L3 – COURS DE CALCUL DIFFÉRENTIEL

Université de Bourgogne - Année 2017–2018

Table des matières

Le propos principal du cours de Calcul Différentiel de L3 est l'étude des deux notions fondamentales suivantes :

1. Celle d'*application différentiable*. Cette notion, qui précise celle d'application continue, est cruciale en analyse comme en géométrie. Il s'agit d'étendre en dimension quelconque la notion de *fonction dérivable*, étudiée en L1. Elle est d'ailleurs déjà introduite et brièvement étudiée en L2, dans le cours de *Fonctions de Plusieurs Variables*. Elle est étudiée de façon beaucoup plus systématique en L3.
2. Celle de *sous-variété différentiable*. C'est l'objet géométrique naturellement associé aux applications différentiables. Alors que les *sous-espaces affines* sont des ensembles "rectilignes" définis comme étant l'ensemble des zéros d'applications linéaires ou affines, les sous-variétés différentiables sont des ensembles "courbés" localement définis comme l'ensemble des zéros d'applications différentiables. Il s'agit d'une collection extrêmement riche d'ensembles, d'une grande utilité. D'ailleurs, au travers de notions comme celle d'*espace tangent*, la connection entre la géométrie différentielle et la géométrie linéaire se fait naturellement.

Première partie

Espaces vectoriels normés de dimension finie

Le Calcul Différentiel admet des développements dans les espaces de dimension infinie, comme par exemple les espaces de fonctions. Cela dépasse les limites du cours de L3. En revanche, une bonne maîtrise des propriétés des espaces vectoriels normés est requise. Cela relève du programme de L2 ; au besoin, une révision s'impose. La première partie de ce cours en rappelle l'essentiel, sans redonner toutes les preuves ni rentrer dans tous les détails.

1 Normes et distances

1.1 Normes et exemples de normes

Définition 1.1.1. Une ***norme*** sur $\mathbb{R}^n$ est une fonction $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_+$ telle que, pour tous $x, y \in \mathbb{R}^n$ et tout $\lambda \in \mathbb{R}$:
1. (*homogénéité*) $\|\lambda x\| = |\lambda| \, \|x\|$.
2. (*inégalité triangulaire*)$\|x + y\| \leq \|x\| + \|y\|$.
3. $\|x\| = 0$ si et seulement si $x = 0$.

Le couple $(\mathbb{R}^n, \|\cdot\|)$ est alors appelé ***espace (vectoriel) normé***.

Définition 1.1.2. Deux normes $\|\cdot\|$ et $|||\cdot|||$ sur $\mathbb{R}^n$ sont dites ***équivalentes*** s'il existe deux réels α et β strictement positifs tels que, pour tout $x \in \mathbb{R}^n$, on ait :

$$\alpha \|x\| \leq |||x||| \leq \beta \|x\| .$$

Exemple 1.1.3. (*Exercice*)

1. Pour tout nombre $p > 1$ la fonction :

$$\|\cdot\|_p : \mathbb{R}^n \longrightarrow \mathbb{R}, \; (x_1, \ldots, x_n) \longmapsto \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}$$

est une norme sur $\mathbb{R}^n$, appelée la *norme p*. La seule propriété délicate à vérifier est l'inégalité triangulaire. On introduit pour cela le *conjugué* de p, c'est à dire le réel $q = p\,(p-1)^{-1}$ si $p > 1$, On a donc $\frac{1}{p} + \frac{1}{q} = 1$. L'inégalité triangulaire est une conséquence de l'*inégalité de Hölder* : *pour tout couple* (p, q) *de réels conjugués, et tout couple de n-uplets* $x = (x_1, \ldots, x_n)$ *et* $y = (y_1, \ldots, y_n)$ *de nombres réels, on a :*

$$\sum_{i=1}^{n} x_i y_i \leq \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} |y_i|^q \right)^{\frac{1}{q}} \; \left(= \|x\|_p \, \|y\|_q \right).$$

On note que l'inégalité classique de Cauchy-Schwartz est le cas particulier de l'inégalité de Hölder lorsque $p = q = 2$.

2. Le cas $p = 1$ est un peu différent. On convient que le conjugué de 1 est ∞, et on peut formuler l'inégalité de Hölder en introduisant la *norme infinie :*

$$\|\cdot\|_\infty : \mathbb{R}^n \longrightarrow \mathbb{R}, \; (x_1, \ldots, x_n) \longmapsto \max_{i=1,\ldots,n} |x_i| .$$

On a alors :

$$\sum_{i=1}^{n} x_i y_i \leq \sum_{i=1}^{n} |x_i| \max_{i=1,\ldots,n} |y_i| \; \left(= \|x\|_1 \, \|y\|_\infty \right).$$

Exercice 1.1.4. (*Exercice*) Montrer que les normes $\|\cdot\|_p$, $p \in [1, +\infty[\cup \{\infty\}$, sont équivalentes. Plus précisément, montrer que, pour tout $x \in \mathbb{R}^n$:

$$1 \leq p \leq q \Longrightarrow \|x\|_\infty \leq \|x\|_p \leq \|x\|_q \leq n^{\frac{1}{p}} \|x\|_\infty .$$

1.2 Distance associée à une norme et notions de topologie

Définition 1.2.1. Une ***distance*** sur $\mathbb{R}^n$ est une application $\mathrm{d} \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ telle que, pour tous $x, y, z \in \mathbb{R}^n$, on ait :
1. (*symétrie*) $\mathrm{d}(x, y) = \mathrm{d}(y, x)$.
2. (*inégalité triangulaire*) $\mathrm{d}(x, z) \leq \mathrm{d}(x, y) + \mathrm{d}(y, z)$.
3. (*séparation*) $\mathrm{d}(x, y) = 0 \Leftrightarrow x = y$.

Le couple $(\mathbb{R}^n, \mathrm{d})$ est appelé un ***espace métrique***.
On déduit aisément de la définition de norme l'énoncé suivant :

Proposition 1.2.2. *Soit* $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_+$ *une norme. Alors l'application* $\mathrm{d}_{\|\cdot\|} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ *définie par* $\mathrm{d}_{\|\cdot\|}(x, y) = \|x - y\|$ *est une distance sur* $\mathbb{R}^n$*. On appelle* $\mathrm{d}_{\|\cdot\|}$ *la* ***distance induite par la norme*** $\|\cdot\|$*.*

Remarque 1.2.3. Il est important de noter ici que les espaces normés sont des espaces métriques très particuliers :de nombreuses distances ne sont pas induites par des normes. Par exemple, une distance très simple, la *distance discrète*, définie sur $\mathbb{R}^n$ par $\mathrm{d}(x, y) = 1$ si $x \neq y$ et $\mathrm{d}(x, x) = 0$, n'est pas induite par une norme (*pourquoi ?*).

De plus, un espace normé est *nécessairement* un espace *vectoriel.* Ca n'est pas le cas des espaces métriques. Par exemple, tout sous-ensemble d'un espace métrique, muni de la distance induite, est encore un espace métrique, *même si ce n'est pas un espace vectoriel.*

La totalité du cours de Calcul Différentiel de L3 a pour cadre les espaces vectoriels $\mathbb{R}^n$ munis d'une norme. L'étude des espaces métriques généraux fait l'objet d'un autre cours.

Rappelons quelques notions classiques :

Définition 1.2.4. On considère l'espace vectoriel normé $(\mathbb{R}^n, \|\cdot\|)$. Soient $a \in \mathbb{R}^n$ et $r > 0$.
1. La ***boule ouverte*** de centre a et de rayon $r > 0$ est l'ensemble $B(a, r) = \{x \in \mathbb{R}^n : \|x - a\| < r\}$.
2. La ***boule fermée*** de centre a et de rayon $r > 0$ est l'ensemble $B_a(r) = \{x \in \mathbb{R}^n : \|x - a\| \leq r\}$.
3. La ***sphère*** de centre a et de rayon r est l'ensemble $\mathcal{S}(a, r) = \{x \in \mathbb{R}^n : \|x - a\| = r\}$. Traditionnellement, on note $\mathcal{S}^1 = \{x \in \mathbb{R}^n : \|x\| = 1\}$, qu'on appelle la ***sphère unité de centre*** $0 \in \mathbb{R}^n$.

Remarque 1.2.5. La forme des *boules* dépend de la norme considérée. *Exercice classique :étudier la forme des boules du plan pour les normes* $\|\cdot\|_p$, $p \in [1, +\infty[\cup \{\infty\}$.

Définition 1.2.6. Soit A un sous-ensemble de l'espace normé $(\mathbb{R}^n, \|\cdot\|)$.
1. L'ensemble A est ***borné*** s'il existe $a \in \mathbb{R}^n$ et $r > 0$ tel que $A \subseteq B(a, r)$.
2. L'ensemble A est ***ouvert*** si, pour tout $a \in A$, il existe $r > 0$ tel que la boule ouverte $B(a, r)$ soit contenue dans A.
3. L'ensemble A est ***fermé*** si le complémentaire $\mathbb{R}^n \setminus A$ de A dans $\mathbb{R}^n$ est un ensemble ouvert.

Exemple 1.2.7. Les ensembles ouverts sont exactement les unions quelconques de boules ouvertes.

Remarque 1.2.8. Les ensembles ouverts et fermés satisfont les propriétés suivantes :
1. La réunion d'une famille quelconque d'ensemble ouverts est un ensemble ouvert.
2. L'intersection d'une famille quelconque d'ensembles fermés est un ensemble fermé.
3. Attention, les boules ouvertes sont des exemples évidents d'ensembles ouverts, les boules fermées sont des exemples évidents d'ensembles fermés, mais les ensembles ouverts et fermés peuvent être (et sont en général) bien plus compliqués que des boules. Par exemple, les *ensembles de Cantor*, qui ne sont pas étudiés dans le cours de Calcul Différentiel de L3, sont introduits dans d'autres cours.

La notion suivante est également d'un usage fréquent :

Définition 1.2.9. Soit (E, d) un espace métrique. Un sous-ensemble $A \subseteq E$ est ***compact*** si, de toute suite d'éléments de A, on peut extraire une sous-suite convergeant dans A.

Dans le cadre des espaces normés de dimension finie qui est celui de ce cours, cette notion s'exprime plus simplement :

Proposition 1.2.10. *Un sous ensemble A d'un espace normé* <u>de dimension finie</u> *est compact si et seulement si il est fermé et borné.*

Remarque 1.2.11. Cette propriété est fausse dans les espaces vectoriels normés de dimension infinie, et, *a fortiori*, dans les espaces métriques généraux.

Exemple 1.2.12. Les boules fermées, ou les unions finies de boules fermées, sont des ensembles compacts. Les *ensembles de Cantor*, mentionnés plus haut, sont également des ensembles compacts.

