métrique) complet*, soit $\Phi : F \rightarrow F$ une application telle que

$$\|\Phi(x) - \Phi(y)\| \leq L \cdot \|x - y\|$$

où $L < 1$: on parle d'une contraction stricte.

Alors il existe une unique point fixe x^* tel que $\Phi(x^*) = x^*$.

Exemple 3.2. Si je pose une carte de France sur la table de mon salon, alors l'application qui à la France associe la carte est contractante. De plus si je traverse avec une aiguille la carte à l'emplacement exact de ma table (une toute petite aiguille), j'obtiens l'unique point fixe de l'application ! (le point réel est au même endroit que sa représentation sur la carte).

(*) : Il faut bien de la **complétude** de l'espace dans les hypothèses : le point est qu'on a pas besoin de connaître la limite d'une suite pour démontrer sa convergence, on montre à la place qu'elle est de Cauchy. Sinon, comme on va voir dans la preuve en bas, il faudrait connaître le point fixe pour démontrer qu'il existe – ce qui est absurde. Mais il ne s'agit pas d'une maladresse de preuve, comme l'exemple suivant illustre.

Exemple 3.3. Pour tout $a > 0$, la suite $x_0 = a$, et $x_{n+1} = \frac{1}{2}(x_n + \frac{a}{x_n})$ tend vers \sqrt{a} dans \mathbb{R} : c'est la "suite de Heron" : cette méthode était déjà connue en Mésopotamie vers Hammurabi I, ~ 1750 avant J.C. Vers 100 après J.C., Héron d'Alexandrie l'a rédigé dans son œuvre "Metrica". Pour $a = 2$, la suite (x_n) est bien rationnelle mais ne peut converger dans \mathbb{Q} - manque de limite ! On voit que la complétude (que \mathbb{R} possède, mais pas \mathbb{Q}) est une hypothèse nécessaire.

Démonstration du théorème. a) Étape 1 : Existence :

Soit $x_0 \in F$ quelconque, et soit

$$x_{n+1} \stackrel{\text{def.}}{=} \Phi(x_n)$$

pour tout $n \geq 1$. Alors

$$\begin{aligned} \|x_{n+1} - x_n\| &= \|\Phi(x_n) - \Phi(x_{n-1})\| \\ &\leq L \cdot \|x_n - x_{n-1}\| \\ &\leq L^2 \|x_{n-1} - x_{n-2}\| \\ &\leq \dots \leq L^n \|x_1 - x_0\|. \end{aligned}$$

De ce fait, par série géométrique de raison $L < 1$,

$$\begin{aligned} \|x_n - x_{n+k}\| &\leq \sum_{j=0}^{k-1} \|x_{n+j} - x_{n+j+1}\| \leq \sum_{j=0}^{k-1} L^j \|x_{n+1} - x_n\| \\ &\leq \sum_{j=0}^{\infty} L^{j+n} \|x_1 - x_0\| = L^n \cdot \|x_1 - x_0\| \sum_{j=0}^{\infty} L^j \\ &=; \frac{L^n}{1-L} \|x_1 - x_0\| \end{aligned}$$

Soit maintenant $\varepsilon > 0$. Il existe N tel que pour tout $n > N$ on a : $\|x_n - x_{n+k}\| \leq \frac{L^n}{1-L} \|x_1 - x_0\| \leq \varepsilon$. C'est précisément la caractérisation d'une suite de Cauchy. Par hypothèse de complétude cette suite (x_n) converge vers une limite x^* :

$$x^* = \lim(x_n) = \lim(x_{n+1}) = \lim \Phi(x_n) \stackrel{\Phi \text{ continue}}{=} \Phi(\lim(x_n)) = \Phi(x^*).$$

D'où l'existence d'un point fixe.

b) Étape 2 : Unicité du point fixe :

Soit y^* un point fixe. On a : $\|x^* - y^*\| = \|\Phi(x^*) - \Phi(y^*)\| \leq L \cdot \|x^* - y^*\|$. Or $L < 1$ donc $\|x^* - y^*\| < \|x^* - y^*\|$ Ceci implique $\|x^* - y^*\| = 0$ et $x^* = y^*$ d'où l'unicité. □

Exemple 3.4. Revenons à l'exemple de la carte de France. Pour trouver le point fixe on plante l'aiguille où on veut sur la carte (disons, à Strasbourg). Alors $x_0 =$ Strasbourg. En traversant le papier on obtient un point sur le bureau, appelé x_1 . On le cherche sur la carte (près de Bordeaux donc), et replante une aiguille, qui traverse le papier et marque le bureau dans un nouveau point x_2 . ETC. La suite (x_n) converge vers le point fixe.

Exercice 3.5. (Question de d'introduction à la recherche) Supposons que l'itération est faite numériquement, que dans chaque itération s'introduit par erreur d'arrondi dans l'évaluation de f .

Les questions naturelles sont : Est-ce que la suite (x_n) converge tout de même vers x^* ? Ou bien, reste-t-elle au moins dans une petite boule autour de x^* à partir d'un certain rang?

Discuter ces questions si on modélise l'erreur par

- a) $x_{n+1} = f(x_n) + \varepsilon_n$ avec $\varepsilon_n \leq \varepsilon$ pour un $\varepsilon > 0$.
- b) $x_{n+1} = f(x_n) + \varepsilon_n$ avec $\varepsilon_n \leq C \|x_n - x^*\|$ pour un $C > 0$.
- c) $x_{n+1} = f(x_n + \varepsilon_n)$, avec $\varepsilon_n \leq \varepsilon$.

Il faut refaire les calculs d'étape 1 de la preuve et voir si on peut "tracer" la propagation d'erreur. S'estompe-t-elle toute seule?

Exercice 3.6. (Question de d'introduction à la recherche 2) Soient (t_n) et (ε_n) des suites dans $[0, \infty)$. Soit $0 \leq L < 1$. Montrer que si $t_{n+1} \leq Lt_n + \varepsilon_n$, alors $t_n \rightarrow 0$.

S'en servir pour montrer le théorème suivant : Soit $0 \leq L < 1$ et E un e.v.n. Soit $f : E \rightarrow E$ une application admettant un point fixe $x^* = f(x^*)$ (ici c'est une hypothèse!). Si $\|x^* - f(x)\| \leq L \|x^* - x\|$ pour tout $x \in E$, alors les itérations successives (avec x_0 quelconque) $x_{n+1} = f(x_n)$ convergent vers x^* . L'idée est donc que "contraction vers le point fixe" suffit. Supposons que l'itération est perturbé numériquement : à la place de la suite (x_n) nous obtenons une suite (z_n) avec $\varepsilon_n = \|x_n - z_n\| \rightarrow 0$ quand $n \rightarrow \infty$. Montrer que $z_n \rightarrow x^*$ tout de même.

3.1.1 Application : google page rank

Considérons l'internet comme un graphe G de taille N , où chaque page est un noeud, et les arêtes les liens. Le modèle part du principe que chaque lien est suivi de la même probabilité (peu importe sa place sur la page, le layout, etc). Soit $s(x)$ le nombre de liens sortants depuis la page x . Alors la probabilité "de transition" q_{xy} de passer de la page x à y via un lien est

$$\begin{cases} 1/s(x) & \text{si la page } x \text{ possède un lien vers } y \\ 0 & \text{sinon} \end{cases}$$

Soit $f(x)$ le probabilité qu'une marche aléatoire sur le graphe visite la page x . Alors, puisque la seule façon de visiter x est de passer par un lien entrant,

$$f(x) = \sum_{y \in G} f(y)q_{yx}$$

En écriture matricielle, $f = Qf$. Le "vecteur d'importance" qui décide de l'ordre de présentation de résultats est simplement la liste des pages (noeuds) ordonné par probabilité décroissante. Plusieurs problèmes se posent. Une page pourrait être isolé, sans liens entrants, ou bien sans liens sortants, l'internet pourrait être disconnexe (plusieurs îlots qui ne se "voient pas"). Alors la matrice Q initiale est trafiqué un peu : d'abord, google relie vers toute page, est sera relie par toute page : cela signifie de poser la valeur $1/N$ dans la colonne et ligne qui correspond à google. d'un coup le graphe représenté devient connexe! On écrira toujours Q pour cette matrice trafiqué.

Ensuite, on pose $T = \lambda Q + (1 - \lambda)E$ où $\lambda \in (0, 1)$ et où E est la matrice constant à $1/N$ sur tous les entrées. Ceci ne change pas la propriété d'avoir une matrice "probabilités de transition". Mais peu importe comme on choisit λ , T devient une application strictement contractive :

$$\|Tx - Ty\| = \lambda \|Q(x - y)\| \leq \lambda \|x - y\|$$

Reste à trouver un sous-ensemble fermé de \mathbb{R}^N qui est laissé invariant par T . Le choix naturel est

$$\Omega = \{x \in \mathbb{R}^N : \forall_i : x_i \geq 0 \text{ et } \sum_i x_i = 1\}$$

donc tous les vecteurs de "probabilité de visite" sur un graphe de taille N ... C'est un fermé (preuve?!). Puisque T (et Q) n'ont que des entrées positives, $|Qx| \leq Q|x|$ où on note $|x| = (|x|_1, |x|_2, \dots)$. Il suit que

$$\sum_i (Tx)_i = \langle \mathbb{1}, Tx \rangle = \langle T^t \mathbb{1}, x \rangle = \langle \mathbb{1}, x \rangle = \sum_i x_i = 1$$

et donc $T : \Omega \rightarrow \Omega$ (pareil pour Q). Le théorème du point fixe s'applique et montre qu'il existe un et un seul point fixe f tel que $f = Tf$. Le calcul peut se faire par itérations successives ou par autres méthodes. Quoi qu'il en soit, il a permis à google (il paraît que $\lambda = 0.8$ est un bon choix) de gagner des milliards de dollars ...

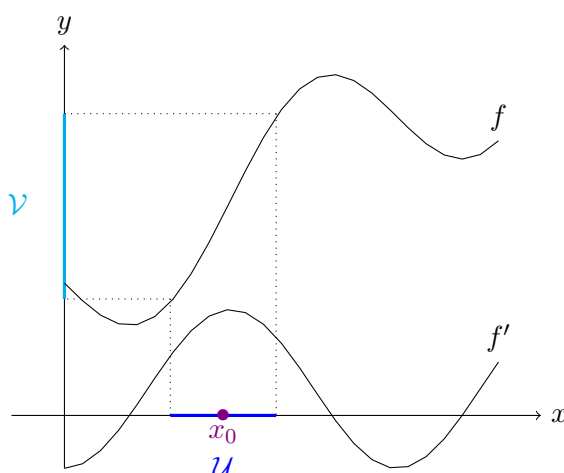
3.2 Le théorème de l'inverse locale

Situation en une dimension :

Proposition 3.7. *Soit*

$$f : (a, b) \rightarrow \mathbb{R}$$

de classe \mathcal{C}^1 . Si en un point x_0 , $f'(x_0) \neq 0$, alors il existe \mathcal{U} un voisinage de x_0 et \mathcal{V} , voisinage de $f(x_0)$ tel que $f : \mathcal{U} \rightarrow \mathcal{V}$ est bijective.



Démonstration. $f'(x_0) \neq 0$. Alors $f'(x_0) > 0$ ou bien $f'(x_0) < 0$. On va supposer $f'(x_0) > 0$ (les deux situations se traitant symétriquement) et montrer qu'alors il existe $r > 0$ tel que $f'(x) > 0$ sur $(x_0 - r, x_0 + r)$. f' est continue, on fixe $\frac{f'(x_0)}{2} = \varepsilon$. Alors il existe $r > 0$ tel que $|x - x_0| < r \Rightarrow |f'(x) - f'(x_0)| < \varepsilon$. Or $f'(x) = f'(x_0) + f'(x) - f'(x_0) \geq f'(x_0) - \varepsilon = \frac{f'(x_0)}{2} > 0$. (i.e : une fonction strictement positive dans un point est strictement positive dans un certain voisinage de ce point). En conclusion, f est strictement monotone donc elle est injective. De plus $\mathcal{V} = f(\mathcal{U})$ est un intervalle (par le T.V.I.) donc f est bijective. \square

Remarque 3.8. *Garde ! On a montré que “si $f' \neq 0$, alors f est inversible”. Ceci ne veut pas dire “si $f'(x) = 0$ alors f n'est pas inversible”. Pensez à $f(x) = x^3$ dans \mathbb{R} qui est strictement croissant, donc inversible (sur tout \mathbb{R} , pourtant, $f'(0) = 0$).*

Exercice 3.9. Appliquer la proposition à $\sin(\cdot)$ sur $[-\pi/2, \pi/2]$, à $\cos(\cdot)$ sur $[0, \pi]$ et à $\tan(\cdot)$ sur $(-\pi/2, \pi/2)$. Qu'obtenez vous ?

Avant d'entamer le cas de plusieurs variables regardons une extension très utile de la série géométrique. On l'appliquera seulement sur \mathbb{R}^n mais la preuve ne change pas d'une virgule dans le c'est à dire re abstrait :

Lemme 3.10. (La série de Neumann*) Soit E un e.v.n. complet et $A = E \rightarrow E$ linéaire telle que $\|A\| = \sup\{\|Ax\|_E : \|x\| \leq 1\} \leq R < 1$. Alors $(\text{Id} - A)$ est inversible,

$$(\text{Id} - A)^{-1} = \sum_{n=0}^{\infty} A^n \quad \text{et donc} \quad \|(\text{Id} - A)^{-1}\| \leq \frac{1}{1 - \|A\|}$$

Démonstration. On pose

$$B_N = \text{Id} + \sum_{n=1}^N A^n$$

*. d'après Karl Gottfried Neumann (7 mai 1832 à Königsberg, mort le 27 mars 1925 à Leipzig), un mathématicien allemand.

Or

$$B_N(\text{Id} - A) = \text{Id} + \sum_{n=1}^N A^n - (A + \sum_{n=1}^N A^n \cdot A) = \text{Id} + A - A - A^2 + A^2 + \dots = \text{Id} - A^{N+1}$$

(somme télescopique). On obtient de même $(\text{Id} - A)B_N = \text{Id} - A^{N+1}$, donc $(\text{Id} - A)B_N = B_N(\text{Id} - A)$. Vu que

$$\|A^N\| \leq \|A\|^N = R^N \longrightarrow 0 \quad \text{quand } N \rightarrow \infty$$

la droite de cette équation tend vers Id , la gauche converge donc également, vers $B(\text{Id} - A) = (\text{Id} - A)B$ où $B = \sum_{n \geq 0} A^n$. Il suit que $B = (\text{Id} - A)^{-1}$! (ceci ressemble beaucoup à une suite géométrique avec des matrices à la place des scalaires). Finalement, on conclut par

$$\|(\text{Id} - A)^{-1}\| \leq \sum_{n=0}^{\infty} \|A^n\| \leq \sum_{n=0}^{\infty} \|A\|^n = \frac{1}{1 - \|A\|}. \quad \square$$

Théorème 3.11. (Théorème d'inversion locale) Soit $\Omega \subseteq \mathbb{R}^n$ ouvert ; $f : \Omega \rightarrow \mathbb{R}^n$ de classe C^r pour un $r \geq 1$. Soit $x_0 \in \Omega$ un point tel que $D_f(x_0)$ est inversible dans $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$.

Alors il existe un ouvert $\mathcal{U} \subseteq \Omega$ contenant x_0 et un ouvert $\mathcal{V} \subseteq \mathbb{R}^n$ contenant $f(x_0)$ tels que :

- i) L'application $f|_{\mathcal{U}} : \mathcal{U} \rightarrow \mathcal{V}$ est bijective. Soit g son inverse.
- ii) g est de classe C^1 . De fait (mais hors programme), g est de classe C^r , i.e. la régularité de l'inverse g est celle de f .
- iii) $D_g(f(x_0)) = (D_f(x_0))^{-1}$.