La proposition suivante donne son intérêt à la notion de couple de normes équivalentes. Sa preuve, facile, est laissée en exercice :

Proposition 1.2.13. *Soient* $\|\cdot\|$ *et* $|\!|\!|\cdot|\!|\!|$ *deux normes équivalentes sur* $\mathbb{R}^n$*. Alors un ensemble* $A \subseteq \mathbb{R}^n$ *est ouvert (*resp. *fermé, compact) pour la norme* $\|\cdot\|$ *si et seulement si A est ouvert (*resp. *fermé, compact) pour la norme* $|\!|\!|\cdot|\!|\!|$*.*

2 Applications continues

Quelques rappels sur les fonctions continues sont nécessaires avant d'entamer l'étude des applications différentiables. De façon générale, dans les définitions et les énoncés, nous désignons par U un ensemble ouvert d'un espace $\mathbb{R}^n$, $n \in \mathbb{N}$.

2.1 Définitions et premières propriétés

Définition 2.1.1. On considère les espaces normés $(\mathbb{R}^p, \|\cdot\|)$ et $(\mathbb{R}^q, |||\cdot|||)$. Soit $U \subseteq \mathbb{R}^p$ un ensemble ouvert. Alors :

1. Une application $f\colon U \subseteq \mathbb{R}^p \to \mathbb{R}^q$ est ***continue*** au point $a \in U$ si :

$$\forall \varepsilon > 0, \exists \eta > 0, \forall x \in U,\ \|x - a\| < \eta \Longrightarrow |||f(x) - f(a)||| < \varepsilon.$$

2. L'application $f\colon U \subseteq \mathbb{R}^p \to \mathbb{R}^q$ est ***continue*** sur U si elle est continue en chaque point de U.

Remarque 2.1.2.

1. Une façon équivalente de formuler la continuité de f au point a est d'écrire :

$$f(a+h) = f(a) + \varepsilon(h),$$

où $a + h \in U$ est ε est une fonction définie au voisinage de $0 \in \mathbb{R}^p$ et de limite nulle en $0 \in \mathbb{R}^p$.

2. (*Exercice*) On n'altère pas la continuité d'une application en remplaçant dans l'espace de départ et dans l'espace d'arrivée les normes par des normes équivalentes. Cela permet de travailler, dans une famille de normes équivalentes, avec celle qui paraît le plus appropriée aux calculs. C'est en particulier le cas de la famille des normes $\|\cdot\|_p$, p$\in [1, \infty[\cup \{\infty\}$.

Exemple 2.1.3. 1. La famille des applications continues sur un ouvert $U \subseteq \mathbb{R}^p$ à valeurs dans $\mathbb{R}^q$ est stable par somme (et différence). Si $q = 1$, elle est également stable par produit et quotient (là où ce quotient est défini). En particulier, les applications polynomiales (et *a fortiori* les applications linéaires) sont continues. Les fractions rationnelles sont continues sur leur domaine de définition.

2. L'application $\|\cdot\| : (\mathbb{R}^p, \|\cdot\|) \to (\mathbb{R}, |\cdot|)$ est continue. En effet, il résulte de l'inégalité triangulaire que, pour $x, y \in \mathbb{R}^p$, $|\|x\| - \|y\|| \leq \|x - y\|$.

2.2 Applications continues et topologie

Exemple 2.2.1. Les fonctions continues jouent un rôle essentiel dans l'étude des propriétés topologiques :

Proposition 2.2.2. *Soit* $f\colon (\mathbb{R}^p, \|\cdot\|) \to (\mathbb{R}^q, |||\cdot|||)$ *une application continue. Alors :*

1. *L'image réciproque par f d'un sous-ensemble ouvert de $\mathbb{R}^q$ est un ouvert de $\mathbb{R}^p$.*
2. *L'image réciproque par f d'un sous-ensemble fermé de $\mathbb{R}^q$ est un fermé de $\mathbb{R}^p$.*

Remarque 2.2.3. Cette proposition est très utile (peut-être d'avantage que la définition initiale) pour déterminer si un sous-ensemble de $\mathbb{R}^p$ est ouvert (ou fermé).

Exemple 2.2.4. Soit $f\colon (\mathbb{R}^p, \|\cdot\|) \to \mathbb{R}$ une application continue. Puisque $[0, +\infty[\ \subset \mathbb{R}$ est un ensemble fermé, l'ensemble

$$\{x \in \mathbb{R}^n : f(x) \geq 0\} = f^{-1}([0, +\infty[)$$

est un sous-ensemble fermé de $\mathbb{R}^p$.

Théorème 2.2.5 (ensembles compacts et applications continues)**.**

1. *L'image (*directe*) d'un ensemble compact par une application continue est également un ensemble compact.*
2. *Soient $A \subset \mathbb{R}^n$ un ensemble compact et $f\colon A \to \mathbb{R}$ une application continue. Alors f est bornée, et "atteint ses bornes" : il existe $a \in A$ et $b \in A$ tels que :*

$$f(a) = \max_{x \in A} f(x) \ \text{ et } f(b) = \min_{x \in A} f(x).$$

Nous avons vu plus haut que toutes les normes $\|\cdot\|_p$ sont équivalentes. En fait *toutes les normes* sur $\mathbb{R}^n$ sont équivalentes, comme l'affirme l'énoncé suivant :

Proposition 2.2.6 (équivalence des normes dans $\mathbb{R}^n$)**.** *Soit $n \geq 1$. Toutes les normes sur $\mathbb{R}^n$ sont équivalentes.*

Démonstration. On considère une norme $\|\cdot\|$ sur $\mathbb{R}^n$ et on montre qu'elle est équivalente à la norme $\|\cdot\|_\infty = \max_{i=1,\ldots,n} |x_i|$. Pour cela, on désigne par $(\mathrm{e}_1, \ldots, \mathrm{e}_n)$ la base canonique de $\mathbb{R}^n$. Pour tout $x = \sum_{i=1}^n x_i \mathrm{e}_i \in \mathbb{R}^n$, on a :

$$\|x\| = \left\| \sum_{i=1}^n x_i \mathrm{e}_i \right\| \leq \sum_{i=1}^n |x_i| \, \|\mathrm{e}_i\| \leq \beta \, \|x\|_\infty \,,$$

où $\beta = \sum_{i=1}^n \|\mathrm{e}_i\|$. Il reste à trouver un nombre $m > 0$ tel que $m \, \|x\|_\infty \leq \|x\|$ pour tout $x \in \mathbb{R}^n$. Or, on déduit de l'inégalité précédente que, pour tous $x, y \in \mathbb{R}^n$:

$$|\|x\| - \|y\|| \leq \|x - y\| \leq \beta \, \|x - y\|_\infty \,,$$

et donc que $\|\cdot\| : (\mathbb{R}^n, \|\cdot\|_\infty) \to (\mathbb{R}, |\cdot|)$ est une fonction continue. Or la sphère unité $\mathcal{S} = \{x \in \mathbb{R}^n : \|x\|_\infty = 1\}$, qui est un ensemble fermé de $(\mathbb{R}^n, \|\cdot\|_\infty)$ en tant qu'image réciproque du singleton $\{1\} \subset \mathbb{R}$ par $\|\cdot\|_\infty$ et évidemment borné est un compact de $(\mathbb{R}^n, \|\cdot\|_\infty)$. Donc l'application $\|\cdot\|$, continue sur $(\mathbb{R}^n, \|\cdot\|_\infty)$ est bornée sur $\mathcal{S}$ et atteint son minimum m en un point $b \in \mathcal{S}$:

$$\|b\| = \min_{x \in \mathcal{S}} \|x\| = m.$$

Or, pour tout $x \in \mathbb{R}^n \setminus \{0\}$, le vecteur $\frac{1}{\|x\|_\infty} x$ appartient à $\mathcal{S}$. Donc $m \leq \frac{1}{\|x\|_\infty} \|x\|$ ou encore $m \, \|x\|_\infty \leq \|x\|$.

Nous avons montré que, pour tout $x \in \mathbb{R}^n$, $m \, \|x\|_\infty \leq \|x\| \leq \beta \, \|x\|_\infty$, c'est à dire que les normes $\|\cdot\|$ et $\|\cdot\|_\infty$ sont équivalentes. $\square$

Remarque 2.2.7. Dorénavant, grâce à la proposition précédente, nous ne mentionnons plus de norme dans nos énoncés ultérieurs. En revanche, cette proposition nous permet, au détour des preuves, d'utiliser la norme de notre choix en fontion de sa commodité. En particulier, nous pouvons nous "libérer" de la classique norme euclidienne, qui n'est pas nécessairement la plus appropriée pour les calculs ou les estimations.

3 Normes d'applications linéaires

Nous avons déjà mentionné que les applications linéaires entre les espaces $\mathbb{R}^n$, $n \in \mathbb{N}$, sont continues. Il s'avère que ces applications jouent un rôle particulier en Calcul Différentiel, il importe de bien les comprendre.

3.1 Normes subordonnées et normes matricielles

L'ensemble $\mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$ des applications linéaires de $\mathbb{R}^p$ dans $\mathbb{R}^q$ est un espace vectoriel sur $\mathbb{R}$ isomorphe à $\mathbb{R}^{pq}$. Cette affirmation est facile à voir. En effet, une fois fixées une base $(\mathrm{e}) = (\mathrm{e}_1, \ldots, \mathrm{e}_p)$ de $\mathbb{R}^n$ et une base $(\varepsilon) = (\varepsilon_1, \ldots, \varepsilon_q)$ de $\mathbb{R}^q$, on peut représenter une application linéaire $L \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$ à l'aide d'une ***matrice*** : il s'agit du tableau de nombres à p colonnes et q lignes obtenu en écrivant en colonnes consécutives les coordonnées dans la base (ε) des images par L des p vecteurs de la base (e). Ce tableau de pq nombres est un élément de $\mathbb{R}^{pq}$. Naturellement, pour des raisons pratiques liées aux manipulations sur les matrices en lien avec les opérations sur les applications linéaires (par exemple la composition des applications linéaires est correspond au produit des matrices) on ne représente pas ces pq nombres par une colonne, mais par un tableau.

Notation 3.1.1. On désigne par $\mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$ l'espace des applications linéaires de $\mathbb{R}^p$ dans $\mathbb{R}^q$, et par $\mathcal{M}_{q,p}(\mathbb{R})$ l'espace des matrices à q lignes et p colonnes. Si $p = q$, on note ces espaces respectivement $\mathcal{L}(\mathbb{R}^p)$ et $\mathcal{M}_p(\mathbb{R})$.

Remarque 3.1.2. On retient la règle simple mais importante : lorsqu'une matrice de $\mathcal{M}_{q,p}(\mathbb{R})$ représente une application linéaire de $\mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$ dans des bases, *le nombre de colonnes est la dimension de l'espace de départ.*

On peut ainsi munir l'espace $\mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$ des mêmes normes que l'espace $\mathbb{R}^{pq}$. Ces normes sont toutes equivalentes. Néanmoins, certaines de ces normes sont plus appropriées que d'autres pour l'analyse :

Proposition 3.1.3. *On munit l'espace $\mathbb{R}^p$ d'une norme $\|\cdot\|$ et l'espace $\mathbb{R}^q$ d'une norme $|\!|\!|\cdot|\!|\!|$. Alors :*

1. *l'application :*
$$N \colon \mathcal{L}(\mathbb{R}^p, \mathbb{R}^q) \longrightarrow \mathbb{R}_+, \; L \longmapsto \max_{\|x\| \leq 1} |\!|\!|L(x)|\!|\!|$$
*est une norme sur $\mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$, dite **subordonnée aux normes** $\|\cdot\|$ et $|\!|\!|\cdot|\!|\!|$.*
2. *Si $p = q$, et qu'on munit $\mathbb{R}^p$ à la source et au but de la même norme $\|\cdot\|$, on dit que la norme N définie comme ci-dessus est **subordonnée** à la norme $\|\cdot\|$.*

Remarque 3.1.4. La preuve de cette proposition est laissée en exercice. On note que le maximum de la définition est bien réalisé, car $\{x \in \mathbb{R}^n : \|x\| \leq 1\}$ est un ensemble fermé borné, donc compact, de $\mathbb{R}^n$, et L est une application continue.

Exercice 3.1.5. Montrer que la définition ci-dessus est équivalente à $N(x) = \max_{\|x\|=1} |||L(x)|||$ (au lieu de prendre le maximum sur la boule unité fermée, on peut prendre le maximum sur la sphère unité).

L'intérêt des normes subordonnées vient de la propriété suivante :

Proposition 3.1.6. *On munit l'espace $\mathbb{R}^p$ d'une norme $\|\cdot\|$ et l'espace $\mathbb{R}^q$ d'une norme $|||\cdot|||$. On note N la norme sur $\mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$ subordonnée aux normes $\|\cdot\|$ et $|||\cdot|||$. Alors, pour toute application linéaire $L \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$, et tout $x \in \mathbb{R}^p$ on a :*

$$|||L(x)||| \leq N(L)\|x\|.$$

On déduit de cette proposition que les normes bien se comportent bien pour les compositions :

Proposition 3.1.7. *On considère les espaces $\mathbb{R}^p$, $\mathbb{R}^q$ et $\mathbb{R}^r$, munis respectivement des normes $\|\cdot\|$, $|||\cdot|||$ et $|||\cdot|||'$. On note N la norme subordonnée à $\|\cdot\|$ et $|||\cdot|||$, N' la norme subordonnée à $|||\cdot|||$ et $|||\cdot|||'$, et N'' la norme subordonnée à $\|\cdot\|$ et $|||\cdot|||'$. Alors :*

1. *pour tout $L \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$ et tout $L' \in \mathcal{L}(\mathbb{R}^q, \mathbb{R}^r)$:*

$$N''(L' \circ L) \leq N'(L')\, N(L).$$

2. *En particulier, si $p = q = r$, si $\|\cdot\| = |||\cdot||| = |||\cdot|||'$ et si N désigne la norme subordonnée à $\|\cdot\|$, on a, pour tous $L, L' \in \mathcal{L}(\mathbb{R}^p)$:*

$$N(L' \circ L) \leq N(L')\, N(L).$$

Remarque 3.1.8.
1. Une norme N sur l'espace $\mathcal{M}_n(\mathbb{R})$ est dite *matricielle* $N(AB) \leq N(A)\, N(B)$ pour tous $A, B \in \mathcal{M}_n(\mathbb{R})$.
2. De même que pour les applications linéaires, on définit sur $\mathcal{M}_n(\mathbb{R})$ la norme N subordonnée à une norme $\|\cdot\|$ sur $\mathbb{R}^n$ par :

$$\forall x \in \mathbb{R}^n,\ N(A) = \max_{\|x\|\leq 1} \|Ax\| = \max_{\|x\|=1} \|Ax\|.$$

   Ici, $A \in \mathcal{M}_n(\mathbb{R})$, x et Mx sont vus comme des *vecteurs colonnes* de $\mathbb{R}^n$.

3.2 Exemples de normes matricielles

Exercice 3.2.1. Dans ces exemples, on considère une matrice carrée $A \in \mathcal{M}_n(\mathbb{R})$. On a donc $A = (a_{ij})_{i,j=1,\ldots,n}$, $a_{ij} \in \mathbb{R}$, où i est l'indice de ligne et j l'indice de colonne. On demande de décrire en termes des éléments de la matrice A les différentes normes sur $\mathcal{M}_n(\mathbb{R})$ subordonnées aux normes classiques sur $\mathbb{R}^n$:

1. $$\|\cdot\|_\infty \longrightarrow \|A\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|.$$

2. $$\|\cdot\|_1 \longrightarrow \|A\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}|.$$

3. $$\|\cdot\|_2 \longrightarrow \|A\|_2 = \sqrt{\rho\left({}^tAA\right)}$$

   où ρ est le *rayon spectral* de la matrice symétrique ${}^tAA$, c'est à dire sa plus grande valeur propre (elle est positive).

Remarque 3.2.2. Il est facile de voir (*exercice*) que pour une norme subordonnée N, alors $N(\mathrm{Id}_n) = 1$, où Id_nest l'identité de $\mathcal{L}(\mathbb{R}^n)$ (ou l'identité de $\mathcal{M}_n(\mathbb{R})$).

Exercice 3.2.3. Montrer qu'une norme matricielle n'est pas nécessairement une norme subordonnée. Pour cela, on considère la *norme de Frobenius* sur $\mathcal{M}_n(\mathbb{R})$, définie par $N_F(A) = \sqrt{\mathrm{tr}\left({}^tAA\right)}$ pour tout $A \in \mathcal{M}_n(\mathbb{R})$.
1. Montrer que N_F est une norme sur $\mathcal{M}_n(\mathbb{R})$.
2. Montrer que N_F est une norme matricielle (on pourra pour cela utiliser l'*inégalité de Cauchy-Schwartz* sur $\mathbb{R}^{n^2}$.

3. Montrer que N_F n'est pas une norme subordonnée.

Deuxième partie

Applications différentiables

Cette partie est consacrée à l'introduction de la première notion fondamentale du cours de Calcul Différentiel :celle d'application différentiable. Plus précisément :

1. Nous définissons cette notion. Nous monstrons de quelle manière elle précise la notion de fonction continue.
2. Nous l'illustrons par plusieurs exemples, de nature variée.
3. Nous étudions les propriétés générales des applications différentiables

4 Définitions et exemples

Sauf s'il y a besoin de précision, nous noterons indifféremment $\|\cdot\|$ les normes sur les espaces $\mathbb{R}^n$, $n \in \mathbb{N}$. De même, nous noterons $\|\cdot\|$ les normes subordonnées sur les espaces $\mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$.

4.1 Applications différentiables, notion de différentielle d'une application

Définition 4.1.1. Soient U un sous-ensemble ouvert de $\mathbb{R}^p$ et $f : U \subset \mathbb{R}^p \to \mathbb{R}^q$.

1. La fonction f est dite ***différentiable au point*** $a \in U$ s'il existe une application linéaire $L \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$, un nombre réel $r > 0$ tel que $B(a, r) \subset U$, et une application $\varepsilon : B(0, r) \subset \mathbb{R}^p \to \mathbb{R}^q$ de limite nulle en 0 telle que :
$$f(a+h) = f(a) + L(h) + \|h\| \varepsilon(h) \ , \forall h \in B_0(\varepsilon) .$$
2. La fonction est dite ***différentiable sur*** U si elle est différentiable en tout point de U.

Remarque 4.1.2. On montre facilement (*exerice*) que la notion de fonction différentiable n'est pas altérée si on remplace la norme $\|\cdot\|$ de la définition par une norme équivalente. Or nous savons depuis la Partie I que *toutes les normes de* $\mathbb{R}^n$ *sont équivalentes.* Il est donc dorénavant inutile de mentionner la norme dans les énoncés de différentiabilité. En revanche, comme pour les fonctions continues, il peut être bon de faire un choix astucieux de norme pour rendres commodes certains calculs.

Commençons par une proposition évidente, qui souligne que la différentiabilité précise la continuité :

Proposition 4.1.3. *Si une application $f : U \subseteq \mathbb{R}^p \to \mathbb{R}^q$ est différentiable au point $a \in U$, alors f est continue au point a.*

Démonstration. En effet, l'application $\eta : h \mapsto L(h) + \|h\| \varepsilon(h)$ est de limite nulle en $0 \in \mathbb{R}^p$. Donc $f : h \mapsto f(a) + \eta(h)$ est bien continue au point a. $\square$

Remarque 4.1.4. Nous voyons là le point précis de la différentiabilité, par rapport à la continuité. Dire que f est continue au point a signifie que $f(a+h)$ est une "petite" perturbation de $f(a)$. Dire que f est différentiable au point a donne une précision essentielle sur la nature de cette petite perturbation :elle est de l'ordre de $L(h)$, où L est une application linéaire.

Exemple 4.1.5 (Contrexemple à la différentiabilité)**.** La fonction $x \to \sqrt{x}$ est continue en $0 \in \mathbb{R}$. Est-elle différentiable en ce point ? Si oui, il existerait $c \in \mathbb{R}$ et une fonction ε de limite nulle en $0 \in \mathbb{R}$ tels que, pour tout $h > 0$ assez petit :$\sqrt{h} = \sqrt{0} + Ch + |h| \varepsilon(h) = Ch + |h| \varepsilon(h)$, et donc $h^{-\frac{1}{2}} = C + \varepsilon(h)$. Cela impliquerait que $h^{-\frac{1}{2}} \underset{h \to 0^+}{\longrightarrow} C$, alors qu'on sait que $h^{-\frac{1}{2}} \underset{h \to 0^+}{\longrightarrow} +\infty$. Ainsi $\sqrt{h}$ est bien une petite perturbation de $\sqrt{0}$, mais elle n'est pas de l'ordre d'un terme linéaire Ch.

Notation 4.1.6 (la notation o). Considérons une application $g : 0 \in U \subseteq \mathbb{R}^n \to \mathbb{R}^p$. On dit qu'une application $\ell : U \subseteq \mathbb{R}^n \to \mathbb{R}^p$ est un $o(g)$ (au voisinage de 0), et on note $\ell = o(g)$, si $\ell = \|g\| \varepsilon$ où ε est une application de limite nulle à l'origine. De façon équivalente, $\ell = o(g)$ si $\|\ell\| \leq \|g\| \eta$, où η est une fonction positive de limite nulle à l'origine.