Exercice 3.12. Avant de lire la preuve, examiner le cas où f est linéaire, i.e. $f(x) = A \cdot x$, la multiplication matricielle avec une matrice carrée A . Que signifient les hypothèses ? Les conclusions ? Le théorème fournit une version "locale" de ce cas, si f est au moins continûment différentiable.

Démonstration. Étape 1 : existence d'un inverse, par un argument de point fixe

On pose $A = D_f(x_0)$. Par hypothèse $f_0 = A^{-1}f$ existe. f_0 assure : $D_{f_0}(x_0) = \text{Id}$. Par translation on se ramène au point 0 : $f_1(x) = f_0(x + x_0) - f_0(x_0)$ On a alors $f_1(0) = 0$ et $D_{f_1}(0) = \text{Id}$ On va "remplacer" f par f_1 et supposer $D_f(0) = \text{Id}$ et $f(0) = 0$. (Il suffira pour retrouver f de faire la translation dans l'autre sens).

Soit $h(x) = x - f(x)$; $h(0) = 0$ et $D_h(0) = \text{Id} - \text{Id} = 0$ (l'application nulle). Par continuité (hypothèse !) de D_h , il existe un $r > 0$ tel que $\|D_h(x)\|_{2 \rightarrow 2} \leq 1/2$ pour $x \in B[0, r]$ c'est à dire $\|x\| \leq r$ †. Par l'inégalité des accroissements finis (2.3), on a

$$\|h(x) - h(y)\|_2 \leq (\max_{\|\xi\| \leq r} \|D_h(\xi)\|_{2 \rightarrow 2}) \cdot \|x - y\|_2 \leq 1/2 \|x - y\|_2$$

pour $x, y \in B[0, r]$. Soit maintenant $\|y\| \leq \frac{r}{2}$. On introduit $h_y(x) = y + x - f(x)$ et pose : $F = B[0, r]$ (pour retrouver la notation du théorème du point fixe) et $x, z \in F$. Alors

$$\|h_y(x) - h_y(z)\| \leq 1/2 \|x - z\| \leq 1/2(\|x\| + \|z\|) \leq r.$$

†. Pour rappel on note : $\|L\|_{p \rightarrow q} = \max\{\|Lx\|_q ; \|x\|_p = 1\}$; ici $p = q = 2$, c'est la norme d'opérateur induite par la norme euclidienne

Donc F est stable sous $h_y : h_y : F \rightarrow F$. De plus $\|D_{h_y}(x)\| = \|D_h(x)\| \leq L = 1/2$ pour $x \in F$: h_y est strictement contractante sur F . En conséquence, par le théorème du point fixe : $\exists! x_y$ tel que $h_y(x_y) = x_y \Leftrightarrow y + x_y - f(x_y) = x_y \Leftrightarrow f(x_y) = y$. La notation x_y nous rappelle que le point fixe x_y dépend de y car h_y dépend de y . On définit une fonction g pour tout point de $B[0, \frac{r}{2}]$ par $g(y) = x_y$. C'est à dire : pour tout y on applique le théorème du point fixe pour trouver $g(y)$! Alors nous avons

$$f(g(y)) = f(x_y) = y \quad \text{c'est à dire} \quad f \circ g = \text{Id}$$

et $g(y) = x_y$ donne par application de f que

$$g(f(y)) = f(x_y) = y \quad \text{c'est à dire} \quad g \circ f = \text{Id}$$

On a donc construit une fonction inverse ! Maintenant il faut montrer qu'elle a toutes les propriétés souhaitées :

Étape 2 : continuité de g :

On va montrer que g est Lipschitzienne (donc continue). Rappel : $g(y) = x$; $f(x) = y$

$$\begin{aligned} \|x - \tilde{x}\| &= \|x - f(x) + f(x) - (\tilde{x} - f(\tilde{x}) + f(\tilde{x}))\| \\ &\leq \|f(x) - f(\tilde{x})\| + \|h(x) - h(\tilde{x})\| \\ &\leq \|f(x) - f(\tilde{x})\| + 1/2 \|x - \tilde{x}\| \end{aligned}$$

Vu que le terme de gauche apparaît avec facteur $1/2$ à droite, on peut les assembler et on obtient

$$\|x - \tilde{x}\| \leq 2 \|f(x) - f(\tilde{x})\|$$

C'est à dire :

$$\|g(y) - g(\tilde{y})\| \leq 2 \|f(g(y)) - f(g(\tilde{y}))\| = 2 \|y - \tilde{y}\|$$

Ainsi, g est 2-Lipschitzienne, donc continue.

Étape 3 : différentiabilité de g :

On choisit $\|y\|, \|\tilde{y}\| \leq \frac{r}{2}$. On veut montrer que $\|g(y) - g(\tilde{y}) - (D_f(x))^{-1}(y - \tilde{y})\|$ est $o(\|y - \tilde{y}\|)$ pour démontrer la différentiabilité de g et trouver en même temps la différentielle. On note $g(y) = x$, $g(\tilde{y}) = \tilde{x}$.

$$\begin{aligned} &\|g(y) - g(\tilde{y}) - [D_f(x)]^{-1}(y - \tilde{y})\| \\ &= \|g(y) - g(\tilde{y}) - [D_f(x)]^{-1}(f(x) - f(\tilde{x}))\| \\ (\text{car } f \text{ diff.able}) &= \|g(y) - g(\tilde{y}) - [D_f(x)]^{-1}((D_f(x))(x - \tilde{x}) + r_f(x - \tilde{x}))\| \\ &= \|g(y) - g(\tilde{y}) - (x - \tilde{x}) - [D_f(x)]^{-1}r_f(x - \tilde{x})\| \\ &= \|[D_f(x)]^{-1}r_f(x - \tilde{x})\| \end{aligned}$$

Mais puisque $\|x\| \leq r$ on a

$$\|[D_f(x)]^{-1}r_f(x - \tilde{x})\| \leq \max\{\|[D_f(\xi)]^{-1}\|_{2 \rightarrow 2}; \|\xi\| \leq r\} \cdot \|r_f(x - \tilde{x})\|.$$

Rappelons le choix de $r > 0$ du début : sur $B[0, r]$, $D_h = \text{Id} - D_f$ est de norme au plus $1/2$. Ainsi, par série de Neumann, voir Lemme 3.10, $D_f = \text{Id} - D_h$ est inversible sur $B[0, r]$ et $\|[D_f(x)]^{-1}\| \leq 2$ pour tout $\|x\| \leq r$. Il suit que

$$\|g(y) - g(\tilde{y}) - (D_f(x))^{-1}(y - \tilde{y})\| \leq 2 \|r_f(x - \tilde{x})\| \leq 4 \frac{\|r_f(x - \tilde{x})\|}{\|x - \tilde{x}\|} \cdot \|y - \tilde{y}\|$$

car $\|x - \tilde{x}\| \leq 2\|y - \tilde{y}\|$ par l'étape 2. Or $\frac{\|r_f(x-\tilde{x})\|}{\|x-\tilde{x}\|} \rightarrow 0$ on conclut que g est différentiable et que, de plus sa différentielle est $D_g(y) = (D_f(x))^{-1}$.

Étape 4 : continuité de la différentielle

On souhaite montrer que $y \rightarrow D_g(y)$ est continue, et que la linéarité de f entraîne celle de g . On a : $D_g(y) = (D_f(x))^{-1} = (D_f(g(y)))^{-1}$ ce qui se traduit par la composition : $D_g(y) = \text{inv} \circ D_f \circ g$, où $\text{inv} : M \mapsto M^{-1}$ est l'inversion.

Tout d'abord on a vu que g est continue (car 2-Lipschitz : il suffit de prendre $\delta = \frac{\varepsilon}{L}$, ici $L = 2$). De plus, D_f est continue par hypothèse. Reste à montrer que inv est continue. Nous savons que \det est polynomial dans les coefficients d'une matrice, du coup le développement de M^{-1} en comatrice montre la continuité. Il y a une autre preuve, un peu plus abstraite, et plus puissante. Soit M inversible. Par la série de Neumann, pour tout $\|H\| < \frac{1}{\|M^{-1}\|}$,

$$\begin{aligned} M^{-1} - (M + H)^{-1} &= M^{-1} - [M(\text{Id} + M^{-1}H)]^{-1} \\ &= \left[\text{Id} - (\text{Id} - (-M^{-1}H))^{-1} \right] M^{-1} \\ &= \left[\sum_{n=1}^{\infty} (-1)^n (M^{-1}H)^n \right] M^{-1} \end{aligned} \tag{3.1}$$

Il suit que

$$\|M^{-1} - (M + H)^{-1}\| \leq \left(\sum_{n=1}^{\infty} (\|M^{-1}\| \|H\|)^n \right) \|M^{-1}\| = \|H\| \left(\frac{\|M^{-1}\|^2}{1 - \|M^{-1}\| \|H\|} \right)$$

On voit que $H \rightarrow 0$ implique $\|M^{-1} - (M + H)^{-1}\| \rightarrow 0$. □

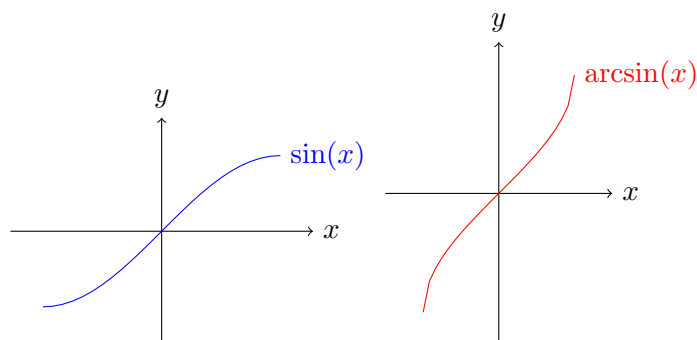
Remarque 3.13. En fait, le développement en série de l'inverse, à savoir

$$(M + H)^{-1} = \left[\sum_{n=0}^{\infty} (-1)^{n+1} (M^{-1}H)^n \right] M^{-1}$$

qui découle de (3.1), montre que inv est de classe C^∞ sur l'ouvert $\text{GL}_n = \det^{-1}(\mathbb{R}^*) \subset \text{Mat}_{n \times n}$, mais comme nous n'avons pas discuté les différentielles d'ordre supérieur, ceci est hors programme pour la Licence 2. Tout comme l'idée de réitérer le procédé et de proche en proche pour obtenir : $f \in \mathcal{C}^r \Rightarrow g \in \mathcal{C}^r$.

Exemple 3.14. $f(x) = \sin(x)$. $f \in \mathcal{C}^\infty$; $D_f : \lambda \rightarrow f'(x).\lambda$ est inversible si et seulement si $f'(x) \neq 0$, c'est à dire pour $x \neq k.\pi/2, k \in \mathbb{Z}^*$. On peut donc définir une fonction inverse sur un voisinage de tout $x \neq k.\pi/2$

Astuce pour avoir une idée de l'inverse d'une fonction : regarder dans un miroir ou à travers d'une feuille où on a dessiné la fonction et faire pivoter la feuille afin d'avoir les axes "dans le bon sens"! Un exemple($\sin(x)$ et $\arcsin(x)$ au voisinage de 0.



Exemple 3.15. $f(x, y) = \begin{pmatrix} e^x \sin(y) \\ e^x \cos(y) \end{pmatrix}$. Les deux fonctions coordonnées sont \mathcal{C}^∞ donc $f \in \mathcal{C}^\infty$. La différentielle est donc représentée par la matrice Jacobienne. Or

$$\mathcal{J}_f(x, y) = \begin{pmatrix} e^x \sin(y) & e^x \cos(y) \\ e^x \cos(y) & -e^x \sin(y) \end{pmatrix} \quad \det(\mathcal{J}_f(x, y)) = -e^{2x} \cdot 1 < 0$$

Donc autour de tout point (x, y) il existe un voisinage \mathcal{U} sur lequel f est bijective sur $f(\mathcal{U}) = \mathcal{V}$. Attention cependant : f est 2π -périodique par rapport à sa deuxième variable, donc non injective sur \mathbb{R}^2 entier.

Exemple 3.16. $f(x, y) = \begin{pmatrix} e^{xy} - e^x + 1 \\ e^x - e^y \end{pmatrix}$. $f \in \mathcal{C}^\infty$ on peut donc calculer D_f représentée par la Jacobienne.

$$\mathcal{J}_f(x, y) = \begin{pmatrix} ye^{xy} - e^x & xe^{xy} \\ e^x & -e^y \end{pmatrix}.$$

On s'intéresse au point $a = (0, 0)$

$$\det(\mathcal{J}_f(0, 0)) = \begin{vmatrix} -1 & 0 \\ 1 & -1 \end{vmatrix} = 1 \neq 0.$$

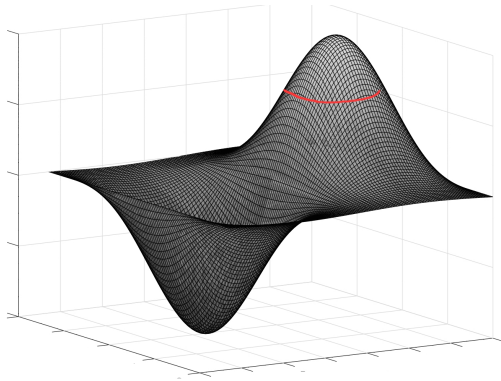
Donc on peut inverser f autour de $(0, 0)$. Il existe \mathcal{U} un voisinage de $(0, 0)$ et \mathcal{V} voisinage de $f(0, 0) = (1, 0)$ tel que $f : \mathcal{U} \rightarrow \mathcal{V}$ est bijective. On appelle $f^{-1} = g$. $D_g(f(0, 0)) = D_g(1, 0) = [D_f(0, 0)]^{-1}$ est donnée par

$$[\mathcal{J}_f(0, 0)]^{-1} = \begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} -1 & 0 \\ -1 & -1 \end{pmatrix} \quad \text{on a donc} \quad \mathcal{J}_g(1, 0) = \begin{pmatrix} -1 & 0 \\ -1 & -1 \end{pmatrix}$$

Remarque 3.17. On peut ainsi calculer des dérivées supérieures : on a $f \circ g = Id$ puis $D_f(g) \circ D_g = Id$. En re-dérivant cette expression on obtient des dérivées supérieures.

3.3 Le théorème des fonctions implicitement définis

L'idée : soit une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. L'ensemble $\{(x, y, z); z = f(x, y)\}$ (donc le graphe de f) est représenté par une surface, comme la surface d'une montagne : chaque ligne de niveau serait une courbe pour un certain z fixé : Question : en un point, puis-je trouver une courbe isocline, c'est à dire où la hauteur est constante ? Et est-ce que je peux caractériser cette courbe ?