Ainsi, dans la Définition 4.1.1, on aurait pu noter $f(a+h) = f(a) + L(h) + o(h)$. Par exemple, si $h \in \mathbb{R}$, $h^2 = o(h)$.

Naturellement, la notation o s'étend au voisinage de tout point, où de l'infini.

Exemple 4.1.7. Une application g telle qu'il existe $c > 0$ avec $\|g(h)\| \leq c\|h\|^2$ au voisinage de 0 est un $o(h)$.

Remarque 4.1.8. La notion $o(h)$ est soumise à une arithmétique particulière. On a notamment $o(h) + o(h) = o(h)$, et $-o(h) = o(h)$.

Proposition 4.1.9. *L'application linéaire L de la Définition 4.1.1 est unique.*

Démonstration. Supposons en effet qu'il existe deux applications linéaires L_1 et L_2 vérifiant la Définition 4.1.1, et notons $L = L_1 - L_2$. On a alors :

$$0 = L(h) + o(h) \text{ et donc } L(h) = o(h).$$

On montre que cela implique que L est nulle. En effet, soit $h_0 \in \mathcal{S}^1$ tel que $\|L(h_0)\| = \|L\|$ (où l'on utilise la même notation $\|\cdot\|$ pour une norme sur $\mathbb{R}^n$ et la norme subordonnée à $\|\cdot\|$ sur $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$). Pour $t \in \mathbb{R}$ suffisament petit, il résulte de l'hypothèse que $\dfrac{\|L(th_0)\|}{|t|\,\|h_0\|} \underset{t \to 0}{\longrightarrow} 0$. Mais d'autre part $\dfrac{\|L(th_0)\|}{|t|\,\|h_0\|} = \|L(h_0)\| = \|L\|$. Donc $\|L\| = 0$, ce qui implique que $L = 0$ et donc $L_1 = L_2$. $\square$

Définition 4.1.10. Soit $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^p$.

1. Si $f$est différentiable au point a, on appelle l'unique application linéaire L de la Définition 4.1.1 la ***différentielle de f au point*** a. On la note $\mathrm{d}_a f$ [a]
2. Si f est différentiable sur U, l'application $\mathrm{d}f\colon U \to \mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$, $x \mapsto \mathrm{d}_x f$ s'appelle ***l'application différentielle*** de f sur U.

a. On trouve diverses notations dans les livres, comme $D_a f$ ou $\mathrm{d}f(a)$.

Remarque 4.1.11.

1. Il faut bien garder à l'esprit que $\mathrm{d}_a f \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$. C'est une chose importante à noter. Alors que l'application η de la Définition 4.1.1 ne s'applique qu'aux éléments h de $B(0,\varepsilon)$, la différentielle $\mathrm{d}_a f$, en tant qu'application linéaire, *s'applique à tous les vecteurs de $\mathbb{R}^n$, et pas seulement aux éléments de $B(0,\varepsilon)$*. Il importe de faire la distinction entre les rôles des différents éléments de cette définition. La fonction f ne s'applique qu'à des *points de U*, comme le point a ou le point $a+h$. *Notons bien que même si f n'est définie que sur l'ouvert U, la différentielle* $\mathrm{d}_a f$ *s'applique à tous les vecteurs de $\mathbb{R}^n$.*
2. Il ne faut pas faire de confusion entre la différentielle $\mathrm{d}_a f$ de f au point a, qui est une application linéaire de $\mathbb{R}^p$ à valeurs dans $\mathbb{R}^p$, et l'application différentielle de f qui est une application (non linéaire en général) définie sur U à valeurs dans $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$.
3. Le cas $n = 1$ est un peu particulier. En effet, une application $L \in \mathcal{L}(\mathbb{R}, \mathbb{R}^q)$ peut être identifiée à un *vecteur* de $\mathbb{R}^q$ par $L \leftrightarrow L(1) \in \mathbb{R}^q$. En effet, pour tout $t \in \mathbb{R}$, on a $L(t) = tL(1)$ et donc l'image par L de tout réel t est le multiple du vecteur $L(1)$ par t lui-même. Ainsi, si on considère une application différentiable $f\colon I \subseteq \mathbb{R} \to \mathbb{R}^q$, où I est in intervalle ouvert de $\mathbb{R}$, alors l'image $f(I)$ est une ***courbe*** de $\mathbb{R}^p$. Dans ce cas, avec $t_0 \in I$, la différentielle $\mathrm{d}_{t_0} f$ est considérée comme le ***vecteur dérivé*** de f au point t_0, et on peut l'écrire $f'(t_0) \in \mathbb{R}^q$. C'est un ***vecteur tangent*** à la courbe $f(I)$ au point $f(t_0)$. Dans cette situation, au lieu de parler la différentielle de f au point a, on peut parler plus classiquement de la dérivée de f au point a, et noter $f'(a) \in \mathbb{R}^q$.
   En mécanique, ce vecteur est le ***vecteur vitesse*** du point mobile au temps t_0, et $f(I)$ est la ***trajectoire*** du point mobile.

Exercice 4.1.12 (à faire)**.** Une application $f = (f_1, \ldots, f_q) : U \subseteq \mathbb{R}^p \to \mathbb{R}^q$ est différentiable en un point $a \in U$ si et seulement si f_i est différentiable au point a pour $i = 1, \ldots, q$.

Définition 4.1.13. L'application $f\colon U \subset \mathbb{R}^p \to \mathbb{R}^q$ est dite ***continûment différentiable*** (ou de ***classe*** $\mathcal{C}^1$) sur U si elle est différentiable sur U et si l'application différentielle $\mathrm{d}f\colon U \to \mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$ est continue.

Exercice 4.1.14. Une application $f = (f_1, \ldots, f_q) : U \subseteq \mathbb{R}^p \to \mathbb{R}^q$ est de classe $\mathcal{C}^1$ sur U si et seulement si f_i est de classe $\mathcal{C}^1$ sur U pour $i = 1, \ldots, q$.

Exemple 4.1.15. Tout application linéaire $L \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$ est différentiable et de classe $\mathcal{C}^1$ sur $\mathbb{R}^p$. En effet, si $a \in \mathbb{R}^p$ et $h \in \mathbb{R}^p$, alors :

$$L(a+h) = L(a) + L(h).$$

On retrouve exactement le développement de $\mathrm{L}(a+h)$ de la Définition 4.1.1, avec un reste $o(h) = 0$. Donc $f$est différentiable au point a et $\mathrm{d}_a f = L$. Donc $\mathrm{d}f$ est l'application constante (donc continue) sur $\mathbb{R}^p$, qui à tout $a \in \mathbb{R}^p$ associe L.

Proposition 4.1.16 (Somme d'applications différentiables)**.** *Soient $f, g \colon U \subseteq \mathbb{R}^p \to \mathbb{R}^q$.*

1. *Si f et g sont différentiables au point $a \in U$, alors $f+g$ est différentiable au point a, et*

$$\mathrm{d}_a(f+g) = \mathrm{d}_a f + \mathrm{d}_a g.$$

2. *Si f et g sont différentiables sur U, alors $f+g$ est différentiables sur U et $\mathrm{d}(f+g) = \mathrm{d}f + \mathrm{d}g$.*
3. *Si f et g sont de classe $\mathcal{C}^1$ sur U, alors $f+g$ est de classe $\mathcal{C}^1$ sur U.*

Démonstration. Exercice ! □

4.2 Dérivées directionnelles et dérivées partielles

Définition 4.2.1. Soient $f : U \subseteq \mathbb{R}^p \to \mathbb{R}^q$ et $a \in U$.

1. Pour tout vecteur $v \in \mathbb{R}^p$, la ***dérivée directionnelle de f au point a dans la direction v*** (ou encore ***dérivée de f au point a selon le vecteur v***) est, s'il existe, le vecteur :

$$f'_v(a) = \lim_{t \to 0} \frac{1}{t}\left(f(a+tv) - f(a)\right) \in \mathbb{R}^q.$$

   On dit alors que f est ***dérivable au point a dans la direction v.***
2. Si on note $(\mathrm{e}) = (\mathrm{e}_1, \ldots, \mathrm{e}_p)$ la base canonique de $\mathbb{R}^p$, alors, pour $j = 1, \ldots, p$, la ***j-ème dérivée partielle de f au point a*** est, s'il existe le vecteur :

$$\lim_{t \to 0} \frac{1}{t}\left(f(a+t\mathrm{e}_j) - f(a)\right) = \lim_{t \to 0} \frac{1}{t}\left(f(a_1, \ldots, a_j + t, \cdots, a_p) - f(a)\right) \in \mathbb{R}^q.$$

   C'est donc la dérivée directionnelle de f au point a dans la direction e_i. On note ce vecteur $\dfrac{\partial f}{\partial x_j}(a)$ ou $f'_{x_j}(a)$.

Remarque 4.2.2. La dérivée directionnelle de f dans la direction v est la dérivée en $0 \in \mathbb{R}$ de l'application $t \mapsto f(a+tv)$.

Proposition 4.2.3. *Soit $f \colon U \subset \mathbb{R}^p \to \mathbb{R}^q$ est différentiable au point $a \in U$. Alors :*

1. *Pour tout $v \in \mathbb{R}^p$, la dérivée directionnelle de f au point a dans la direction v est égale à $\mathrm{d}_a f(v)$.*
2. *L'application f admet des dérivées partielles au point a et on a, pour tout $h = (h_1, \ldots, h_p) \in \mathbb{R}^p$:*

$$\mathrm{d}_a f(h) = \sum_{j=1}^{p} h_j \frac{\partial f}{\partial x_j}(a) = \sum_{j=1}^{p} h_j f'_{x_j}(a) \in \mathbb{R}^q.$$

3. *Notons $f = (f_1, \ldots, f_q)$ les* composantes *de f. La matrice de la différentielle* $\mathrm{d}_a f$ *dans les bases canoniques de $\mathbb{R}^p$ et $\mathbb{R}^q$ est la* ***matrice jacobienne*** *:*

$$J_a(f) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(a) & \cdots & \dfrac{\partial f_1}{\partial x_p}(a) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_q}{\partial x_1}(a) & \cdots & \dfrac{\partial f_q}{\partial x_p}(a) \end{pmatrix} \in \mathcal{M}_{q,p}(\mathbb{R}).$$

Démonstration. **1.** On a, en posant $h = tv$ pour t au voisinage de $0 \in \mathbb{R}$:

$$f(a+tv) = f(a) + \mathrm{d}_a f(tv) + \|tv\| \varepsilon(tv) = f(a) + t\mathrm{d}_a f(v) + |t| \|v\| \varepsilon(tv)$$
$$\frac{f(a+tv) - f(a)}{t} = \mathrm{d}_a f(v) + \frac{|t|}{t} \|v\| \varepsilon(tv) \underset{t\to 0}{\longrightarrow} \mathrm{d}_a f(v).$$

Donc $f'_v(a) = \mathrm{d}_a f(v)$.