Généralisons : Soit $\Omega \subseteq \mathbb{R}^n$ un ouvert, $m < n$ et $f : \Omega \rightarrow \mathbb{R}^m$ et un "ensemble de niveaux" (qui n'est pas nécessairement une ligne) :

$$N_c = \{x \in \Omega \text{ tel que } f(x) = c\}.$$

En se plaçant dans l'exemple, où $n=2$ et $k=1$ cela nous donnera la ligne de niveau "locale" pour $f(x, y) = c$. On cherche à effectuer une résolution locale : de certaines variables par rapport aux autres : on a $x = (x_1, \dots, x_n)$. On pose $x = (x', x'')$ où $x' = (x_1, \dots, x_k)$ et $x'' = (x_{k+1}, \dots, x_n)$. On cherche alors une application "locale", $\gamma : \mathcal{U} \rightarrow \mathbb{R}^{n-k}$ définie dans le voisinage \mathcal{U} d'un point x'_0 tel que

$$f(x', \gamma(x')) = c.$$

Le cas des applications linéaires

Exemple 3.18. Résoudre certaines variables par rapport à d'autres peut être une démarche "perdue d'avance", par des arguments de dimension (théorème du rang). Commençons avec un exemple : soit

$$f(x, y, z) = x + y + z,$$

qui est une fonction linéaire. L'équation $f(x, y, z) = 42$ définit un hyperplan affine $H = N_{42}$: il ne pourra pas être complètement décrit, même localement, avec deux fonctions g et h dépendant d'un seul paramètre : quoi qu'on fasse avec g et h , l'ensemble $\{(t, g(y), h(t)) : t \in \mathbb{R}\}$ ne sera qu'une courbe (un-dimensionnelle) dans l'hyperplan (deux-dimensionnel). En revanche, on peut facilement résoudre *une seule* des 3 variables par rapport aux *deux autres* : par exemple,

$$f(x, y, 42 - x - y) = 42 \quad \text{pour tout } (x, y) \in \mathbb{R}^2$$

Exemple 3.19. Soit $g : \mathbb{R}^3 \rightarrow \mathbb{R}^2$. Pour rester dans le cas linéaire, prenons

$$g(x, y, z) = \begin{pmatrix} x + y + z \\ x - y \end{pmatrix}$$

Alors le système d'équations $g(x, y, z) = c = (42, 36)$ est un système de deux équations (indépendantes) pour 3 inconnues. L'ensemble des solutions, intersection de deux hyperplans, est

donné par une droite affine, notamment

$$\gamma(t) = \begin{pmatrix} 32 \\ -4 \\ 14 \end{pmatrix} + t \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \quad t \in \mathbb{R}$$

qui se paramétrise facilement en fonction d'une seule des trois variables : par exemple

$$g(x, x - 36, -2x + 78) = (42, 36)^t \quad \forall x \in \mathbb{R}$$

ce qui résoud (y, z) par rapport à x . De même on peut résoudre y par rapport à (x, z) ou z par rapport à (x, y) .

La situation observée dans les deux exemples est générique pour des applications linéaires : soit A une matrice $m \times (k + m)$ de plein rang (rang m) et $f(x) = A.x$. Le "plein rang" sert à assurer $f(\mathbb{R}^{k+m}) = \mathbb{R}^m$, i.e. à atteindre tout $c \in \mathbb{R}^m$ comme valeur sous f . Quitte à permuter des colonnes, nous pouvons alors supposer que

$$A = \left(\begin{array}{ccc|ccc} a_{1,1} & \cdots & a_{1,k} & a_{1,k+1} & \cdots & a_{1,k+m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,k} & a_{m,k+1} & \cdots & a_{m,k+m} \end{array} \right) = (A_1|A_2)$$

avec la deuxième matrice, A_2 , inversible (elle est de rang m et de taille $m \times m$). Alors pour tout $c \in \mathbb{R}^m$ et tout $x' \in \mathbb{R}^k$ il existe un (unique) $x'' \in \mathbb{R}^m$ tel que

$$A \begin{pmatrix} x' \\ x'' \end{pmatrix} = A_1 x' + A_2 x'' = c$$

notamment $x'' = A_2^{-1}(c - A_1 x')$. Ainsi, si $f(x', x'') = A(x', x'')^t$,

$$\forall x' \in \mathbb{R}^k : \quad f(x', A_2^{-1}(c - A_1 x')) = c.$$

La variable "plein rang" (ici x'') permet donc une "résolution", on peut l'écrire comme fonction de l'autre variable (ici x'). Puisque la structure est linéaire, on peut résoudre "globalement", sur tout \mathbb{R}^k , x'' comme fonction de x' . Cela change dans

Le cas général.

Le cas d'une fonction $f \in \mathcal{C}^1$ générale (non-linéaire) est évidemment plus complexe ; entre autre on ne peut pas espérer une résolution "globale". Le théorème suivant est frappant : il montre que si la *linéarisation* $A = D_f(a)$ d'une fonction f autour d'un point a permet une résolution de certaines variables (accumulées par permutation en x'' , comme ci-dessus) par rapport aux autres (les variables de x'), alors la résolution de x'' comme fonction de x' reste est vraie pour la fonction f , au moins dans un voisinage de a .

Théorème. 3.20. *Soit $\mathcal{U} \subseteq \mathbb{R}^k$, et $\mathcal{V} \subseteq \mathbb{R}^m$ deux ouverts, $f : \mathcal{U} \times \mathcal{V} \rightarrow \mathbb{R}^m$ une fonction de classe \mathcal{C}^r avec $r \geq 1$, et $(a, b) \in \mathcal{U} \times \mathcal{V}$ tels que $f(a, b) = c$ (ici $a \in \mathbb{R}^k; b \in \mathbb{R}^m; c \in \mathbb{R}^m$). Si $D_2 f(a, b) : \mathbb{R}^m \rightarrow \mathbb{R}^m$ est inversible, il existe un voisinage $\mathcal{U}_0 \subseteq \mathcal{U}$ de a et une fonction $g : \mathcal{U}_0 \rightarrow \mathcal{V}$ de classe \mathcal{C}^r tels que $g(a) = b$ et $f(x, g(x)) = c$. Autrement dit, $\{(x, g(x)) : x \in \mathcal{U}_0\} \subseteq N_c$. Pour \mathcal{U}_0 suffisamment petit, g est unique.*

Il est important que les variables à résoudre en utilisant les autres soient de la même dimension (ici m) que celle de l'espace d'arrivée, déjà de façon pratique, car sinon, la matrice Jacobienne \mathcal{J} n'est pas quadratique, et donc pas inversible.

Démonstration. On peut supposer $c = 0$ (au pire on prend $f(x, y) - c = 0$). Soit $A = (D_2f)(a, b)$. En remplaçant f par $A^{-1} \cdot f$ on peut supposer $(D_2f)(a, b) = \text{Id}$. Soit

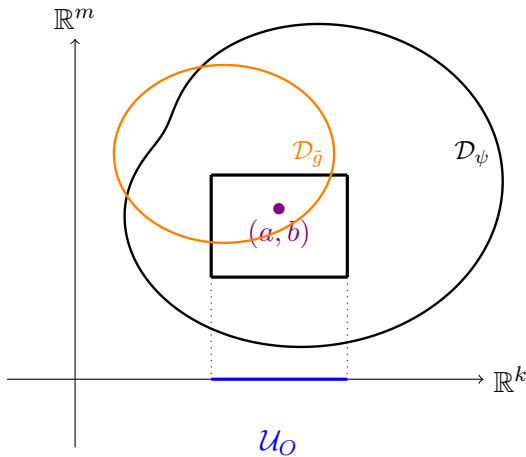
$$\varphi : \begin{cases} \mathcal{U} * \mathcal{V} \rightarrow \mathbb{R}^k * \mathbb{R}^m \\ (x, y) \rightarrow (x, f(x, y)) \end{cases}$$

Or $f \in \mathcal{C}^r$ et $r \geq 1$, il en est de même pour φ et D_φ est représentée par

$$\mathcal{J}_\varphi = \left(\begin{array}{c|c} I_k & 0 \\ \hline (*) & A = I_m \end{array} \right)$$

à l'effet que $\det \mathcal{J}_\varphi = \det I_k * \det I_m = 1 \neq 0$ donc \mathcal{J}_φ est inversible. Par le théorème de l'inversion locale, il existe un voisinage de (a, b) dans $\mathbb{R}^k \times \mathbb{R}^m$ sur lequel φ est inversible. Soit ψ l'inverse locale de φ . Cette fonction s'écrit : $\psi(x, y) = (x, h(x, y))$ avec $x \in \mathbb{R}^k$ et $y \in \mathbb{R}^m$. On pose $g(x) = h(x, 0)$. Puisque f et donc φ sont \mathcal{C}^r par le théorème d'inversion locale $\psi \in \mathcal{C}^r$; on a ainsi $h \in \mathcal{C}^r$ et donc $g \in \mathcal{C}^r$. De plus, $(x, f(x, g(x))) = \varphi(x, g(x))$ par construction de φ . En utilisant $\varphi \circ \psi = \text{Id}$ on a donc

$$\varphi(x, g(x)) = \varphi(x, h(x, 0)) = \varphi(\psi(x, 0)) = (x, 0)$$



Ce qui montre que $f(x, g(x)) = 0 \forall x \in \mathcal{U}_0$, voisinage de (a, b) inclus dans le domaine de définition de ψ . On vient de démontrer l'existence d'une fonction f telle que $f(x, g(x)) = 0$ (la constante c du début).

Unicité de cette fonction : Soit \tilde{g} une autre fonction telle que $f(x, \tilde{g}(x)) = 0$. Alors $(x, f(x, \tilde{g}(x))) = \varphi(x, \tilde{g}(x)) = 0 = \varphi(x, g(x))$. En appliquant ψ on trouve $g = \tilde{g}$ sur l'intersection de leurs domaines de définition. \square

Remarque 3.21. Pour la clarté de la preuve nous avons “ordonné” les variables x_{k+1}, \dots, x_{k+m} à résoudre par rapport aux variables (x_1, \dots, x_k) en deux “blocs”. Ceci n'est évidemment pas limitant quant à l'utilisation. Si, par exemple $f(x, y, z) = x^2 + \sin(y) + \cos(z)$, $\nabla f(0, 0, 0) = (0, 1, 0)$, on peut donc exprimer y en fonction de (x, z) .

3.3.1 Dérivation implicite

Dès que $h(x) = f(x, g(x)) = c$ sur \mathcal{U}_0 on a $D_h(x) = 0$ sur \mathcal{U}_0 . Mais :

$$D_h(x) = (D_1f)(x, g(x)) + (D_2f)(x, g(x)) \circ D_g(x)$$

d'où :

$$D_g(x) = -[(D_2f)(x, g(x))]^{-1} \cdot (D_1f)(x, g(x))$$

Exemple 3.22. Soit

$$f : \begin{cases} \mathbb{R}^2 \rightarrow \mathbb{R} \\ (x, y) \rightarrow x^2 - 3xy + y^3 - 7 \end{cases}$$

On a $f(4, 3) = 0$ et $(\nabla f)(x, y) = (2x - 3y, -3x + 3y^2)$, donc $(\nabla f)(4, 3) = (-1, 15)$. La dérivée par rapport à la seconde variable est non nulle, donc d'après le théorème des fonctions implicitement définies il existe un voisinage de 4 ($\mathcal{U}_0 = (4 - \delta, 4 + \delta)$) et une fonction $g : \mathcal{U}_0 \rightarrow \mathbb{R}$, \mathcal{C}^∞ (car $f \in \mathcal{C}^\infty$) avec $g(4) = 0$ et $f(x, g(x)) = 0$. On a donc

$$f_x(x, g(x)) + f_y(x, g(x)) \cdot g'(x) = 0, \forall x \in \mathcal{U}_0$$

d'où $g'(4) = -(-1) * \frac{1}{15} = \frac{1}{15}$ (car $(f_y(4, 3))^{-1} = \frac{1}{15}$). On re-dérive, on obtient l'expression :

$$f_{xx}(x, g(x)) + f_{xy}(x, g(x)) \cdot g'(x) + f_{xy}(x, g(x)) \cdot g'(x) + f_y(x, g(x)) \cdot g''(x) + f_{yy}(x, g(x)) \cdot g'(x) \cdot g'(x) = 0$$

ce qui donne $g''(4)$:

$$g''(4) = -(f_{xx}(4, 3) + f_{xy}(4, 3)g'(4) + f_{xy}(4, 3) \cdot g'(4) + f_{yy}(4, 3 \cdot g'(4)^2)) \cdot (f_y(4, 3))^{-1} = \frac{14}{125}$$

Ceci permet d'avoir une bonne approximation de g autour de 4, en effet, on utilise un développement de Taylor, et, connaissant $g'(4)$ et $g''(4)$ on a :

$$g(4 + h) = 3 + \frac{h}{15} + \frac{h^2}{2} * \frac{14}{125} + o(h^2).$$

Exemple 3.23. Soit

$$f(x, y, u, v) = \begin{pmatrix} x^2 + 2x - 2y + u \cos(x) + \arctan(v) \\ xy + \cos(xv) - e^{u^2} + (x-1)u \end{pmatrix}$$

so that

$$\mathcal{J}_f = \begin{pmatrix} 2x + 2 - u \sin(x) & -2 & \cos(x) & \frac{1}{1+v^2} \\ y - v \sin(xv) + u & x & -2ue^{u^2} - 1 & -x \sin(xv) \end{pmatrix}$$

Au point $(0, 0, 0, 0)$ on obtient

$$\mathcal{J}_f(0, 0, 0, 0) = \left(\begin{array}{cc|cc} 2 & -2 & 1 & 1 \\ 0 & 0 & -1 & 0 \end{array} \right)$$

où $D_1 f(0, 0, 0, 0)$ est représentée par $\begin{pmatrix} 2 & -2 \\ 0 & 0 \end{pmatrix}$ est non inversible. Mais $\begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}$ qui représente $D_2 f(0, 0, 0, 0)$ est, elle, inversible. Donc d'après le théorème il existe un voisinage \mathcal{U}_0 de $(0, 0)$ (les deux premières coordonnées) et une fonction $g : \mathcal{U}_0 \rightarrow \mathbb{R}^2$ telle que $g(0, 0) = (0, 0)$ (l'image des deux premières coordonnées du vecteur est égale aux deux autres, on se rappelle $g(x') = x''$) et $f(x, y, g(x, y)) = 0$. On calcule la jacobienne de g en $(0, 0)$:

$$\mathcal{J}_g(0, 0) = - \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 2 & -2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ -2 & 2 \end{pmatrix}$$

D'où :

$$g(0+x, 0+y) = g(0, 0) + \begin{pmatrix} 0 & 0 \\ -2 & 2 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} + o(\sqrt{x^2 + y^2}) = \begin{pmatrix} 0 \\ -2x + 2y \end{pmatrix} + o(\sqrt{x^2 + y^2})$$

3.4 Le théorème de l'immersion locale

Jusque là nous avons étudié le comportement local d'une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ dans les cas $n = m$ (inversion locale) et $n > m$ (fonctions implicitement définis). Reste un cas : $n < m$! Il découle comme corollaire des fonctions implicites :

Théorème. 3.24. *Soit $\Omega \subset \mathbb{R}^n$ un ouvert et $f : \Omega \rightarrow \mathbb{R}^{n+k}$ une fonction de classe \mathcal{C}^1 . Si en un point $a \in \Omega$, $\mathcal{J}_f(a)$ a plein rang, $f(U)$ est localement graphe d'une fonction. Précisément : Soit $f(a) = b = (b_1, b_2)$ avec $b_1 \in \mathbb{R}^n$ et $b_2 \in \mathbb{R}^k$. Il alors un existe un voisinage V de b_1 dans \mathbb{R}^n et une fonction $g : V \rightarrow \mathbb{R}^k$ tel que $f(U)$ est le graphe de g sur V .*

Démonstration. On écrit $f = (f_1, \dots, f_{n+k})$: les fonctions f_i sont les fonctions coordonnées de f . Quitte à les permuter on peut supposer que la matrice $n \times n$

$$A := \left(\frac{\partial f_i}{\partial x_k} \right)_{1 \leq i, k \leq n}$$

est inversible – ceci est l'hypothèse du “plein rang” : la matrice Jacobienne de f a $n + k$ lignes et n colonnes, et plein rang signifie qu'il existent bien n vecteurs lignes dans la Jacobienne de f qui soient indépendants : on les permute sur les n premiers coordonnées.

Notons alors $\pi_1 : \mathbb{R}^{n+k} \rightarrow \mathbb{R}^n$ la projection sur les n premiers coordonnées, et π_2 la projection sur les k derniers. Alors $h = (\pi_1 \circ f)$ satisfait que $\mathcal{J}_h(a) = A$, qui est inversible. Par le théorème de l'inversion locale, Théorème 3.11, il existe un voisinage U de a tel que $h : U \rightarrow h(U) \stackrel{\text{def.}}{=} V$ est bijective, et h et h^{-1} sont de classe \mathcal{C}^1 . Observons que $V = (h^{-1})^{-1}(U)$, donc V image réciproque d'un ouvert par l'application continue h^{-1} . Ainsi V est donc un voisinage ouvert de $h(a) = \pi_1(f(a)) = b_1$. Sur V on définit

$$g \stackrel{\text{def.}}{=} \pi_2 \circ f \circ h^{-1}$$

Observons que $\pi_1 \circ f \circ h^{-1}(x) = (\pi_1 \circ f)(\pi_1 \circ f)^{-1} = \text{Id}$ sur V ce qui donne pour $x \in V$

$$f(h^{-1}(x)) = (\pi_1(f(h^{-1}(x))), \pi_2(f(h^{-1}(x)))) = (x, g(x))$$

Ainsi, $f(U) = f(h^{-1}(V)) = \{(x, g(x)) : x \in V\}$, comme souhaité. □

4 Applications

4.1 Extrema sous contrainte

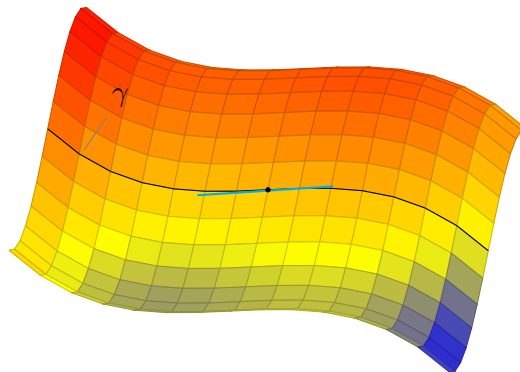
Soit $\Omega \subseteq \mathbb{R}^n$ un ouvert, $m < n$ et $f : \Omega \rightarrow \mathbb{R}^m$ une fonction. Soit $F : \Omega \rightarrow \mathbb{R}^m$ différentiable, et

$$M = \{x \in \Omega : F(x) = 0\}. \quad (4.1)$$

Le but est de maximiser/minimiser f sur un ensemble de points M . Par exemple quel est le maximum de $f(x, y) = x^2 + 3y^3$ pour les points $\{(x, y) \in \Omega; x^2 + y^2 = 1\}$ – ici $F(x, y) = x^2 + y^2 - 1$. on introduit un objet “géométrique” utile :

$$T_a M \stackrel{\text{def.}}{=} \{v \in \mathbb{R}^m : \exists \varepsilon > 0; \exists \gamma : (-\varepsilon, \varepsilon) \rightarrow M \text{ dérivable avec } \gamma(0) = a \text{ et } \gamma'(0) = v\}$$

$T_a M$ est appelé l’espace tangent à M au point a . Attention : “l’espace tangent” n’est pas l’ensemble des tangentes à M en un point, mais l’ensemble de ses “directions”. Il faudrait former $\{a + v : v \in T_a M\}$ pour obtenir l’ensemble des tangentes à M qui passent par a .



Cette translation est très utile, grâce à elle, $T_a M$ a une structure d’espace vectoriel : La fonction constante $\gamma(t) = a$ est dans M et passe par a , et donc $0 \in T_a M$.

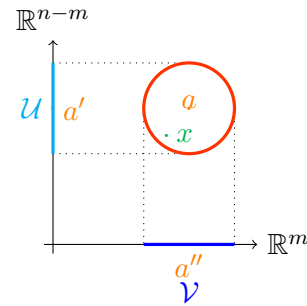
Soit $v \in T_a M$ et γ tel que $\gamma'(0) = v$. Regardons $\tilde{\gamma}(t) = \gamma(\lambda.t)$ on a $\tilde{\gamma}'(0) = \lambda(\gamma'(\lambda.t))(0) = \lambda.v \in T_a M$. Donc $T_a M$ est stable par la multiplication par un scalaire.

Soient $v, w \in T_a M$ avec les courbes correspondantes γ_v, γ_w . On introduit $\gamma = \frac{\gamma_v(2t) + \gamma_w(2t)}{2}$. Alors $\gamma(0) = \frac{a+a}{2} = a$ et $\gamma'(0) = v + w$, donc $v + w \in T_a M$: $T_a M$ est un sous-espace vectoriel de \mathbb{R}^n .

Le premier résultat est auxiliaire pour notre but d’extrema sous contrainte.

Proposition 4.1. Soit $\Omega \subseteq \mathbb{R}^n$ un ouvert, $m < n$, $F : \Omega \rightarrow \mathbb{R}^m$ différentiable. Soit M donné par (4.1) et $a \in M$. Supposons que D_F est continue en a telle que $D_f(a) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ est surjective. Alors $\ker(D_f(a)) = T_a M$.

Démonstration.



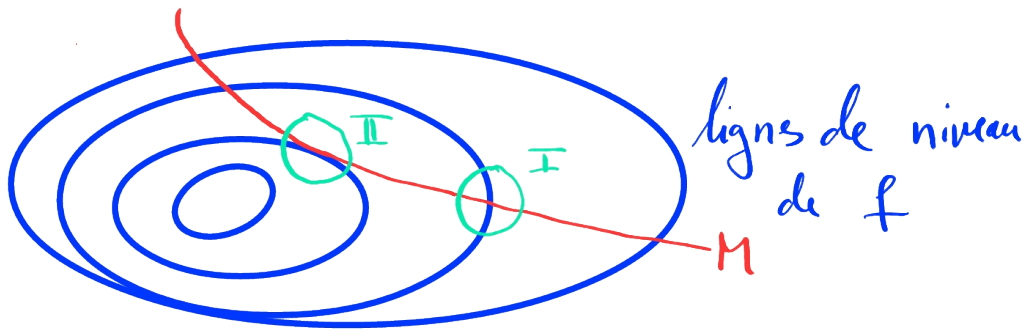
$u \in \mathbb{R}^{n-m}$ fixé. $\gamma(t) = (a' + tu, g(a' + tu))$ où $|t| < \varepsilon$ et où ε est choisi tel que $a' + tu \in \mathcal{U} \quad \forall |t| < \varepsilon$. On a donc $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$ et par construction $\gamma(0) = a$ et $\gamma'(0) = (u, D_g(a')(u)) \in T_a M$ par définition de $T_a M$. Soit

$$\Phi : \begin{cases} \mathbb{R}^{n-m} \rightarrow \ker(D_F(a)) \\ u \rightarrow (u, D_g(a')(u)) \in T_a M \subseteq \ker(D_F(a)) \end{cases}$$

Clairement, Φ est linéaire. L'image d'un espace vectoriel par Φ est donc un sous-espace du noyau de $D_F(a) : \Phi(\mathbb{R}^{n-m}) \subseteq T_a M \subseteq \ker(D_F(a))$. De plus : $\dim(\mathbb{R}^{n-m}) = n - m$ et $\dim(\ker(D_F(a))) = n - m$ par le théorème du rang. Donc :

$$\Phi(\mathbb{R}^{n-m}) = T_a M = \ker(D_F(a)) \quad \square$$

Nous sommes en principe prêts pour énoncer le théorème principal. Regardons d'abord une illustration d'extrema sous contrainte. On cherche le maximum de f sur M . Les lignes bleues sont les lignes de niveau $N_c = \{x : f(x) = c\}$ d'une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. Comme une carte de randonnées de montagne, où $f(x, y)$ est la hauteur du terrain au point (x, y) . Imaginons un max. local de f au centre des cercles.



Si, comme dans la situation (I), la courbe M (i.e. un chemin de randonnée indiqué sur la carte) coupe ces lignes non-tangentiellement, il n'y a pas espoir d'un extremum local : sur les deux cotés se trouvent des valeurs plus grandes et petites. En revanche, dans la situation (II) nous avons un extremum local : avant et après le point de contact a les valeurs de f sont plus petites : c'est un maximum local : pas de f sur Ω mais de f sur la courbe M . Rappelons que M est donné par les points $M = \{(x, y) : F(x, y) = 0\}$. Par une résolution locale, $\{(x, g(x)) : |x - a'| < \delta\} \subset M$, alors la tangente de M au point (a', a'') a la direction $(1, g'(a'))$. Mais $g'(a') = -F_x(a)/F_y(a)$ par dérivation implicite. La tangente a donc la direction

$$(1, -F_x(a)/F_y(a)) \perp (F_x(a), F_y(a)) \in \mathbb{R}^2$$

c'est à dire qu'elle est perpendiculaire au gradient de F en a . Mais le gradient de f est perpendiculaire à la tangente à la ligne de niveau N_c (la même tangente!), car cette dernière indique, elle, la direction de la plus forte variation de f , voir remarque 2.16. Il suit que $\nabla f(a)$ et $\nabla F(a)$ sont colinéaires en tout extremum local de f sur M . Cette structure se généralise à m contraintes simultanées de fonctions f définies sur des ouverts de \mathbb{R}^n .

Théorème. 4.2. Multiplicateurs de Lagrange Soit $\Omega \in \mathbb{R}^n$ ouvert, $F : \Omega \rightarrow \mathbb{R}^m$ différentiable, $m < n$; $M = \{x \in \Omega : F(x) = 0\}$; D_F continue en $a \in M$ et $D_F(a)$ surjective. Si $f : \Omega \rightarrow \mathbb{R}$ est différentiable et admet un extremum local en $x = a$ alors :

$$(\nabla f)(a) = \sum_{j=1}^m \lambda_j (\nabla F_j)(a) \quad (4.2)$$

C'est à dire que si on écrit F avec ces fonctions coordonnées, $F = (F_1, \dots, F_m)^t$, le gradient de f en a est une combinaison linéaire des gradients des F_j .

Démonstration. Observons que la ligne numéro j de la matrice Jacobienne de F est de la forme $(\frac{\partial F_j}{\partial x_k})_{k=1 \dots m} = \nabla F_j(a)$, donc

$$\mathcal{J}_F(a) = \begin{pmatrix} (\nabla F_1)(a) \\ (\nabla F_2)(a) \\ \dots \\ (\nabla F_m)(a) \end{pmatrix}$$

Par hypothèse, $\mathcal{J}_F(a)$ est de plein rang : $rg = m$. Les vecteurs $(\nabla F_i)(a)$ sont donc linéairement indépendants. L'espace vectoriel $V \stackrel{\text{def.}}{=} \{(\nabla F_j)(a) : j = 1, \dots, m\}$ est ainsi de dimension m .

On veut montrer : $(\nabla f)(a) = \sum_{j=1}^m \lambda_j (\nabla F_j)(a)$, pour certains λ_j , donc $(\nabla f)(a) \in V$. On montre ceci en montrant d'abord $\nabla f(a) \in (T_a M)^\perp$ et puis $V = (T_a M)^\perp$.

Par hypothèse on sait que f admet un extremum local au point $a \in M$. Soit $v \in (T_a M)$ et γ courbe dérivable dans M telle que $\gamma(0) = a$ et $\gamma'(0) = v$. Montrons que $(\nabla f)(a) \in (T_a M)^\perp$: par hypothèse, $f(\gamma(t))$ admet un extremum au point $t = 0$: $\frac{df(\gamma(t))}{dt}|_{t=0} = 0 \Leftrightarrow D_f(\gamma(0))(\gamma'(0)) = 0 \Leftrightarrow \langle (\nabla f)(a) | v \rangle = 0$ Ceci montre que le gradient de f au point a est perpendiculaire à v pour tout $v \in T_a M$. Donc $\nabla f(a) \in (T_a M)^\perp$. Attaquons l'égalité des deux sous-espaces vectoriels $(T_a M)^\perp = V = \text{vect}(\nabla F_1(a), \dots, \nabla F_m(a))$.

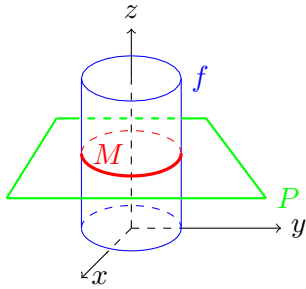
" \supseteq " : Soit $\gamma : (-\varepsilon; \varepsilon) \rightarrow M$ à l'effet que $F_j(\gamma(t)) = 0_{\mathbb{R}^m}, \forall |t| < \varepsilon$ et tout $j = 1 \dots m$. Ainsi $\frac{df(\gamma(t))}{dt}|_{t=0} = 0 \Leftrightarrow \langle (\nabla F_j)(a) | v \rangle = 0$ pour tout $v \in T_a M$. Ceci montre que $V \subseteq (T_a M)^\perp$.

" \subseteq " : $\dim(T_a M) = \dim(\ker(D_F(a))) = n - m$ donc $\dim((T_a M)^\perp) = \dim(\mathbb{R}^n) - \dim(T_a M) = n - (n - m) = m$ et $\dim V = m$. Ainsi, " \supseteq " avec égalité de dimensions montre $V = (T_a M)^\perp$.

On a établi $(\nabla f)(a) \in (T_a M)^\perp = V$. Ce qui signifie que ∇f est une combinaison linéaire d'une base de V i.e. (4.2). \square

Exemple 4.3. $f(x, y) = x^2 + y^2$. On souhaite maximiser f sur un plan $F(x, y) = x + y - 1$ Ici $M = \{(x, y) \in \mathbb{R}^2; F(x, y) = 0\}$ est la droite $y = 1 - x$ de \mathbb{R}^2 . On sait qu'en tout point extremal de f sur M : $(\nabla f)(x, y) = (2x, 2y) = \lambda(1, 1)$ pour un λ à déterminer. On a $2x = 1 = 2y$ donc $x = y$ Ainsi $(x, x) \in M \Leftrightarrow x + x - 1 = 0 \Leftrightarrow x = 1/2$ alors $f(1/2, 1/2) = 1/2$ est un extremum de f . Comme $\lim_{\|(x, y)\| \rightarrow +\infty} f = +\infty$, nécessairement $1/2$ est un minimum local. (c'est à dire que pour tous les points $(x, 1 - x)$ le minimum de f est atteint en $x = 1/2$ et vaut $1/2$)

Exemple 4.4.



Soit $f(x, y, z) = x + 2y + z$ et $F(x, y, z) = \begin{pmatrix} x^2 + y^2 - 1 \\ y + z - 1 \end{pmatrix}$. M est l'intersection d'un plan P et d'un cylindre. On a

$$(\nabla f)(x, y, z) = (1, 2, 1)$$

et

$$\mathcal{J}_F(x, y, z) = \begin{pmatrix} 2x & 2y & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

On observe $\text{rg } \mathcal{J}_F = 2$ si $x \neq 0$ ou $y \neq 0$. Or $(0, 0, z) \notin M$: partout sur M on a plein rang, donc si f admet un extremum local ce sera une combinaison linéaire des gradients de F :

$$(\nabla f) = (1, 2, 1) = \lambda(2x, 2y, 0) + \mu(0, 1, 1)$$

Ceci nous donne le système suivant :

$$\begin{cases} \mu = 1 \\ 2\lambda x = 1 \\ 2\lambda y + 1 = 2 \end{cases}$$

On remarque que $\lambda \neq 0$ car $(1, 2, 1) \neq (0, 1, 1)$ ce qui nous donne $x = y$. Avec la contrainte $x^2 + y^2 = 1$ on obtient $(x, y) \in \{(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}), (-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})\}$. La seconde contrainte $z = 1 - y$ nous permet de déterminer deux seuls points : $a = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 1 - \frac{1}{\sqrt{2}})$ et $b = (-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 1 + \frac{1}{\sqrt{2}})$ soit $f(a) = 1 + \sqrt{2}$ et $f(b) = 1 - \sqrt{2}$.