2. Pour $j = 1, \ldots, p$, la j-ème dérivée partielle de f au point a est par définition la dérivée directionnelle de f au point $a$dans la direction e_j. Donc d'après le point précédent :

$$\frac{\partial f}{\partial x_j}(a) = \mathrm{d}_a f(\mathrm{e}_j) \in \mathbb{R}^q.$$

Or $h = (h_1, \ldots, h_p) = \sum_{j=1}^{p} h_j \mathrm{e}_j$. Donc :

$$\mathrm{d}_a f(h) = \mathrm{d}_a f\left(\sum_{j=1}^{p} h_j \mathrm{e}_j\right) = \sum_{j=1}^{p} h_j \mathrm{d}_a f(\mathrm{e}_j) = \sum_{j=1}^{p} h_j \frac{\partial f}{\partial x_j}(a).$$

3. On note $(\mathrm{e}) = (\mathrm{e}_1, \ldots, \mathrm{e}_p)$ et $(\varepsilon) = (\varepsilon_1, \ldots, \varepsilon_q)$ les bases canoniques respectives de $\mathbb{R}^p$ et $\mathbb{R}^q$. Ainsi, $f = \sum_{i=1}^{q} f_i \varepsilon_i$. Donc $\mathrm{d}_a f = \sum_{i=1}^{q} \mathrm{d}_a f_i \varepsilon_i$ (*attention, ne pas perdre le fil : puisque f_i est une fonction définie sur $U \subseteq \mathbb{R}^p$ à valeurs réelles, $\mathrm{d}_a f_i \in \mathcal{L}(\mathbb{R}^p, \mathbb{R})$; d'autre part, $\varepsilon_i \in \mathbb{R}^q$, donc $\mathrm{d}_a f_i \varepsilon_i \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$*). Donc, pour tout $j = 1, \ldots, p$, on a :

$$\mathrm{d}_a f(\mathrm{e}_j) = \sum_{i=1}^{p} \mathrm{d}_a f_i(\mathrm{e}_j) \varepsilon_i = \sum_{i=1}^{p} \frac{\partial f_i}{\partial x_j}(a) \varepsilon_i.$$

Donc, les éléments de la j-ème colonne de la matrice de $\mathrm{d}_a f$ dans les bases canoniques, qui est constituée des composantes de $\mathrm{d}_a f(\mathrm{e}_j)$ dans la base (ε), sont le nombres $\frac{\partial f_i}{\partial x_j}(a)$. □

4.3 Exemples

Exemple 4.3.1. On considère l'application $f : \mathbb{R}_d[X] \to \mathbb{R}$ définie par $f(P) = \int_0^1 P(t)^2 \, \mathrm{d}t$, où $\mathbb{R}_d[X]$ est l'espace des polynômes à coefficients réels de degré inférieur ou égal à d.

1. Montrer que f est différentiable sur $\mathbb{R}_d[X]$.
2. Donner ses dérivées partielles pour tout $P \in \mathbb{R}_d[X]$.
3. Ecrire la matrice jacobienne de f au point 1.

Réponses :

1. Notons que $\mathbb{R}_d[X]$ est un espace vectoriel réel de dimension $d+1$, et donc isomorphe à $\mathbb{R}^{d+1}$ par l'application $P(X) = a_d X^d + \cdots + a_1 X + a_0 \mapsto (a_0, a_1, \ldots, a_d) \in \mathbb{R}^{d+1}$. Pour répondre à la première question, on applique directement la Définition 4.1.1, c'est à dire qu'on étudie l'accroissement $f(P+H)$, où $P \in \mathbb{R}_d[X]$ et H est un "petit" élément de $\mathbb{R}_d[X]$. On a :

$$f(P+H) = \int_0^1 (P(t) + H(t))^2 \, \mathrm{d}t = \int_0^1 P(t)^2 \, \mathrm{d}t + 2\int_0^1 P(t) H(t) \, \mathrm{d}t + \int_0^1 H(t)^2 \, \mathrm{d}t.$$

   Le premier terme de cette somme $f(P)$, et le deuxième terme est donné par l'application linéaire $L \colon \mathbb{R}_d[X] \to \mathbb{R}$ définie par $L(H) = 2\int_0^1 P(t) H(t) \, \mathrm{d}t$. Donc, afin de vérifier qu'on satisfait bien la définition d'application différentiable, nous devons prouver que le dernier terme est un $o(H)$ lorsque H tend vers $0 \in \mathbb{R}_d[X]$. Le "signe favorable" est le fait que H apparaisse élevé au carré dans ce dernier terme. Pour cela, rappelons qu'on peut munir $\mathbb{R}_d[X]$ de la norme de notre choix. C'est le moment de le faire. Pour $P \in \mathbb{R}_d[X]$, nous posons $\|P(X)\|_\infty = \sup_{t\in[0,1]} |P(t)|$. On vérifie (*exercice*) qu'il s'agit d'une norme sur $\mathbb{R}_d[X]$. On a alors :

$$\left| \int_0^1 H(t)^2 \, \mathrm{d}t \right| \leq \|H\|_\infty^2,$$

   ce qui implique que $\int_0^1 H(t)^2 \, \mathrm{d}t = o(H)$, et que f est différentiable au point $P \in \mathbb{R}_d[X]$.
2. Ses dérivées partielles en P sont, d'après la Proposition 4.2.3 :

$$\mathrm{d}_P f(1) = 2\int_0^1 P(t) \, \mathrm{d}t, \ \mathrm{d}_P f(X) = 2\int_0^1 tP(t) \, \mathrm{d}t, \ \mathrm{d}_P f\left(X^2\right) = 2\int_0^1 t^2 P(t) \, \mathrm{d}t, \ldots$$

3. Si $P = 1$, on a :

$$\mathrm{d}_1 f\left(X^j\right) = 2\int_0^1 t^j \mathrm{d}t = \frac{2}{j+1}.$$

Donc :

$$J_1 f = 2 \cdot \left(1, \frac{1}{2}, \ldots, \frac{1}{d+1}\right).$$

Exemple 4.3.2. Si $f\colon U \subset \mathbb{R}^p \to \mathbb{R}$, on a : $\mathrm{d}_a f(h) = f'_{x_1}(a) h_1 + \cdots + f'_{x_p}(a) h_p$. Soit par exemple $f\colon \mathbb{R}^2 \to \mathbb{R}$ définie par $f(x,y) = \mathrm{e}^x + \sin x \cos y$. On a :

$$f'_x(x,y) = \mathrm{e}^x + \cos x \cos y, \quad f'_y(x,y) = -\sin x \sin y.$$

Donc $J_{(0,0)} f = \begin{pmatrix} 2 & 0 \end{pmatrix}$ et $\mathrm{d}_{(0,0)} f(h_1, h_2) = 2h_1$.

Exemple 4.3.3. Une fonction peut avoir des dérivées directionnelles en un point dans toutes les directions sans différentiable, *ni même continue*, en ce point. Par exemple, la fonction $f\colon \mathbb{R}^2 \to \mathbb{R}$ définie par :

$$f(x,y) = \begin{cases} \dfrac{y^2}{x} & \text{si } x \neq 0 \\ 0 & \text{sinon} \end{cases}$$

est dérivable en $(0,0) \in \mathbb{R}^2$ dans toutes les directions. En effet, si $v = (v_1, v_2)$, on a :

$$\begin{aligned} \text{si } v_1 \neq 0,\ \lim_{t\to 0} \frac{f(tv_1, tv_2) - f(0,0)}{t} &= \lim_{t\to 0} \frac{t^2 v_2^2}{t^2 v_1} = \frac{v_2^2}{v_1}, \\ \text{si } v_1 = 0,\ \lim_{t\to 0} \frac{f(tv_1, tv_2) - f(0,0)}{t} &= \lim_{t\to 0} \frac{f(0, tv_2) - f(0,0)}{t} = 0. \end{aligned}$$

Cependant, f n'est pas continue à l'origine. En effet, la limite en $(0,0)$ de f le long des axes de coordonnées est nulle, alors que la limite de f en $(0,0)$ le long de la parabole d'équation $\left\{x = y^2\right\}$ est égale à 1.

5 Premières propriétés des applications différentiables

5.1 Propriétés algébriques et composition

Nous avons vu Proposition 4.1.16 que la somme de deux applications différentiables est différentiables. Il y a d'autres propriétés algébriques similaires :

Proposition 5.1.1 (Produit et quotient de fonctions différentiables)**.** *Soient $f, g\colon U \subseteq \mathbb{R}^p \to \mathbb{R}$ deux applications différentiables au point $a \in U$. Alors :*

1. *Le produit fg est différentiable au point a, et*

$$\mathrm{d}_a(fg) = g(a)\,\mathrm{d}_a f + f(a)\,\mathrm{d}_a g.$$

2. *Si $g(a) \neq 0$, le quotient $\frac{f}{g}$ est différentiable au point a, et*

$$\mathrm{d}_a\left(\frac{f}{g}\right) = \frac{1}{g^2(a)}\left(g(a)\,\mathrm{d}_a f - f(a)\,\mathrm{d}_a g\right).$$

Proposition 5.1.2 (Composition d'applications différentiables)**.** *Soient $f\colon U \subset \mathbb{R}^p \to \mathbb{R}^q$ et $g\colon V \subset \mathbb{R}^q \to \mathbb{R}^r$ deux applications telles que f soit différentiable au point $a \in U$, g soit différentiable au point $b = f(a) \in V$ et $f(U) \subseteq V$. Alors :*

1. *l'application composée $g \circ f$ est différentiable au point a et*
2. $\mathrm{d}_a(g \circ f) = \mathrm{d}_b g \circ \mathrm{d}_a f \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^r)$.

Remarque 5.1.3. On peut déduire de cette proposition un calcul des dérivées partielles de $g \circ f$ au point, en remarquant que

$$J_a(g \circ f) = J_b(g) \times J_a(f).$$

Corollaire 5.1.4 (Réciproque d'une application différentiable)**.** *Soit $f\colon U \subseteq \mathbb{R}^p \to V \subseteq \mathbb{R}^q$ une application différentiable bijective dont l'application réciproque f^{-1} est également différentiable. Alors :*
1. *Les dimensions p et q sont égales.*
2. *Pour tout $a \in U$, $\mathrm{d}_{f(a)} f^{-1} = (\mathrm{d}_a f)^{-1}$.*

Démonstration. Puisque $f^{-1} \circ f = \mathrm{Id}_{\mathbb{R}^p}$ et $f \circ f^{-1} = \mathrm{Id}_{\mathbb{R}^q}$, il résulte de la Proposition 5.1.2 que

$$\mathrm{d}_{f(a)} f^{-1} \circ \mathrm{d}_a f = \mathrm{Id}_{\mathbb{R}^p} \text{ et } \mathrm{d}_a f \circ \mathrm{d}_{f(a)} f^{-1} = \mathrm{Id}_{\mathbb{R}^q}.$$

Cela signifie que l'application linéaire $\mathrm{d}_a f$ est inversible (et donc $p = q$) et d'inverse $\mathrm{d}_{f(a)} f^{-1}$. □

Remarque 5.1.5. L'enoncé suivant utilise la notion d'*ouvert connexe*, qui est introduite en topologie. Dans le cadre de cours de Calcul Différentiel, il suffit de savoir qu'un ouvert $U \subseteq \mathbb{R}^p$ est connexe si et seulement si il est *connexe par arcs* : pour tout couple de points $a, b \in U$, il existe un *chemin continu* $\gamma\colon [0,1] \subset \mathbb{R} \to U$ tel que $\gamma(0) = a$ et $\gamma(1) = b$, c'est à dire joignant a et b.

Proposition 5.1.6. *On suppose l'ouvert $U \subset \mathbb{R}^p$ **connexe** (par exemple, U est une boule ouverte). Soit $f\colon U \subset \mathbb{R}^p \to \mathbb{R}^q$ une application différentiable. Si la différentielle de f est identiquement nulle, alors f est constante sur U.*

Démonstration. Soient a et b deux points de U et $\gamma\colon [0,1] \subset \mathbb{R} \to U$ joignant a et b. On considère l'application continue $g\colon [0,1] \subset \mathbb{R} \to \mathbb{R}^q$ définie par $g(t) = f(\gamma(t))$. Il résulte de la Proposition 5.1.2 que g est dérivable sur $]0,1[$, de dérivée :

$$g'(t) = \mathrm{d}_{\gamma(t)} f(\gamma'(t)), \ t \in]0,1[.$$

D'après l'hypothèse, g' est nulle sur $]0,1[$ donc elle est constante sur $]0,1[$. Par continuité, on déduit que $g(0) = g(1)$, c'est à dire que $f(a) = f(b)$.

On a bien montré que f est constante dans U. □

5.2 Caractérisation des applications de classe $\mathcal{C}^1$.

Théorème 5.2.1. *Soit $f\colon U \subseteq \mathbb{R}^p \to \mathbb{R}^q$ une fonction continue. Alors :*
1. *Si les dérivées partielles de f existent au voisinage de a et qu'elles sont continues au point a, alors f est différentiable au point a.*
2. *L'application $f\colon U \subset \mathbb{R}^p \to \mathbb{R}^q$ est de classe $\mathcal{C}^1$ sur U si et seulement si elle admet des dérivées partielles continues en tout point de U.*

Démonstration. Grâce à l'Exercice 4.1.14, il suffit de démontrer le théorème pour une *fonction* $f\colon U \subseteq \mathbb{R}^p \to \mathbb{R}$. De plus, pour simplifier les notations, nous supposons (sans perte de généralité) que $p = 2$. Nous avons donc $f\colon (x,y) \in U \subseteq \mathbb{R}^2 \mapsto f(x,y) \in \mathbb{R}$.

1. On simplifie à nouveau les notations en supposant $a = (0,0)$ (ce qu'on peut faire quitte travailler avec la fonction $g\colon (x,y) \mapsto f(x_0 + x, y_0 + y)$, c'est à dire en faisant une translation par le vecteur (x_0, y_0) *à la* source). On a :

$$f(h) - f(a) = f(k,l) = (f(k,l) - f(k,0)) + (f(k,0) - f(0,0)) = \Delta_y + \Delta_x.$$

La preuve consiste à estimer convenablement les deux accroissements Δ_x et Δ_y. Pour cela, on considère la fonction $f_x\colon t \mapsto f(t,0)$ sur $[k,0] \subset \mathbb{R}$. Elle est continue sur $[0,k]$, et dérivable sur $]0,k[$, avec $f_x'(t) = \frac{\partial f}{\partial x}(t,0)$ en tout point de $]0,k[$. On peut lui appliquer le *Théorème des Accroissements Finis* (voir cours de L1) :il existe $t_x \in\subseteq]0,1[$ tel que :

$$f_x(1) - f_x(0) = f_x'(t_x) \text{ c'est à dire :}$$
$$f(k,0) - f(0,0) = k f_x'(t_x) \text{ et donc}$$
$$\Delta_x = f_x'(t_x) = k\frac{\partial f}{\partial x}(t_x, 0).$$

De même, il existe $t_y \in]0,1[$ tel que :

$$\Delta_y = l\frac{\partial f}{\partial y}(k, t_y).$$

C'est à ce point de la démonstration qu'on utilise la continuité des dérivées partielles $\frac{\partial f}{\partial x}$ et $\frac{\partial f}{\partial y}$ au point $(0,0)$, et le fait que (t_x, t_y) tende vers $(0,0)$ quand $h=(k,l)$ tend vers $(0,0)$. Cela donne :

$$\frac{\partial f}{\partial x}(t_x,0) = \frac{\partial f}{\partial x}(0,0) + \varepsilon_x(k,l)$$
$$\frac{\partial f}{\partial y}(k,t_y) = \frac{\partial f}{\partial y}(0,0) + \varepsilon_y(k,l)$$

où ε_x et ε_y désignent deux fonctions limite nulle à l'origine. Ainsi :

$$\begin{aligned} f(a+h) - f(a) &= k\frac{\partial f}{\partial x}(0,0) + l\frac{\partial f}{\partial y}(0,0) + k\varepsilon_x(k,l) + l\varepsilon_y(k,l) \\ &= k\frac{\partial f}{\partial x}(0,0) + l\frac{\partial f}{\partial y}(0,0) + o(h). \end{aligned}$$

En effet, $|k\varepsilon_x(k,l) + l\varepsilon_y(k,l)| \leq \|h\|_\infty (|\varepsilon_x(h)| + |\varepsilon_y(h)|) = \|h\|_\infty \varepsilon(h)$. Donc f est différentiable en $(0,0)$, de matrice jacobienne :

$$J_{(0,0)}f = \left(\frac{\partial f}{\partial x}(0,0), \frac{\partial f}{\partial y}(0,0)\right).$$

2. D'après l'hypothèse et le point 1. de l'énoncé, la fonction f est différentiable en tout point de U. De plus, l'application qui a tout $(x,y) \in U$ associe la matrice jacobienne $\left(\frac{\partial f}{\partial x}(x,y), \frac{\partial f}{\partial y}(x,y)\right)$ est continue, ce qui permet de conclure que f est continûment différentiable sur U. □

Troisième partie

Les quatre grands théorèmes

Ces quatre théorèmes développés dans cette partie sont les "énoncés royaux" du cours de Calcul Différentiel de L3. Ils recouvrent abordent les propriétés profondes des applications différentiables, et permettent en particulier d'aborder la seconde notion principale de ce cours, celle de *sous-variété différentielle*. On les nomme couramment :
1. Le théorème du point fixe (de Banach).
2. L'inégalité des accroissements finis.
3. Le théorème d'inversion locale
4. Le théorème des fonctions implicites

6 Le Théorème du Point Fixe

Il ne s'agit pas, à proprement parler, d'un théorème de Calcul Différentiel. C'est plus naturellement un énoncé de topologie. Néanmoins, en raison de son importance dans le cadre du cours de Calcul Différentiel en L3, et dans bien d'autres thèmes d'analyse (comme celui des Equations Différentielles, enseignées au second semestre de L3), nous en donnons ici non seulement l'énoncé, mais également la preuve.

Contrairement au reste de ce cours, le cadre naturel du Théorème du point fixe est celui des espaces métriques (et pas spécialement des espaces vectoriels normés).

6.1 Définitions et résultats préliminaires

Définition 6.1.1. Soit (X, d) un espace métrique. Une application $f\colon X \to X$ est appelée ***contraction*** s'il existe une constante réelle $0 \leq k < 1$ telle que

$$\forall x,y \in X, \quad \mathrm{d}(f(x), f(y)) \leq k\mathrm{d}(x,y).$$

Le nombre k est appelé ***rapport*** de la contraction f (il n'est pas unique).

Définition 6.1.2. Soient X un ensemble, $f\colon X \to X$ une application et $a \in X$. Alors l'***orbite*** du point a par l'application f est la suite de points $\{f^n(a), n \in \mathbb{N}\}$, où f^n désigne l'application $f \circ \cdots \circ f$ composée n fois.

Définition 6.1.3. Soient X un ensemble et $f\colon X \to X$ une application. Un point $a \in X$ tel que $f(a) = a$ est appelé ***point fixe*** de f.

Proposition 6.1.4. *Soient (X, d) un espace métrique, $f\colon X \to X$ une contraction de rapport k et $a, b \in X$. Alors, pour tout $n \in \mathbb{N}$, on a*

$$\mathrm{d}(f^n(a), f^n(b)) \leq k^n \mathrm{d}(a, b).$$

Démonstration. La preuve se fait par récurrence sur n. Pour $n = 0$, le résultat est évident. On suppose l'affirmation vraie pour $n \in \mathbb{N}$. Alors

$$\begin{aligned} \mathrm{d}\left(f^{n+1}(a), f^{n+1}(b)\right) &= \mathrm{d}(f(f^n(a)), f(f^n(b))) \\ &\leq k\mathrm{d}(f^n(a), f^n(b)) \\ &\leq k^{n+1}\mathrm{d}(a, b) \end{aligned}$$

grâce à l'hypothèse de récurrence. □

Proposition 6.1.5. *Une contraction f sur un espace métrique (X, d) est une application continue sur X.*

Démonstration. Soit $0 \leq k < 1$ le rapport de la contraction f. Soient $a \in X$ et $(x_n) \in X$ une suite de limite a. On a

$$\mathrm{d}(f(x_n), f(a)) \leq k\mathrm{d}(x_n, a) \underset{n\to\infty}{\longrightarrow} 0.$$

Donc $f(x_n)$ tend vers $f(a)$ et f est continue au point a. □

Remarque 6.1.6. Si f est une contraction sur un espace métrique (X, d), et si f admet un point fixe, alors ce point fixe est unique. En effet, soient $a, b \in X$ deux points fixes de f, on a

$$\mathrm{d}(a, b) = \mathrm{d}(f(a), f(b)) \leq k\mathrm{d}(a, b),$$

donc $\mathrm{d}(a, b) = 0$ et $a = b$.

On note de plus que toute application contractante est continue.