Nature de ces points : M est fermé comme intersection de deux fermés, et il est borné donc M est compact. De plus f est continue, elle atteint ses bornes ; puisque deux extrema doivent exister et on en a trouvé deux, il faut juste voir "qui est qui". Par comparaison de $f(a)$ et $f(b)$ on voit $f(a) = \max_M(f)$ et $f(b) = \min_M(f)$.

En pratique, on cherche à maximiser sur une boule, on a donc deux cas : ou bien les extrema sont dedans et on utilise le critère du gradient qui s'annule, ou bien les extrema sont sur les bords et on utilise la méthode des multiplicateurs de Lagrange.

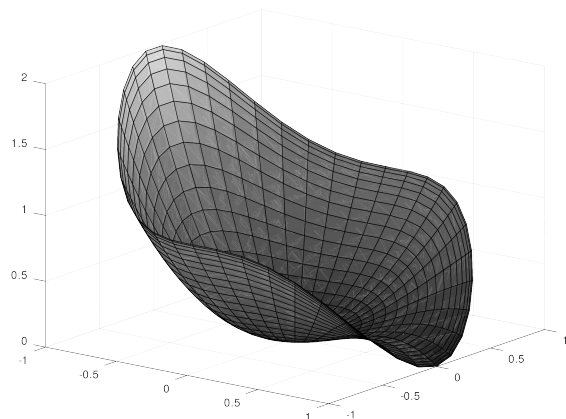
Exemple 4.5. Soit $f(x, y) = e^{-x^2 - y^2}$. On cherche le maximum sur la boule fermée $K = \{x^2 + y^2 \leq 1\}$ (ceci est un compact, et f continue, le max. existe donc). On voit que $f(x, y) = \exp(-\|(x, y)\|_2^2)$. Le maximum est donc pris en zéro (valeur 1) et la fonction est constante sur le bord (minimum de f sur K).

Deuxième exemple : soit $g(x, y) = x^2 + y^2 - x^3$. On cherche le maximum sur $K = \{x^2 + y^2 \leq 1\}$ (voir image à droite).

A l'intérieur : $\nabla g(x, y) = (2x - 3x^2, 2y)$ les points critiques sont $(0, 0)$ et $(2/3, 0)$. L'étude de la matrice Hessienne révèle que $(0, 0)$ est un min. local, et $(2/3, 0)$ un point selle.

Au bord : on pose $F(x, y) = x^2 + y^2 - 1$. Alors $\nabla F(x, y) = \lambda \nabla g(x, y)$ donne

$$(2x, 2y) = \lambda \cdot (2x - 3x^2, 2y) \quad \text{et} \quad x^2 + y^2 = 1$$



(dans le dessin le plan est parallèle à (x, y) !!)

D'abord $\lambda \neq 0$ car sinon $(x, y) = (0, 0)$ ce qui contredit la deuxième équation. D'où soit $y = 0$ et donc $x = \pm 1$ soit $\lambda = 1$, puis $x = 0$ puis $y = \pm 1$. On a donc 4 nouveaux points sur le cercle :

$$g(0, 1) = 1 = g(0, -1), \quad g(1, 0) = 0 \quad g(-1, 0) = 2$$

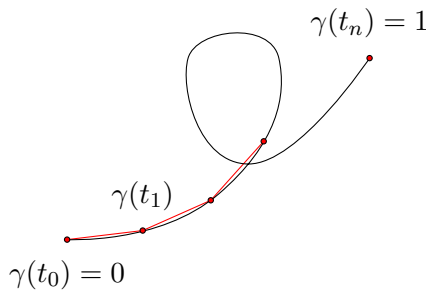
Le maximum de g sur K est donc 2, pris au point $(-1, 0)$. Les points $(0, \pm 1)$ sont des points "fantômes" trouvés par accident (des "points selle sur le bord"). On les ignore. Le min. de g sur K est 0, valeur prise en $(0, 0)$ et en $(1, 0)$.

4.2 Intégrales curvilignes

Avant de lire cette section, il est utile de re-lire la section sur es courbes paramétrées !

4.2.1 Intégrales sur une courbe

Soit $\gamma : [0, 1] \rightarrow \mathbb{R}^n$ de classe \mathcal{C}^1 . On peut approcher la courbe par des segments rectilignes et mesurer le poids appliqué en chaque point t_i , quantifié par $f(\gamma(t_i))$.



$$\begin{aligned} & \sum_{i=0}^{n-1} f(\gamma(t_i)) \cdot \|\gamma(t_{i+1}) - \gamma(t_i)\|_2 \\ &= \sum_{i=0}^{n-1} f(\gamma(t_i)) \cdot (t_{i+1} - t_i) \cdot \frac{\|\gamma(t_{i+1}) - \gamma(t_i)\|_2}{(t_{i+1} - t_i)} \\ &= \sum_{i=0}^{n-1} f(\gamma(t_i)) \cdot (t_{i+1} - t_i) \cdot \left\| \frac{\gamma(t_{i+1}) - \gamma(t_i)}{(t_{i+1} - t_i)} \right\|_2 \end{aligned}$$

Or $\left\| \frac{\gamma(t_{i+1}) - \gamma(t_i)}{(t_{i+1} - t_i)} \right\|_2 \xrightarrow{t_{i+1} \rightarrow t_i} \|\gamma'(t_i)\|_2$, avec une erreur d'approximation qui est $o(|t_{i+1} - t_i|)$, on obtient en limite l'objet suivant, qui nous sert comme définition :

Définition 4.6. Soit $\gamma : [a, b] \rightarrow \mathbb{R}^n; \in \mathcal{C}^1$, alors

$$\int f d\gamma := \int_a^b f(\gamma(t)) \|\gamma'(t)\|_2 dt$$

Remarque 4.7. Soit une reparamétrisation telle que $\tilde{\gamma} = \gamma \circ \varphi$ où $\varphi : [a, b] \rightarrow [\alpha, \beta]$ est strictement monotone, \mathcal{C}^1 (et donc φ^{-1} aussi). Alors

$$\int f d\tilde{\gamma} = \int_a^b f(\tilde{\gamma}(t)) \|\tilde{\gamma}'(t)\|_2 dt = \int_a^b f(\gamma(\varphi(t))) \|\tilde{\gamma}'(\varphi(t))\|_2 \varphi'(t) dt$$

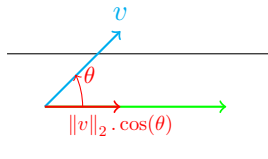
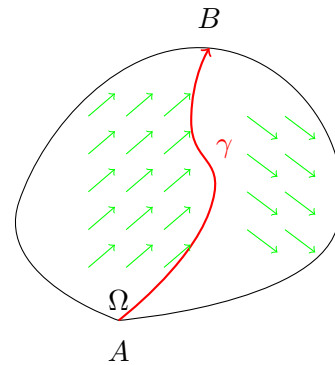
(on suppose φ strictement croissante, si elle est strictement décroissante il suffit d'inverser les bornes). On pose : $\varphi(t) = s$, alors $\varphi'(t) dt = ds$ et

$$\int f d\tilde{\gamma} = \int_{\varphi(a)=\alpha}^{\varphi(b)=\beta} f(\gamma(s)) \cdot \|\gamma'(s)\|_2 ds$$

L'intégrale ne dépend pas d'un changement de paramètre.

4.2.2 Champs vectoriels

Soit $f : \Omega \in \mathbb{R}^n \rightarrow \mathbb{R}^n$ continue. Imaginons que l'on parte du point A pour atteindre le point B , en subissant des forces sur le chemin (exemple : un bateau qui va de Calais à Douvre et qui subit les vents variables sur la Manche).



On a la relation suivante : "travail = force subie \times chemin parcouru" : imaginons un train soumis à un vent de travers : comme le train est sur des rails, la force qui s'exerce réellement sur lui est représentée en rouge sur le dessin.

D'autre part, on a pour deux vecteurs a, b une décomposition (unique) $b = \alpha a + \beta a^\perp$ où $\langle a^\perp | a \rangle = 0$. Il suit (par le même dessin, remplacer u, v par a, b !) que $\langle a | b \rangle = \alpha \|a\|^2 = \|a\| \cdot \|b\| \cdot \cos(\theta)$, ce qui donne :

$$\begin{aligned} \sum_{i=0}^{n-1} \langle f(\gamma(t_i)) | (\gamma(t_{i+1}) - \gamma(t_i)) \rangle &= \sum_{i=0}^{n-1} \left\langle f(\gamma(t_i)) \left| \frac{\gamma(t_{i+1}) - \gamma(t_i)}{t_{i+1} - t_i} \right. \right\rangle \cdot (t_{i+1} - t_i) \\ &\xrightarrow{n \rightarrow \infty} \int_a^b \langle f(\gamma(t)) | \gamma'(t) \rangle dt \end{aligned}$$

Définition 4.8. Soit $f : \Omega \rightarrow \mathbb{R}^n$, continue et $\gamma \in \mathcal{C}^1$ alors :

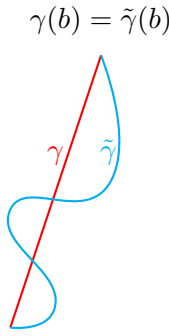
$$\int f.d\gamma = \int_a^b \langle f(\gamma(t)) | \gamma'(t) \rangle dt$$

De même que dans le premier cas, ce résultat ne varie pas en cas de changement de paramètre : $t = \varphi(s), dt = \varphi'(s)ds$ donne

$$\int f.d\gamma = \int_a^b \langle f(\gamma(t)) | \gamma'(t) \rangle dt = \int_\alpha^\beta \langle f(\gamma(\varphi(s))) | \gamma'(\varphi(s))\varphi'(s) \rangle ds$$

puis, si on appelle $\xi = \gamma \circ \varphi$ ceci est $\int_\alpha^\beta \langle f(\xi(s)) | \xi'(s) \rangle ds = \int f.d\xi$.

4.2.3 Application : Primitives de champs de vecteurs



$$\gamma(a) = \tilde{\gamma}(a)$$

Quand est-ce qu'on peut "primitiver" une fonction $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$? L'idée est de trouver une fonction $u : \Omega \rightarrow \mathbb{R}$ avec $f = \nabla u$! Imaginons que ce serait le cas, avec $u : \Omega \rightarrow \mathbb{R}$ et $u \in \mathcal{C}^1$. Alors

$$\int f \cdot d\gamma = \int_a^b \langle (\nabla u)(\gamma(t)) | \gamma'(t) \rangle dt = \int_a^b \frac{d}{dt} u(\gamma(t)) dt = u(\gamma(b)) - u(\gamma(a))$$

par le théorème fondamental d'analyse. Cela signifie que cette intégrale curviligne ne dépend pas du chemin : elle dépend uniquement des points de départ et d'arrivée!

Corollaire 4.9. ("Théorème du gradient") Soit $f = \nabla u$, $u : \Omega \rightarrow \mathbb{R}$; $u \in \mathcal{C}^1$;

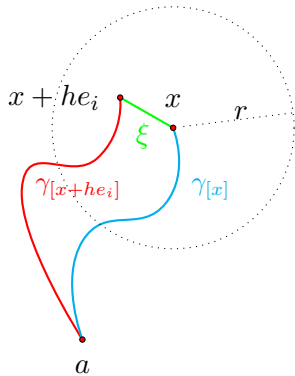
$$\int f \cdot d\gamma = u(\gamma(b)) - u(\gamma(a))$$

De façon surprenante nous avons une réciproque directe

Proposition 4.10. ("Théorème réciproque du gradient") Soit $\Omega \subset \mathbb{R}^n$ ouvert et $f : \Omega \rightarrow \mathbb{R}^n$ une fonction de continue. Supposons $\forall x, y \in \Omega$; $\exists \gamma : [0, 1] \rightarrow \Omega$ tel que $\gamma(0) = x$ et $\gamma(1) = y$. Fixons un point $a \in \Omega$ et soit $\gamma_{[x]}$ une courbe qui relie a à x . On définit $u(x) \stackrel{\text{def.}}{=} \int f \cdot d\gamma_{[x]}$. Si cette intégrale ne dépend pas du chemin emprunté on a : $\nabla u = f$ pour une fonction u de classe \mathcal{C}^1 .

Démonstration. On fixe pour chaque x une courbe $\gamma_{[x]}$. La fonction $u(x) \stackrel{\text{def.}}{=} \int f \cdot d\gamma_{[x]}$ ainsi définie est bien définie en tout point. Alors u est continue, et partiellement dérivable :

L'argument est semblable pour les deux assertions : il existe $r > 0$ avec $B(x, r) \subseteq \Omega$. Par indépendance du chemin, et pour $\|h\| < r$ on peut remplacer $\gamma_{[x+h]}$ par l'enchaînement de $\gamma_{[x]}$ avec le segment droit ξ qui relie x avec $x+h$ (paramétrisé sur $[0, 1]$ par $\xi(s) = x + sh$). Le point est que les boules sont convexes : le segment droit entre x et $x+h$ est donc bien contenu dans Ω ! Soit $\varepsilon > 0$. Par continuité de f proche de x , il existe $\delta > 0$ tel que, pour tout $\|h\| \leq \min(r, \delta)$, $\|f(x+h) - f(x)\| \leq \varepsilon$. Il suit



$$\begin{aligned} |u(x+h) - u(x)| &= \left| \int f \cdot d\gamma_{[x+he_i]} - \int f \cdot d\gamma_{[x]} \right| = \left| \int f \cdot d\xi \right| \\ &= \left| \int_0^1 \langle f(x+sh) | h \rangle ds \right| \\ (\text{Cauchy-Schwarz}) &\leq \int_0^1 \|f(x+sh)\|_2 \|h\|_2 ds \\ &\leq (\|f(x)\|_2 + \varepsilon) \|h\|_2 \leq \varepsilon \end{aligned}$$

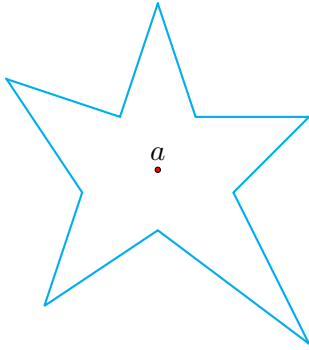
dès que $\|h\| < \min(\delta, r, \frac{\varepsilon}{\|f(x)\|_2 + 1})$. Ce qui montre que u est continue. Par le même argument de changement de chemin on a pour $h = te_i$ avec $|t| < r$ et la paramétrisation $\xi(s) = x + se_i$, avec

$s \in [0, t]$:

$$\begin{aligned} \frac{\partial u}{\partial x_j}(x) &= \lim_{t \rightarrow 0} \frac{u(x + t.e_j) - u(x)}{t} = \lim_{t \rightarrow 0} \frac{1}{t} \left(\int f.d\xi \right) \\ &= \lim_{t \rightarrow 0} \frac{1}{t} \int_0^t \langle f(x + se_j) | e_j \rangle ds = f_j(x) \end{aligned}$$

pour tout $i \in \{1, \dots, n\}$. Il suit que $\nabla u = f$. Puisque les dérivées partielles de u (les composantes de f donc) sont continues, u est de classe \mathcal{C}^1 . \square

Observons en particulier que si $\gamma(a) = \gamma(b)$, alors l'intégrale curviligne est nulle. Cette réciproque développe la plus grande importance dans l'analyse complexe (fonctions de $\Omega \subset \mathbb{C}$ vers \mathbb{C} où l'on démontre indépendance de l'intégrale curviligne si Ω est "simplement connexe" ce qui veut dire (en prose) qu'on peut rétracter toute courbe paramétrée dans Ω par des courbes continues intermédiaires en un point. En gros, on interdit des "trous" dans Ω . Mais ce sujet devra hélas attendre bien longtemps ... on y reviendra en master 1 ! Une classe de domaines simplement connexes est celle des domaines "étoilés" :

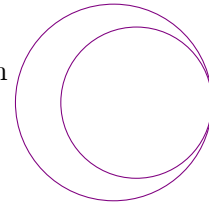


Définition 4.11. (Ensemble étoilé)

Un ensemble Ω est dit étoilé s'il existe un point $a \in \Omega$ tel que :

$$\forall x \in \Omega; \{a + t(x - a) : t \in [0, 1]\} \subseteq \Omega.$$

C'est à dire : tout segment droit reliant a à un point de Ω est inclus dans Ω .