Le théorème principal de cette partie est du originellement à Banach (1922). La preuve que nous présentons est plus récente, plus courte, et due à Palais (2006). Elle repose sur l'emploi du petit résultat astucieux suivant, qui mène à une preuve "efficace" du Théorème du point fixe :

Lemme 6.1.7 (Inégalité fondamentale de contraction)**.** *Soit $f\colon (X, \mathrm{d}) \to X$ une contraction de rapport k. Alors, pour tout couple de points $x, y \in X$, on a :*

$$\mathrm{d}(x, y) \leq \frac{\mathrm{d}(f(x), x) + \mathrm{d}(f(y), y)}{1 - k}.$$

Démonstration. Pour tous $x, y \in X$, on a, en utilisant l'inégalité triangulaire :

$$\begin{aligned} \mathrm{d}(x, y) &\leq \mathrm{d}(x, f(x)) + \mathrm{d}(f(x), f(y)) + \mathrm{d}(f(y), y) \\ &\leq \mathrm{d}(x, f(x)) + k\mathrm{d}(x, y) + \mathrm{d}(f(x), y), \end{aligned}$$

d'où l'on déduit *l'inégalité fondamentale de contraction* :

$$\mathrm{d}(x, y) \leq \frac{\mathrm{d}(f(x), x) + \mathrm{d}(f(y), y)}{1 - k}.$$

□

6.2 Le Théorème du point fixe

Théorème 6.2.1 (du point fixe)**.** *Soient* (X, d) *un espace métrique* ***complet****, et* $f : X \to X$ *une contraction. Alors* f *admet un (unique) point fixe.*

Démonstration. On raisonne sur les itérés f^n de f. On considère un point quelconque $a \in X$, et on applique l'inégalité fondamentale de contraction aux points $f^n(a)$ et $f^m(a)$, pour $n, m \in \mathbb{N}$. On obtient :

$$\begin{aligned}\mathrm{d}\left(f^n(a), f^m(a)\right) &\leq \frac{\mathrm{d}\left(f^{n+1}(a), f^n(a)\right) + \mathrm{d}\left(f^{m+1}(a), f^m(a)\right)}{1-k} \\ &= \frac{\mathrm{d}\left(f^n(f(a)), f^n(a)\right) + \mathrm{d}\left(f^m(f(a)), f^m(a)\right)}{1-k} \\ &\leq \frac{k^n \mathrm{d}\left(f(a), a\right) + k^m \mathrm{d}\left(f(a), a\right)}{1-k} = \frac{(k^n + k^m)}{1-k} \mathrm{d}\left(f(a), a\right).\end{aligned}$$

Puisque $k < 1$, ceci montre que la suite $(f^n(a))$ est de Cauchy, donc convergente vers une limite $c \in X$, puisque X est complet. On voit aisément que le point c est un point fixe de f. En effet, pour tout $n \in \mathbb{N}$, $f^n(a) = f\left(f^{n-1}(a)\right)$; on obtient donc, en passant à la limite et en utilisant la continuité de f, l'égalité $c = f(c)$. □

Remarque 6.2.2. L'intérêt de la présentation ci-dessus et qu'elle permet d'introduire une *règle d'arrêt*. En effet, en fixant n et en faisant tendre m vers $+\infty$, il vient :

$$\mathrm{d}\left(f^n(a), c\right) \leq \frac{k^n}{1-k} \mathrm{d}\left(f(a), a\right).$$

Donc si on cherche à approcher le point fixe c avec une précision inférieure ou égale à $\varepsilon > 0$, il suffit de calculer $f(a)$ et l'écart $e = \mathrm{d}(f(a), a)$, et de choisir un nombre N d'itérations tel que

$$\frac{k^N}{1-k} e < \varepsilon, \text{ c'est à dire } N > \frac{\ln \varepsilon + \ln(1-k) - \ln \mathrm{d}\left(f(a), a\right)}{\ln k}.$$

6.3 Le Théorème de point fixe à paramètres.

Le propos de cette section est de considérer une famille $(f_\lambda)_{\lambda \in \Lambda}$ de contractions sur un espace métrique complet. Chacune d'elles admet un point fixe unique c_λ. On veut alors montrer que si la famille (f_λ) dépend continûment du paramètre $\lambda \in \Lambda$, alors le point fixe c_λ dépend également continûment de λ.

Pour simplifier les notations nous supposerons que X est un sous-ensemble fermé (*et donc complet*) de $\mathbb{R}^p$, que Λ est un sous-ensemble de $\mathbb{R}^m$, et nous noterons indifféremment $\|\cdot\|$ les normes sur $\mathbb{R}^p$ et sur $\mathbb{R}^m$.

Théorème 6.3.1 (point fixe à paramètres)**.** *Soient* X *un sous-ensemble fermé de* $\mathbb{R}^p$*, et* $\Lambda \subseteq \mathbb{R}^m$*. Soit* $f : X \times \Lambda \to X$ *une application continue et* ***uniformément contractante****, c'est à dire qu'il existe* $k \in]0, 1[$ *tel que pour tous* $x, y \in E$ et tout $\lambda \in \Lambda$ *:*

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

Alors :

1. *Pour tout* $\lambda \in \Lambda$*, l'application* $f(\cdot, \lambda) : x \mapsto f(x, \lambda)$ *admet un unique point fixe* $c_\lambda \in E$*.*
2. *De plus, l'application* $\lambda \mapsto c_\lambda$ *est continue sur* Λ*.*

Démonstration. **1.** L'existence et l'unicité pour tout $\lambda \in \Lambda$ du point fixe c_λ résultent du Théorème du point fixe.

2. Soient $\lambda, \lambda_0 \in \Lambda$. On a :

$$\begin{aligned}\|c_\lambda - c_{\lambda_0}\| &= \|f(c_\lambda, \lambda) - f(c_{\lambda_0}, \lambda_0)\| \\ &\leq \|f(c_\lambda, \lambda) - f(c_{\lambda_0}, \lambda)\| + \|f(c_{\lambda_0}, \lambda) - f(c_{\lambda_0}, \lambda_0)\| \\ &\leq k \|c_\lambda - c_{\lambda_0}\| + \|f(c_{\lambda_0}, \lambda) - f(c_{\lambda_0}, \lambda_0)\|\end{aligned}$$

donc

$$\|c_\lambda - c_{\lambda_0}\| \leq \frac{\|f(c_{\lambda_0}, \lambda) - f(c_{\lambda_0}, \lambda_0)\|}{1-k}.$$

Donc la continuité de $\lambda \mapsto c_\lambda$ en λ_0 découle de la continuité de f au point $(c_{\lambda_0}, \lambda_0)$. □

7 L'inégalité des accroissements finis

7.1 Le théorème des accroissements finis : fonction à valeurs dans $\mathbb{R}$

Théorème 7.1.1. *Si $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}$ est continue sur le segment $[a,b] \subset U$ est différentiable en tout point de $]a,b[$, alors il existe $c \in]a,b[$ tel que*
$$f(b) - f(a) = \mathrm{d}_c f(b-a).$$

Démonstration. On définit $x : [0,1] \subset \mathbb{R} \to [a,b]$ par $x(t) = (1-t)a + tb$. On considère $g : [0,1] \to \mathbb{R}$ définie par $g(t) = f(x(t))$. On a :
$$g'(t) = \mathrm{d}_t(f \circ x) = \mathrm{d}_{x(t)} f \circ \mathrm{d}_t x = \mathrm{d}_{x(t)} f(b-a).$$
D'après le Théorème des accroissements finis en une variable appliqué à g, il existe $t_0 \in]0,1[$ tel que
$$g'(t_0) = \frac{g(1) - g(0)}{1-0} = f(b) - f(a).$$
On en déduit l'égalité voulue, en posant $c = x(t_0)$. □

7.2 Fonction vectorielle définie sur un segment de $\mathbb{R}$

Remarque 7.2.1. Le théorème des accroissements finis n'est pas valide pour une fonction vectorielle. Par exemple, $f\colon [0,2\pi] \to \mathbb{R}^2$, $t \mapsto (\cos t, \sin t)$. Sa dérivée ne s'annulle jamais, mais $f(2\pi) - f(0) = 0$.

Cependant :

Théorème 7.2.2 (Inégalité des accroissements finis sur un intervalle de $\mathbb{R}$)**.** *Si la fonction $f\colon [a,b] \subset \mathbb{R} \to \mathbb{R}^m$ est continue sur $[a,b]$ et différentiable sur $]a,b[$, alors :*
$$\|f(b) - f(a)\| \leq \left(\sup_{c \in]a,b[} \|f'(c)\|\right) \cdot (b-a).$$

Démonstration. On introduit les fonctions $g\colon [a,b] \to \mathbb{R}$ et $h\colon [a,b] \to \mathbb{R}$ définies par :
$$\forall t \in [a,b], \quad g(t) = \|f(t) - f(a)\|, \quad h(t) = g(t) - (t-a)\frac{g(b) - g(a)}{b-a}.$$
Ces deux fonctions sont continues sur $[a,b]$. La quantité $h(t)$ est la "différence verticale" entre la valeur $g(t)$ et la *corde* qui joint les deux extrémités du graphe de g. Si h est constamment nulle tout point de $]a,b[$ est maximum ou minimum. Sinon, h atteint son maximum (ou son minimum) en un point de $]a,b[$.

Si h atteint son maximum en c, on peut donc trouver une suite $(t_n) \in]a,b[$ de limite c, avec $t_n < c$, telle que :
$$0 \leq \frac{h(t_n) - h(c)}{t_n - c} = \frac{g(t_n) - g(c)}{t_n - c} - \frac{g(b) - g(a)}{b-a}.$$
Ainsi :
$$\frac{g(b) - g(a)}{b-a} \leq \frac{g(t_n) - g(c)}{t_n - c}.$$
Si h atteint son minimum en c, on raisonne avec une suite $t_n > c$.

Ainsi :
$$\begin{aligned}
\frac{\|f(b) - f(a)\|}{b-a} = \frac{g(b) - g(a)}{b-a} &\leq \frac{g(t_n) - g(c)}{t_n - c} \\
&= \left|\frac{\|f(t_n) - f(a)\| - \|f(c) - f(a)\|}{t_n - c}\right| \\
&\leq \left\|\frac{f(t_n) - f(c)}{t_n - c}\right\| \underset{n \to +\infty}{\longrightarrow} \|f'(c)\| \leq \sup_{t \in]a,b[} \|f'(t)\|.
\end{aligned}$$
On en déduit l'inégalité voulue. □

7.3 Le cas général : fonction de $\mathbb{R}^n$ dans $\mathbb{R}^m$

Dans la suite, les espaces $\mathbb{R}^n$ et $\mathbb{R}^m$ sont munis d'une norme notée $\|\cdot\|$ et l'espace $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ est muni de la norme subordonnée : $\|L\| = \sup_{\|x\|=1} \|f(x)\|$.