"luné", mais pas étoilé

Remarque 4.12. Si $f : \Omega \in \mathbb{R}^n$ est telle que $f = \nabla u$ et si $u \in \mathcal{C}^2$ alors $\text{Hess}(u) = \mathcal{J}_f$ est symétrique.

Cette condition caractérise les gradients de fonctions de classe \mathcal{C}^2 dans des domaines étoilés.

Théorème. 4.13. ("Lemme de Poincaré*") Soit Ω ouvert, étoilé et $f : \Omega \rightarrow \mathbb{R}^n$ de classe $f \in \mathcal{C}^1$. Supposons que sa matrice Jacobienne \mathcal{J}_f soit symétrique en tout point, c'est à dire

$$\forall x \in \Omega : \quad \frac{\partial f_k}{\partial x_j}(x) = \frac{\partial f_j}{\partial x_k}(x).$$

Alors f est un gradient d'une fonction $u \in \mathcal{C}^2(\Omega)$.

Démonstration. Quitte à translater, on peut prendre $a = 0$. Soit $\gamma_{[x]} = tx$, $t \in [0, 1]$ et $g : [0, 1] \rightarrow \mathbb{R}^n$ telle que $g(t) = t.f(tx)$. Alors $g'(t) = f(tx) + t.\mathcal{J}_f(tx).x$. Rappelons les projections coordonnées π_k . Alors $g_k = \pi_k \circ g$ et $f_k = \pi_k \circ f$ sont les fonctions coordonnées de f, g respectivement. Avec cette notation, en explicitant le produit matrice-vecteur,

$$g'_k(t) = f_k(tx) + t \sum_{j=1}^n \frac{\partial f_j}{\partial x_k}.x_j$$

*. Henri Poincaré est un mathématicien, physicien théoricien et philosophe des sciences français, né le 29 avril 1854 à Nancy et mort le 17 juillet 1912 à Paris.

Définissons maintenant

$$u(x) \stackrel{\text{def.}}{=} \int f \cdot d\gamma_{[x]} = \int_0^1 \langle f(tx) | x \rangle dt$$

Pour dériver u il faut interchanger dérivée et intégrale. Attention, c'est une opération d'échange de deux limites, cela ne se fait pas sans précautions particulières †. Deux ingrédients clé ; les détails restent un exercice avec indications pour mes lecteurs (ce n'est pas difficile) :

- i) si f_n converge uniformément vers f , i.e. si $\sup\{|f_n(x) - f(x)| : a \leq x \leq b\}$ tend vers zéro, alors $\lim \int f_n = \int \lim f_n = \int f$ (intégrales sur $[a, b]$!).
- ii) si f est de classe \mathcal{C}^1 , alors f' est continue, donc uniformément continue sur $[a, b]$. Or, $\frac{1}{h}(f(x+h) - f(x)) - f'(x) = \frac{1}{h} \int_0^h f'(x+s) - f'(x) ds$ et le fait que $f'(x+s) - f'(x)$ devient uniformément petit, nous savons que les quotients différentiels f_n obtenu (disons) par le choix $h = 1/n$ convergent uniformément vers f' .

Combinons les deux : On considère $\varphi(t, x) = \langle f(tx) | x \rangle$ comme fonction de deux variables (t, x_k) , car les autres variables $x_j, j \neq k$ sont "gelés". Puisque la dérivée partielle par rapport à x_k est continue (comme fonction des deux variables (t, x_k)) et puisqu'on intègre sur $t \in [0, 1]$, nous avons bien $\frac{d}{dx_k} \int \varphi(x_k, t) dt = \int \frac{\partial \varphi}{\partial x_k}(t, x_k) dt$. Appliquons en plus exemple 2.10 (e) :

$$\frac{\partial u}{\partial x_k}(x) = \int_0^1 \frac{\partial}{\partial x_k} \langle f(tx) | x \rangle dt = \int_0^1 \left\langle \frac{\partial}{\partial x_k} f(tx) | x \right\rangle + \langle f(tx) | \frac{\partial x}{\partial x_k} \rangle dt$$

Or $\frac{\partial x}{\partial x_k} = e_k$ on a $\langle f(tx) | \frac{\partial x}{\partial x_k} \rangle = \langle f(tx) | e_k \rangle = f_k(tx)$ d'un coté et par dérivée de compositions

$$\left\langle \frac{\partial}{\partial x_k} (f(tx)), x \right\rangle = \langle \mathcal{J}_f(tx) \cdot t e_k, x \rangle = t \sum_{j=1}^n \frac{\partial f_k}{\partial x_j}(tx) \cdot x_j \stackrel{\text{sym. } \mathcal{J}_f}{=} t \sum_{j=1}^n \frac{\partial f_j}{\partial x_k}(tx) \cdot x_j$$

On a donc

$$\frac{\partial u}{\partial x_k}(x) = \int_0^1 (f_k(tx) + t \sum_{j=1}^n \frac{\partial f_j}{\partial x_k} \cdot x_j) dt = \int_0^1 g'_k(t) dt = g_k(1) - g_k(0) = f_k(x).$$

Il suit que $f = \nabla u$. Vu que $\nabla u = f$ est de classe \mathcal{C}^1 , u est de classe \mathcal{C}^2 , comme énoncé. Observons en bonus que le corollaire 4.9 nous assure (après-coup !) que l'intégrale curviligne

$$\int f d\gamma = f(\gamma(b)) - f(\gamma(a))$$

ne dépend pas du chemin γ emprunté, mais uniquement des points de départ et d'arrivée. Le choix de $\gamma_{[x]}$ n'était qu'une ruse de preuve ... \square

Le "lemme de Poincaré" n'est pas énoncé dans toute sa généralité. Au lieu de domaines étoilés on peut le montrer par exemple. sur des domaines "contractibles", i.e. pour lesquelles il existe une fonction continue $\Phi : [0, 1] \times \Omega \rightarrow \Omega$ et un point $a \in \Omega$ qui satisfait

$$\forall x \in \Omega : \quad \Phi(0, x) = x \quad \text{et} \quad \forall x \in \Omega : \quad \Phi(1, x) = a.$$

en prose, l'identité peut être continûment déformée en application constante. Pour un domaine étoilé avec centre a , on a $\Phi(t, a+h) = a + (1-t)h$ qui contracte le domaine en un point. Nous avons visé sur cette situation relativement simple avec un preuve "à la main" (et beaucoup plus simple).

†. on se rappelle que la double-limite de la suite numérique $a_{n,k} = \frac{n-k}{n+k}$ dépend de l'ordre de prise de limite : d'abord $n \rightarrow \infty$ ou d'abord $k \rightarrow \infty$?

Exemple 4.14. Soit $f(x, y) = (e^x + \sin(y), x \cos(y) + y^2)$. Clairement, $f \in \mathcal{C}^1$. Est-ce un gradient ? On teste les dérivées partielles “croisées” : $\frac{\partial f_1}{\partial y} = \cos(y) = \frac{\partial f_2}{\partial x}$. C’est bon ! En effet, $u(x, y) = e^x + x \sin(y) + y^3/3$ satisfait $\nabla u = f$.

Exercice 4.15. Vérifier si les fonctions suivantes sont des gradients. Si oui, calculer u_i avec $\nabla u_i = f_i$.

- a) $f_1(x, y) = (y^2/x + 2xy^3 - \frac{1}{1+x^2}, 2y \ln(x) + 3x^2y^2 - \sin y)$
- b) $f_2(x, y) = (4x^3y - \frac{1}{1+x^2} + e^y, x^4 + xe^y + x)$
- c) $f_3(x, y) = (y^3 + 2xy + 3x^2 + 2xy^2, 4y^3 + x^2 + 2x^2y + 3xy^2)$.
- d) $f_4(x, y) = (x^2 \arcsin y, \frac{x^3}{3\sqrt{1-y^2}} - \ln(y))$.
- e) $f_5(x, y) = (\ln(x) + 2xye^y + \frac{x}{\sqrt{1-x^2}}, x^2(1+y)e^y + \frac{1}{\sqrt{1-y^2}})$.

4.3 Intégrales de surface

4.3.1 La vérité sur les déterminants

Le titre est un peu provocateur. Mais je souhaite vraiment attirer l’attention sur le fait que les déterminants ne sont pas juste un outil en algèbre linéaire. C’est une notion clé à l’interface d’algèbre, d’analyse et de la géométrie, et de la théorie de la mesure.

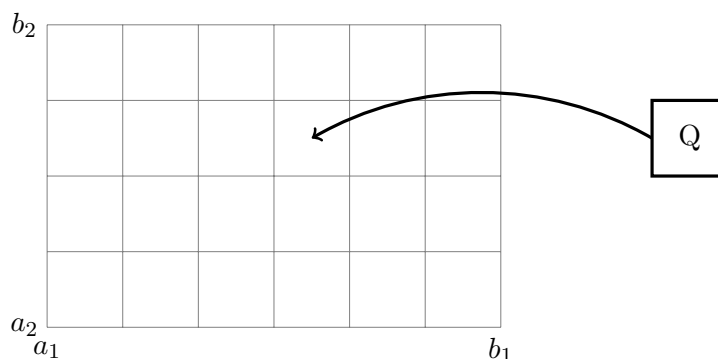
On va essayer de mesurer des volumes d’objets dans \mathbb{R}^n . Les idées communément acceptées sont que ni une translation ni une rotation devrait changer le volume d’un objet. Puis on se met d’accord sur le fait que le volume d’une union disjointe de deux objets est la somme de leurs volumes. On appelle cela “additivité”. Finalement, on peut fixer le volume du cube d’unité $Q = [0, 1]^n$ comme on veut, mais la valeur $\text{Vol.}(Q) = 1$ semble opportune. Ces idées ont plus de conséquences que l’on pense ! D’abord il on va montrer qu’un “pavé”

$$P = \prod_{i=1 \dots n} [a_i, b_i)$$

avec $a_i, b_i \in \mathbb{Z}$ a le volume

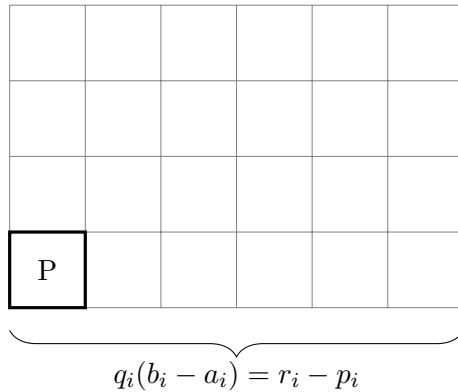
$$\text{Vol.}(P) = \prod_i (b_i - a_i) \tag{4.3}$$

il suffit en effet de le découper en translations de cubes d’unité et d’utiliser l’additivité !



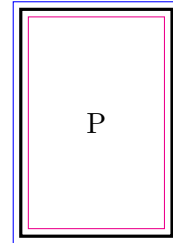
Combien de copies translattées de Q voyez-vous ? Je compte $(b_2 - a_2)(b_1 - a_1)$! En dimension n on a $\prod_i (b_i - a_i)$ copies translattées de Q . Ce qui prouve (4.3) pour $a_i, b_i \in \mathbb{Z}$.

Ensuite, on autorise $a_i, b_i \in \mathbb{Q}$: en les écrivant comme $a_i = \frac{p_i}{q_i}$ et $b_i = \frac{r_i}{q_i}$ (on peut choisir le même dénominateur) on peut créer un nouveau “super-pavé” en accolant des copies translattées de P où un met q_i copies en direction x_i .



Le résultat est une union disjointe de longueur $r_i - p_i \in \mathbb{N}^*$ en direction x_i , donc de volume $\prod_i (r_i - p_i)$. Mais il s’agit de $\prod_i q_i$ copies disjointes de P : en prenant le quotient, on tombe sur (4.3), comme avant, lorsque $a_i, b_i \in \mathbb{Q}$.

Observons que l’additivité entraîne une monotonie du volume : si $A \subset B$, alors $\text{Vol.}(A) \leq \text{Vol.}(B)$. En en c’est à dire rant un pavé de cotés $a_i, b_i \in \mathbb{R}$ par des pavés *rationnels* depuis l’intérieur et d’autres par l’extérieur, on étend (4.3) à toutes sortes de pavés : Si ℓ_i sont les longueurs réelles de P , et $r_{i,n} < \ell_i < R_{i,n}$ avec $\lim_{n \rightarrow \infty} r_{i,n} = \lim_{n \rightarrow \infty} R_{i,n} = \ell_i$, alors $\prod_i r_{i,n} \rightarrow \prod_i \ell_i$ et $\prod_i R_{i,n} \rightarrow \prod_i \ell_i$ quand $n \rightarrow \infty$. Nous avons démontré



Lemme 4.16. Soit $P = \prod_{i=1..n} [a_i, b_i]$ avec $a_i < b_i$. Alors $\text{Vol.}(P) = \prod_{i=1..n} (b_i - a_i)$.

Observons que l’on peut ajouter ou enlever des faces “sous-dimensionnelles” de pavés : Puisque pour tout $\varepsilon > 0$,

$$\{a\} \times [a_2, b_2] \times \dots \times [a_n, b_n] \subset [a - \varepsilon, a + \varepsilon] \times [a_2, b_2] \times \dots \times [a_n, b_n]$$

et le pavé de gauche a un volume qui tend vers zéro avec $\varepsilon \rightarrow 0$, il suit que

$$\text{Vol.}(\{a\} \times [a_2, b_2] \times \dots \times [a_n, b_n]) = 0$$

L’idée est qu’un objet qui rentre dans une boite arbitrairement plate ne peut avoir de volume. Par additivité, il suit

$$\text{Vol.}(\prod_{i=1..n} [a_i, b_i]) = \text{Vol.}(\prod_{i=1..n} [a_i, b_i]) = \text{Vol.}(\prod_{i=1..n} (a_i, b_i))$$

Les prochains objets qui nous intéressent sont des parallélépipèdes. Pour $v_1, \dots, v_n \in \mathbb{R}^n$, on pose $P(v_1, \dots, v_n) = \{\sum_i t_i v_i : 0 \leq t_i \leq 1\}$, le parallélépipède engendré par ces vecteurs. Il est utile de se rendre compte qu’un parallélépipède est précisément l’image du cube d’unité Q sous une application linéaire L . Ainsi, P est non-dégénéré (“sous-dimensionnel”) si et seulement si L est bijectif. On étend donc les volumes de pavés à leurs “déformations linéaires”. Ceci est intéressant

en soi, mais aussi en vue que l'approximation d'ordre 1 d'une fonction \mathcal{C}^1 est, à translation près, une application linéaire — la différentielle ! Définissons donc :

$$D(v_1, \dots, v_n) = \text{Vol.}(P(v_1, \dots, v_n))$$

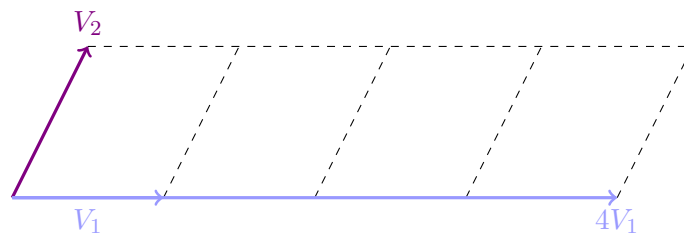
Quelles propriétés a D ? Observons $P(e_1, \dots, e_n) = Q$, le cube d'unité. Il suit que

$$D(e_1, \dots, e_n) = 1$$

et plus généralement, $D(\lambda_1 e_1, \dots, \lambda_n e_n) = \prod_i \lambda_i$. Ensuite, observons que si v_1, \dots, v_n sont liés, quitte à faire une rotation, on tombe dans la situation des faces sous-dimensionnelles, et donc $D(v_1, \dots, v_n) = 0$. De l'autre côté, si v_1, \dots, v_n sont libres, on trouve certainement un petit cube translaté dans $P(v_1, \dots, v_n)$. Ce cube a un volume positif, ce qui montre par monotonie que $D(v_1, \dots, v_n) \neq 0$. On peut donc résumer :

$$D(v_1, \dots, v_n) = 0 \text{ si et seulement si } v_1, \dots, v_n \text{ sont liés}$$

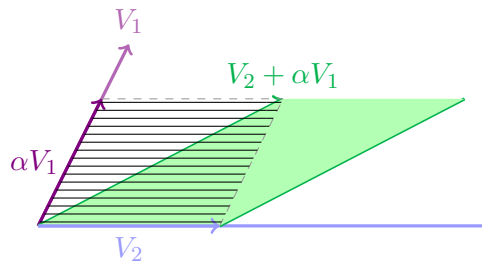
Ensuite, il est curieux de regarder $D(\alpha v_1, v_2, \dots, v_n)$. Comme pour les pavés, on commence avec $\alpha \in \mathbb{N}$, pour ensuite obtenir $\alpha \in \mathbb{Q}_+$, puis, par limite $\alpha \in \mathbb{R}_+$. Le dessin illustre la situation $\alpha = 4$. On a 4 copies translatés du parallélépipède :



On voit que

$$D(\alpha v_1, v_2, \dots, v_n) = \alpha D(v_1, v_2, \dots, v_n)$$

et évidemment la même chose marche avec les autres directions v_k . Ce qui nous gêne ici, est que si $\alpha < 0$, la notion habituelle de volume, d'être positif, suggère de mettre $|\alpha|$ à droite. Mais ceci détruit des propriétés algébriques. Et si on enlevait la valeur absolue pour $\alpha < 0$? ! Certes, cela donnerait des "volumes" négatifs, mais on peut s'en occuper plus tard avec une seule valeur absolue. En revanche, on gagne une propriété qui n'a pas de prix : linéarité ! En effet, observons d'abord que $D(v_1, v_2, v_3, \dots, v_n) = D(v_1, \alpha v_1 + v_2, v_3, \dots, v_n)$:



Les deux parallélogrammes ont la même base, et la même hauteur. Donc le même volume / air. On pensera à une pile de livres bien droite d'un coté, et la même pile de livres un peu translattés, de travers (sans renverser la pile). Le volume est inchangé. La même chose se passe en dimension n et entre v_i et v_k : on peut ajouter ou enlever $\xi_i v_i$ à v_k sans changer les volumes. Cela est utilisé plusieurs fois dans le calcul suivant : Soit $x = \sum \xi_i v_i$.

$$\begin{aligned}
 D(v_1, x + v_2, v_3, \dots, v_n) &= D(v_1, \xi_1 v_1 + (\xi_2 + 1)v_2 + \sum_{i>2} \xi_i v_i, v_3, \dots, v_n) \\
 &= D(v_1, (\xi_2 + 1)v_2, v_3, \dots, v_n) \\
 &= (\xi_2 + 1)D(v_1, v_2, v_3, \dots, v_n) \\
 &= D(v_1, v_2, v_3, \dots, v_n) + D(v_1, \xi_2 v_2, v_3, \dots, v_n) \\
 &= D(v_1, v_2, v_3, \dots, v_n) + D(v_1, \xi_1 v_1 + \xi_2 v_2 + \sum_{i>2} \xi_i v_i, v_3, \dots, v_n) \\
 &= D(v_1, v_2, v_3, \dots, v_n) + D(v_1, x, v_3, \dots, v_n)
 \end{aligned}$$

Le même calcul montre que D , quitte à autoriser des volumes “négatifs”, est linéaire dans chacune de ses n arguments! On parle de “multi-linéaire”.

$$\boxed{(v_1, \dots, v_n) \mapsto D(v_1, \dots, v_n) \text{ est multilinéaire}}$$

Finalement,

$$\begin{aligned}
 D(v_1, v_2, v_3, \dots, v_n) &= D(v_1 - v_2, v_2, v_3, \dots, v_n) \\
 &= D(v_1 - v_2, v_2 + (v_1 - v_2), v_3, \dots, v_n) \\
 &= D(v_1 - v_2, v_1, v_3, \dots, v_n) \\
 &= D(-v_2, v_1, v_3, \dots, v_n) \\
 &= -D(v_2, v_1, v_3, \dots, v_n)
 \end{aligned}$$

ce qui montre “l’alternance” : pour chaque permutation de deux arguments, on a un facteur (-1) qui se forme.

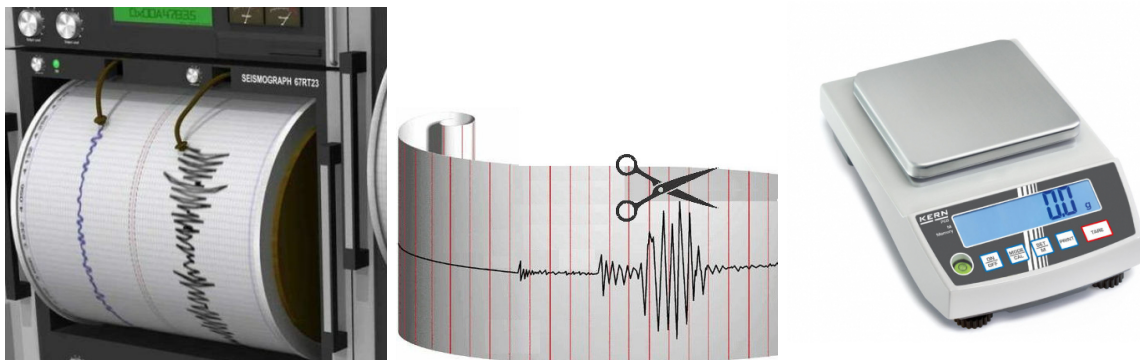
$$\boxed{(v_1, \dots, v_n) \mapsto D(v_1, \dots, v_n) \text{ est alternée}}$$

On l’aura compris : avec la renormalistaion $D(e_1, \dots, e_n) = 1$, D est unique, et on appelle $D(v_1, v_2, v_3, \dots, v_n) = \det(v_1, v_2, v_3, \dots, v_n)$ le déterminant. Le volume du parallélépipède est

$$\text{Vol.}(P(v_1, \dots, v_n)) = |\det(v_1, \dots, v_n)|.$$

Quel petit prix à payer, cette valeur absolue!

Remarque 4.17. “Intégration chimiste”. Les chimistes ont de bonnes balances, ce qui donner une idée rigolotte :



Imaginons un papier qui se déroule à vitesse constante alors que l'on trace une courbe dessus. Comment déterminer la surface entre le bord du papier et la courbe ? Il suffit de découper le morceau de papier de le peser, et de diviser le résultat obtenu par la masse du papier au m^2 . On obtient une approximation de l'aire voulue.

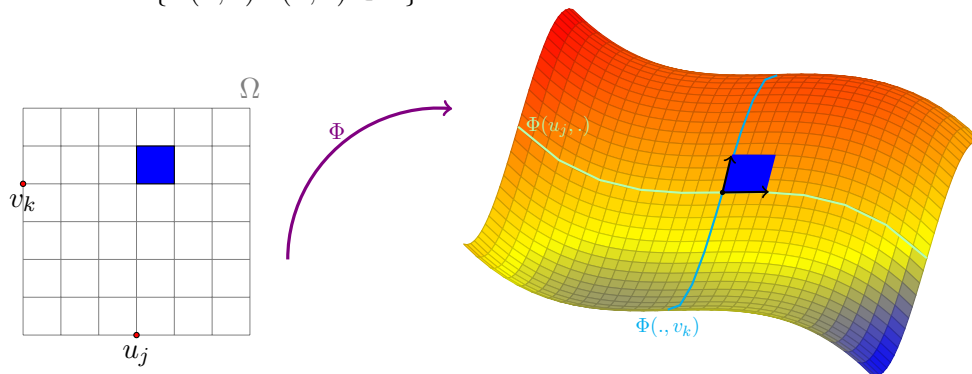
Ceci semble une belle blague, mais c'est une façon efficace pour mesurer des aires de surface dans \mathbb{R}^3 : on leur ajoute une "troisième dimension" : elle devient alors des volumes, qui se mesurent (à peu près) avec des déterminants (l'équivalence de volume et masse est superflu maintenant ... avantage des maths par rapport au monde réel). Au travail !

4.3.2 Intégrale de fonctions scalaires sur une surface

Le but est de calculer l'aire d'une surface. On intègre une coordonnée après l'autre, en travaillant sur des rectangles en $x \times y$. Les intégrales sont donc "itérés". Une double intégrale est interprétée comme

$$\int_a^b \int_c^d f(x, y) dy dx \stackrel{\text{def.}}{=} \int_a^b g(x) dx \quad \text{où} \quad g(x) = \int_c^d f(x, y) dy$$

Plus généralement, on considère $\Omega = [a, b] \times [c, d]$ et $\Phi : \Omega \rightarrow \mathbb{R}^3$, de classe \mathcal{C}^1 et on s'intéresse à la surface $\{\Phi(u, v) : (u, v) \in \Omega\}$.



Dans \mathbb{R}^3 , supposons $\Phi_u(u, v)$ et $\Phi_v(u, v)$ linéairement indépendantes. Alors $\Phi_u(u_j, v_k)$ et $\Phi_v(u_j, v_k)$ forment un parallélogramme. L'aire de ce parallélogramme est une bonne approximation de ce morceau de surface, si Φ est assez régulière et le maillage assez fin.

Première question : comment calculer l'aire de ce parallélogramme ? Puisque l'on est dans \mathbb{R}^3 on se ramène au calcul d'un volume (voir la remarque de "l'intégrale chimiste"). On prend une hauteur de 1, ce qui fait que l'aire cherchée sera égale au volume calculé.

Deuxième question : comment calculer ce volume ? Pour $a, b, c \in \mathbb{R}^3$ (où même $a_1, \dots, a_n \in \mathbb{R}^n$), soit $P(a, b, c)$ le parallélépipède engendré par a, b, c . Alors, $\text{Vol.}(P) = |\det(a \dot{ : } b \dot{ : } c)|$, voir la section précédente ! On introduit le produit vectoriel (aussi appelé produit extérieur, ou cross product, noté \times ou \wedge) :

Définition 4.18. Soient a, b deux vecteurs de \mathbb{R}^3 . Alors

$$a \wedge b \stackrel{\text{def.}}{=} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \wedge \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_2 b_3 - b_2 a_3 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix}$$

Ceci provient de la solution abstraite qui consiste à trouver n tel que $\langle a | n \rangle = 0$ et $\langle b | n \rangle = 0$. Si on pose $c \stackrel{\text{def.}}{=} a \wedge b$, on a

$$\begin{aligned} c_1 &= + \det \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix} = \det \begin{vmatrix} a_1 & b_1 & 1 \\ a_2 & b_2 & 0 \\ a_3 & b_3 & 0 \end{vmatrix} \\ c_2 &= - \det \begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix} = \det \begin{vmatrix} a_1 & b_1 & 0 \\ a_2 & b_2 & 1 \\ a_3 & b_3 & 0 \end{vmatrix} \\ c_3 &= + \det \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = \det \begin{vmatrix} a_1 & b_1 & 0 \\ a_2 & b_2 & 0 \\ a_3 & b_3 & 1 \end{vmatrix} \end{aligned}$$

En particulier, le produit scalaire (Euclidien) de $a \wedge b$ avec un vecteur $d = (d_1, d_2, d_3)$ est la somme des $c_i d_i$, donc par linéarité du déterminant juste $\det(a \dot{ : } b \dot{ : } d) = \text{Vol.}(P(a, b, d))$!

Revenons au problème : on veut calculer l'aire du parallélogramme engendré par (Φ_u, Φ_v) qui est égale au volume du parallélépipède engendré par (Φ_u, Φ_v, n) , où $n = \lambda(\Phi_u \wedge \Phi_v)$ est un vecteur "normal" : donc orthogonal aux deux autres, et normé de longueur 1. Ce qui impose $\lambda = \pm \frac{1}{\|\Phi_u \wedge \Phi_v\|}$. Nous sommes prêts à assembler les pièces :

$$\begin{aligned} \text{Aire du parallélogramme}(\Phi_u, \Phi_v) &= \text{Vol.}(P(\Phi_u, \Phi_v, n)) \\ &= |\det(\Phi_u, \Phi_v, n)| \\ &= |\lambda \cdot \det(\Phi_u, \Phi_v, \Phi_u \wedge \Phi_v)| \\ &= |\lambda| \cdot \langle \Phi_u \wedge \Phi_v, \Phi_u \wedge \Phi_v \rangle \\ &= \|\Phi_u \wedge \Phi_v\|_2. \end{aligned}$$

Définition 4.19. On définit une fonction f représentant un "poids" posé sur chaque petite aire (Φ_u, Φ_v) de la surface. Alors, en sommant toutes les "cases" affectées par le poids $f(\Phi(u_j, v_k))$ on obtient

$$\sum_{j=1}^J \sum_{k=1}^K f(\Phi(u_j, v_k)) (u_{j+1} - u_j) (v_{k+1} - v_k) \|\Phi_u \wedge \Phi_v\|_2$$

et lorsque $J, K \rightarrow \infty$ cela donne, par sommes de Riemann,

$$\int_a^b \int_c^d f(\Phi(u, v)) \cdot \|\Phi_u \wedge \Phi_v\|_2 \cdot dv \cdot du := \int f \cdot dS$$

où Φ est une paramétrisation de la surface.

Exemple 4.20. Soit

$$\Phi : \begin{cases} [0; \pi] \times [0; 2\pi] & \rightarrow \mathbb{R}^3 \\ (u, v) & \mapsto \begin{pmatrix} r \sin(u) \cos(v) \\ r \sin(u) \sin(v) \\ r \cos(u) \end{pmatrix} \end{cases}$$

avec $r > 0$. Ce sont les coordonnées sphériques en 3 dimensions. On calcule

$$\Phi_u = \begin{pmatrix} r \cos(u) \cos(v) \\ r \cos(u) \sin(v) \\ -r \sin(u) \end{pmatrix} \quad \text{et} \quad \Phi_v = \begin{pmatrix} -r \sin(u) \sin(v) \\ r \sin(u) \cos(v) \\ 0 \end{pmatrix}$$

d'où

$$\Phi_u \wedge \Phi_v = \begin{pmatrix} r^2 \sin^2(u) \cos(v) \\ r^2 \sin^2(u) \sin(v) \\ r^2 \sin(u) \cos(u) \end{pmatrix}$$

Ce qui donne :

$$\int_0^\pi \int_0^{2\pi} 1 \cdot \|\Phi_u \wedge \Phi_v\|_2 \, dv \, du = \int_0^\pi \int_0^{2\pi} r^2 \sin(u) \, dv \, du = 2\pi r^2 \int_0^\pi \sin(u) \, du = 4\pi r^2$$

... l'aire d'une sphère !

Exemple 4.21. Un cas particulier est que $\Phi(u, v) = (u, v, \varphi(u, v))$. Dans ce cas, on intègre sur une surface qui est une partie du graphe de φ . Alors $\Phi_u = (1, 0, \varphi_u)^t$ et $\Phi_v = (0, 1, \varphi_v)^t$, puis $\Phi_u \wedge \Phi_v = (-\varphi_u, -\varphi_v, 1)$ donc

$$\|\Phi_u \wedge \Phi_v\|_2 = \sqrt{1 + \varphi_u^2 + \varphi_v^2}$$

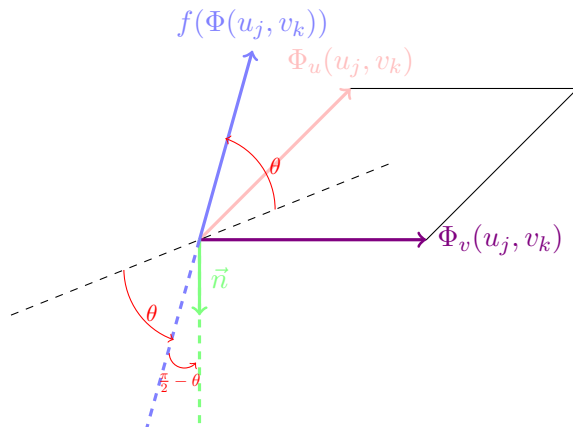
puis

$$\int_S f \, dS = \int_a^b \int_c^d f(u, v, \varphi(u, v)) \sqrt{1 + \varphi_u^2(u, v) + \varphi_v^2(u, v)} \, dv \, du.$$

4.3.3 Intégration d'un champ de vecteurs sur une surface

L'idée : on imagine un filet de pêche, ce qui nous intéresse est de connaître le flux qui passe au travers du filet.

Soit la surface S définie par Φ et le flux par une fonction f . La quantité qui traverse S dépend de l'angle de $f(\Phi(u_j, v_k))$ avec le parallélogramme $P(\Phi_u, \Phi_v)$. Pour obtenir le flux à travers l'élément de surface, pensons en effet à un courant d'eau à vitesse constante (mesuré en m/s) à travers le filet, et cela pendant une seconde. Le flux est donc égal au volume du parallélépipède engendré par $(f(\Phi(u, v))$ et P !



Il faut donc, pour chaque élément de surface $\Phi = \Phi(u_j, v_k)$, mesurer

$$\text{Vol.}(P(f(\Phi), \Phi_u, \Phi_v)) = \langle (f \circ \Phi) | \Phi_u \wedge \Phi_v \rangle.$$

On multiplie ces quantités avec $(u_{j+1} - u_j)(v_{k+1} - v_k)$ pour tenir compte de la taille du maillage considéré, puis somme sur j et k : en limite $J, K \rightarrow \infty$ on obtient finalement

$$\int_S f \cdot dS \stackrel{\text{def.}}{=} \int_a^b \int_c^d \langle f \circ \Phi | \Phi_u \wedge \Phi_v \rangle dv \cdot du = \int_a^b \int_c^d \det(f \circ \Phi, \Phi_u, \Phi_v) dv \cdot du \quad (4.4)$$

Ceci est retenu comme **définition** de l'intégrale de surface. Revenons sur le cas particulier $\Phi(u, v) = (u, v, \varphi(u, v))$. On se rappelle du dernier exemple que $\Phi_u \wedge \Phi_v = (-\varphi_u, -\varphi_v, 1)$. Ainsi,

$$\int_S f \cdot dS = \int_a^b \int_c^d (f_3(u, v, \varphi(u, v)) - f_1(u, v, \varphi(u, v))\varphi_u(u, v) - f_2(u, v, \varphi(u, v))\varphi_v(u, v)) dv \cdot du$$

Exemple 4.22. Soit $\varphi(u, v) = 1$ sur $[-1, 1] \times [-1, 1]$. La surface est donc un carré dans l'espace \mathbb{R}^3 , parallèle aux axes (x, y) en "hauteur" $z = 1$. Soit $f(x, y, z) = (0, 0, 1)^t$, donc un flux à vitesse constante 1 "vers le haut". On s'attend à ce qu'une seconde, le flux soit $2 \times 2 \times 1$, et en effet,

$$\int_S f \cdot dS = \int_{-1}^1 \int_{-1}^1 1 \, dv \, du = 4$$

Regardons $f(x, y, z) = (1, 0, 0)^t$ ou $f(x, y, z) = (0, 1, 0)^t$. Le flot est "horizontal", on s'attend à ce que rien ne traverse la surface. Et en effet, puisque $\varphi_u = \varphi_v = 0$, la formule le montre.

4.4 Généralisation à la dimension n

Dans l'inégrale de surface d'une fonction scalaire nous avons construit l'air du parallélépipède engendré par les dérivées partielles, en rajoutant un vecteur normal $\nu(\cdot)$ et de mesurer le volume. Nous avons vu que cela se fait en dimension n à l'aide d'un déterminant.

Pour une fonction vectorielle, la deuxième formulation dans (4.4) avec le déterminant permet également une formulation indépendante de la dimension. On généralise donc ainsi :

Définition 4.23. Soit $Q \stackrel{\text{def.}}{=} [a_1, b_1] \times \dots \times [a_{n-1}, b_{n-1}]$ et $\Phi : Q \rightarrow \mathbb{R}^n$ de classe \mathcal{C}^1 et $S = \Phi(Q)$ une "hypersurface". Soit $u : S \rightarrow \mathbb{R}$ une fonction scalaire. On pose

$$\int_S u \, dS \stackrel{\text{def.}}{=} \int_{a_1}^{b_1} \dots \int_{a_{n-1}}^{b_{n-1}} (u \circ \Phi)(x) | \det(\nu(\Phi(x)), \Phi_{x_1}, \dots, \Phi_{x_{n-1}}) | dx_1 \dots dx_{n-1} \quad (4.5)$$

où on écrit $x = (x_1, \dots, x_{n-1})$. Soit maintenant $F : S \rightarrow \mathbb{R}^n$ une fonction vectorielle. On pose

$$\int_S F \, dS \stackrel{\text{def.}}{=} \int_{a_1}^{b_1} \dots \int_{a_{n-1}}^{b_{n-1}} \det((F \circ \Phi)(x), \Phi_{x_1}, \dots, \Phi_{x_{n-1}}) dx_1 \dots dx_{n-1}. \quad (4.6)$$

4.4.1 Intégration par parties sur un domaine "pavé"

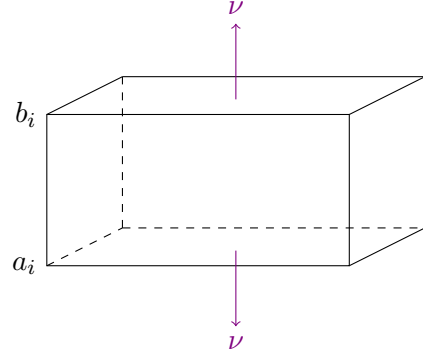
On souhaite faire une intégration par parties en plusieurs dimensions. En une dimension, la formule d'intégration par parties repose sur le fait que $\int_a^b h'(t) dt = h(b) - h(a)$ qu'on applique ensuite à $h = u.v$. On va pour suivre la même démarche. Il faut donc d'abord une extension à plusieurs variables de l'identité ci-dessus. L'évaluation de h en a et b sur la droite devient alors une intégrale de surface.

Théorème. 4.24. Soit $\Omega = [a_1, b_1] \times \dots \times [a_n, b_n]$ et $u : \Omega \rightarrow \mathbb{R}, \in C^1$. Alors

$$\int_{\Omega} \frac{\partial u}{\partial x_i} dx = \int_{\partial\Omega} u(\cdot) \nu_i(\cdot) dS \quad (4.7)$$

où $\partial\Omega$ est la frontière de Ω et ν_i est la i -ième coordonnée du vecteur normal extérieur sur $\partial\Omega$.

La difficulté de la preuve réside uniquement à décortiquer la notation et de se rendre compte que finalement on ne fait rien de compliqué. Observons déjà que l'intégrale à gauche n'est rien qu'une intégrale "itérée" sur un pavé, et que le vecteur normal extérieur qui apparaît sur la droite est *constant* sur chaque face, et de signe opposé sur les faces opposées.



Démonstration. Rappelons que la variable $i \in \{1 \dots n\}$ est celle qui apparaît dans (4.7) et donc fixé. On indexera les autres coordonnées avec la lettre k . L'intégrale de gauche de (4.7) est une intégrale sur le pavé plein d'une fonction scalaire, qui est juste donné par intégrations "successives" d'une variable après l'autre. Par re-arrangement d'ordre d'intégration, on a donc la gauche de (4.7) qui est

$$\begin{aligned} \int_{\Omega} \frac{\partial u}{\partial x_i} dx &= \left(\prod_{k \neq i} \int_{a_k}^{b_k} \right) \left(\int_{a_i}^{b_i} \frac{\partial u}{\partial x_i} dx_i \right) dx_1, \dots, dx_{i-1} dx_{i+1} \dots dx_n \\ &= \left(\prod_{k \neq i} \int_{a_k}^{b_k} \right) \left(u(x_1, \dots, x_{i-1}, b_i, x_{i+1}, \dots, x_n) - \right. \\ &\quad \left. u(x_1, \dots, x_{i-1}, a_i, x_{i+1}, \dots, x_n) \right) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n \end{aligned}$$

C'était simple. Regardons maintenant la droite de (4.7) : le bord $\partial\Omega$ est une union de $2n$ faces, et l'intégrale sur le bord est à lire comme la somme de ces $2n$ intégrales sur les $2n$ faces, qui sont données par

$$\begin{aligned} F_k^{(a)} & \stackrel{\text{def.}}{=} \prod_{j < k} [a_j, b_j] \times \{a_k\} \times \prod_{k < j} [a_j, b_j] \\ F_k^{(b)} & \stackrel{\text{def.}}{=} \prod_{j < k} [a_j, b_j] \times \{b_k\} \times \prod_{k < j} [a_j, b_j] \end{aligned}$$

pour $k = 1 \dots n$. Observons que $\nu_i(x) = +1$ sur $F_i^{(b)}$, $\nu_i(x) = -1$ sur $F_i^{(a)}$ et $\nu_i(x) = 0$ sur les autres faces. Ainsi,

$$\int_{\partial\Omega} u(\cdot) \nu_i(\cdot) dS = \int_{F_i^{(b)}} u dS - \int_{F_i^{(a)}} u dS$$

Il faut paramétrer ces deux surfaces. Soit

$$\Phi^{(a)} : \begin{cases} \prod_{k \neq i} [a_k, b_k] & \rightarrow S = \partial\Omega \\ (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) & \mapsto (x_1, \dots, x_{i-1}, a_i, x_{i+1}, \dots, x_n) \end{cases}$$

et de même $\Phi^{(b)}$, qui porte b_i à la place de a_i . Observons que $\frac{\partial \Phi^{(a)}}{\partial x_k} = e_k$ pour $k \neq i$ et de même pour $\Phi^{(b)}$. Rappelons nous que le vecteur normal extérieur ν sur les deux faces restantes $\pm e_i$. Quoi qu'il en soit, pour $x \in F_i^{(a)} \cup F_i^{(b)}$, on aura

$$|\det(\nu(\Phi(x)), \Phi_{x_1}, \dots, \Phi_{x_{i-1}}, \Phi_{x_{i+1}}, \dots, \Phi_{x_n})| = |\det(\pm e_i, e_1, \dots, e_{i-1}, e_{i+1}, \dots, e_n)| = 1$$

ce qui montre, en utilisant la définition (4.5) le résultat. \square

4.4.2 Formules de Green et le théorème de la divergence

On applique (4.7) au produit de deux fonctions scalaires $f, g : \Omega \rightarrow \mathbb{R}$: Or, $\frac{d}{dx_i}(f \cdot g) = \frac{df}{dx_i} g + f \cdot \frac{dg}{dx_i}$, ceci donne une formule "d'intégration par parties"

$$\int_{\Omega} f_{x_i} g \, dx = \int_{\partial\Omega} f g \nu_i \, dS - \int_{\Omega} f g_{x_i} \, dx$$

Si on remplace ici g par g_{x_i} et somme sur $i = 1, \dots, n$, on obtient

$$\int_{\Omega} \langle \nabla f, \nabla g \rangle \, dx = \int_{\partial\Omega} f \langle \nabla g, \nu \rangle \, dS - \int_{\Omega} f \Delta g \, dx$$

où $\Delta g = \sum_i \frac{d^2}{dx_i^2}(g)$ est le **Laplacien** de g , la trace de la Hessienne de g . On rappelle finalement que $\langle \nabla g(x), a \rangle = \frac{\partial g}{\partial a}(x)$, la dérivée directionnelle en direction a . Il suit l'autre notation de la précédente égalité,

$$\int_{\Omega} \langle \nabla f, \nabla g \rangle = \int_{\partial\Omega} f \frac{\partial g}{\partial \nu} \, dS - \int_{\Omega} f \Delta g.$$

Cette dernière identité est appelée **l'identité de Green**. Une autre conséquence de (4.7) est le **théorème de la divergence** ou le **théorème de Gauss** : soit $F : \Omega \rightarrow \mathbb{R}^n$ un champs de vecteurs, $F = (f_1, \dots, f_n)$. On pose

$$\operatorname{div}(F) := \sum_{i=1}^n \frac{\partial f_i}{\partial x_i} = \operatorname{trace}(\mathcal{J}_F)$$

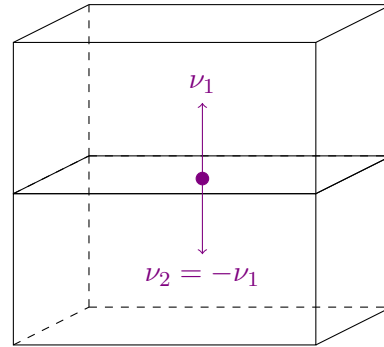
et on l'appelle la divergence. Lorsqu'on intègre $\operatorname{div}(F)$ sur Ω , on a, en sommant sur i la formule (4.7)

$$\int_{\Omega} \operatorname{div}(F) \, dx = \sum_{i=1}^n \int_{\partial\Omega} f_i \nu_i \, dS = \int_{\partial\Omega} \langle F, \nu \rangle \, dS$$

Le produit scalaire $\langle F(x), \nu(x) \rangle$ décrit le flux de F à travers la surface $\partial\Omega$ au point x . Ainsi, l'intégrale de droite mesure le changement de la quantité décrite par F à travers la surface – l'égalité nous dit que ce changement est mesuré par une quantité scalaire, $\operatorname{div}(F)$ sur Ω . Pour cette raison, le théorème de Gauss a une grande importance pour dériver des équations différentielles en utilisant lois de conservation en physique.

Pour le moment, tout repose sur le théorème 4.24. Il faudra l'étendre sur des domaines plus intéressants que des pavés. Cela nécessite de paramétrer à la fois le domaine et son bord : la preuve devient alors plus technique, et on n'a pas le temps de le faire proprement dans ce cours. Cela dit, on peut le rendre "plausible" avec une idée d'approximation.

La première observation est que si Ω est une union de deux pavés, alors les vecteurs normaux extérieurs sur les deux faces qui se touchent ont un signe opposé. Les intégrales respectives s'annulent donc mutuellement. Cela permet de construire des domaines qui sont unions de pavés et d'étendre tous les résultats à ce type de domaine "gratuitement". Reste la partie "floue" de l'argument :



Avec un peu de "handwaiving" on est prêt à accepter sans preuve formelle que l'on peut approximer des domaines suffisamment régulières par des petits pavés, comme indiqué ici, en illustration 2D