Théorème 7.3.1 (Inégalité des accroissements finis)**.** *Si $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ est continue sur le segment $[a,b] \subset U$ et est différentiable en tout point de $]a,b[$, alors*

$$\|f(b) - f(a)\| \leq \left(\sup_{t\in]a,b[} \|\mathrm{d}_t f\|\right) \cdot |b-a|.$$

Démonstration. On introduit la fonction d'une variable réelle $x\colon [0,1] \to \mathbb{R}^n$ définie par $x(t) = (1-t)a + tb$, et la fonction $g = f \circ x$. Cette fonction est continue sur $[0,1]$ et dérivable en tout point de $]0,1[$, avec $g'(t) = \mathrm{d}_{x(t)} f(b-a)$.

On pose $M = \sup_{t\in]0,1[} \|\mathrm{d}_t f\|$. On suppose $M < +\infty$ (sinon le résultat est évident). On a alors

$$\|g'(t)\| \leq M\|b-a\|, \quad \forall t \in]0,1[$$

D'après l'inégalité des accroissements finis pour une fonction vectorielle d'une variable réelle, on a :

$$\|f(b) - f(a)\| = \|g(1) - g(0)\| \leq M\|b-a\|,$$

ce qui est le résultat annoncé. □

Remarque 7.3.2. Si, avec les notations précédentes, f est une fonction différentiable sur l'ouvert ***convexe*** $U \subseteq \mathbb{R}^n$ telle que $\|\mathrm{d}_x f\| \leq M$ pour tout $x \in U$, alors :

$$\|f(y) - f(x)\| \leq M\|y-x\|, \quad \forall x, y \in U.$$

8 Théorème d'inversion locale

8.1 Difféomorphismes, et difféomorphismes locaux

Définition 8.1.1. Une application $f\colon U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^n$ est appelée un ***difféomorphisme*** (de classe $\mathcal{C}^1$) de U sur V si :
1. f est une bijection de U sur V,
2. f est de classe $\mathcal{C}^1$ sur U,
3. l'application réciproque f^{-1} est de classe $\mathcal{C}^1$ sur V.

L'application f est appelée ***difféomorphisme local*** au point $a \in U$ s'il existe un voisinage ouvert $U_1 \subseteq U$ de a et un voisinage ouvert $V_1 \subseteq V$ de $f(a)$ tels que f soit un difféomorphisme de U_1 sur V_1.

Remarque 8.1.2. On voit que dans la définition de difféomorphisme, la dimension de l'espace de départ égale la dimension de l'espace d'arrivée. Ce fait est incontournable. Imaginons par exemple un difféomorphisme $f : U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^p$. Soit $a \in U$. On a $f^{-1} \circ f = \mathrm{Id}_U$. Donc on a au point :

$$\mathrm{d}_{f(a)} f^{-1} \circ \mathrm{d}_a f = \mathrm{Id}_{\mathbb{R}^n}.$$

De même, on a $f \circ f^{-1} = \mathrm{Id}_V$, donc :

$$\mathrm{d}_a f \circ \mathrm{d}_{f(a)} f^{-1} = \mathrm{Id}_{\mathbb{R}^p}.$$

Ce qui prouve que les applications linéaires $\mathrm{d}_a f$ et $\mathrm{d}_{f(a)} f^{-1}$ sont inverses l'une de l'autre. D'après le Théorème du rang, on en déduit que $n = p$.

Théorème 8.1.3 (inversion locale)**.** *Soit $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ une application de classe $\mathcal{C}^1$ et un point $a \in U$ tel que $\mathrm{d}_a f \in \mathrm{GL}_n(\mathbb{R})$. Alors f est un difféomorphisme local au point a.*

Démonstration. L'idée de la démonstration est la suivante. Etant donné un point y proche de $f(a)$, on cherche à montrer que l'équation $y = f(x)$ admet une unique solution x proche de a. On se souvient que $f(x) = f(a) + \mathrm{d}_a f(x-a) + \cdots$ Donc on peut écrire

$$y = f(x) \Longleftrightarrow x = a + (\mathrm{d}_a f)^{-1}(y - f(a)) + \cdots \Longleftrightarrow x = a - (\mathrm{d}_a f)^{-1}(f(a) - y) + \cdots$$

Ce qui donne bien une solution, proche de a. La démonstration rigoureuse repose sur le Théorème du point fixe.

Pour tout $y \in \mathbb{R}^n$, on considère l'application $\varphi_y \colon U \to \mathbb{R}^n$ définie par

$$\varphi_y(x) = x - (\mathrm{d}_a f)^{-1}(f(x) - y).$$

On note que $f(x) = y \Leftrightarrow x$ est un point fixe de φ_y. On cherche donc à montrer que φ_y est une contraction au voisinage de a. C'est une application de classe $\mathcal{C}^1$, et on a :

$$\mathrm{d}_a\varphi_y = \mathrm{Id} - (\mathrm{d}_a f)^{-1} \circ \mathrm{d}_a f = 0.$$

Puisque f est de classe $\mathcal{C}^1$, φ_y est également de classe $\mathcal{C}^1$. Donc, par continuité, il existe $r > 0$ tel que $\bar{B}(a,r) \subset U$ tel que :

$$x \in \bar{B}(a,r) \Longrightarrow \|\mathrm{d}_x\varphi_y\| \leq \frac{1}{2}.$$

Donc, d'après l'inégalité des accroissements finis, pour tout $y \in \mathbb{R}^n$ et tout $x_1, x_2 \in \bar{B}(a,r)$, on a :

$$\|\varphi_y(x_1) - \varphi_y(x_2)\| \leq \frac{1}{2}\|x_1 - x_2\|.$$

Ainsi φ_y est une contraction. Nous devons montrer qu'elle envoie $\bar{B}(a,r)$ dans elle-même, si y est assez proche de $f(a)$. Pour cela, on utilise la continuité de $(\mathrm{d}_a f)^{-1}$: il existe $\delta > 0$ tel que

$$\|y - f(a)\| < \delta \Longrightarrow \left\|(\mathrm{d}_a f)^{-1}(f(a) - y)\right\| \leq \frac{r}{2}.$$

Donc pour $x \in \bar{B}(a,r)$ et $y \in B(f(a), \delta)$ on a :

$$\|\varphi_y(x) - a\| \leq \|\varphi_y(x) - \varphi_y(a)\| + \|\varphi_y(a) - a\| \leq r.$$

On peut alors appliquer le théorème du point fixe à φ_y dans la boule $\bar{B}(a,r)$: pour tout $y \in B(f(a), \delta)$ il existe un unique $x \in \bar{B}(a,r)$ tel que $\varphi_y(x) = x$, c'est à dire tel que $f(x) = y$.

Notons $g(y)$ ce point fixe : g est un bijection de $B(f(a), \delta)$ sur son image W. En utilisant la version à paramètre du théorème du point fixe et le fait que $(x,y) \mapsto \varphi_y(x)$ est continue, de rapport de contraction uniforme $1/2$, on en déduit que g est continue.

Pour montrer que g est de classe $\mathcal{C}^1$, il faut d'abord montrer qu'elle est différentiable. Prouvons le au point $f(a)$: en tout autre point de $B(f(a), \delta)$ les hypothèses du théorème restent vérifiées, donc la preuve sera la même. On a

$$g(f(a) + h) = a + \varepsilon(h)$$

grâce à la continuité de g. Puisque $f(g(f(a)+h)) = f(a) + h$ d'une part, et

$$f(g(f(a)+h)) = f(a + \varepsilon(h)) = f(a) + \mathrm{d}_a f(\varepsilon(h)) + o(\varepsilon(h)),$$

il vient $h = \mathrm{d}_a f(\varepsilon(h)) + o(\varepsilon(h))$, et donc $\varepsilon(h) = (\mathrm{d}_a f)^{-1}(h - o(h)) = (\mathrm{d}_a f)^{-1}(h) + o(h)$, on a :

$$g(f(a) + h) = a + (\mathrm{d}_a f)^{-1}(h) + o(h),$$

ce qui montre le résultat.

Enfin, les coefficients de la matrice de $(\mathrm{d}_a f)^{-1}$ dans la base canonique sont continus, puisque obtenus à l'aide de fractions rationnelles à partir des coefficients de la matrice de $\mathrm{d}_a f$. Ainsi g est de classe $\mathcal{C}^1$. $\square$

8.2 De l'inversion locale aux difféomorphismes

Théorème 8.2.1. *Soit $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ une application de classe $\mathcal{C}^1$ telle que :*
i) $f\colon U \to f(U)$ *est une bijection,*
ii) f *est un difféomorphisme local en chacun de ses points.*
Alors :
1. $f(U)$ *est ouvert,*
2. f *est un difféomorphisme de U sur $f(U)$.*

Démonstration. Pour tout $a \in U$, f admet un inverse local au voisinage de a. Donc il existe un voisinage $V \subseteq U$ de $f(a)$ tel que tout élément y de ce voisinage admet un antécédant. Donc $V \subseteq f(U)$: $f(U)$ étant un voisinage de chacun de ses points, il est donc ouvert. L'application f^{-1} est bien définie sur $f(U)$ car f est une bijection, et elle est de classe $\mathcal{C}^1$ en tout point de $f(U)$ toujours d'après le théorème d'inversion local.

On en déduit que f est un difféomorphisme de U sur $f(U)$. □

9 Le Théorème des fonctions implicites

9.1 La résolution d'un système d'équations

C'est une question extrêmement naturelle : que signifie précisément *résoudre un système d'équations*

$$\begin{aligned} f_1(u) &= 0 \\ f_2(u) &= 0 \\ &\vdots \\ f_m(u) &= 0 \end{aligned}$$

où les inconnues u appartiennent à un certain espace $\mathbb{R}^p$. La réponse est apparemment simple :résoudre le système signifie décrire l'ensemble de ses solutions ! Mais la question se repose à nouveau : que signifie exactement *décrire* cet ensemble de solutions ?

La réponse est donnée précisément dans le cadre des systèmes d'équations linéaires. Donnons deux exemples élémentaires :

Exemple 9.1.1. Il s'agit de résoudre le système :

$$\begin{aligned} f_1(x,y,z) &= x + 2y + 2z = 1 \\ f_2(x,y,z) &= x + y + z = 0. \end{aligned}$$

Ici on a $u = (x,y,z) \in \mathbb{R}^3$. La matrice associée à ce système est $A = \begin{pmatrix} 1 & 2 & 2 \\ 1 & 1 & 1 \end{pmatrix}$. C'est une matrice de *rang* 2 : le plus grand mineur non nul que l'on peut extraire de cette matrice est de taille 2×2. Par exemple, le mineur correspondant aux deux premières colonnes de A, $\begin{vmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{vmatrix}(x,y,z) = \begin{vmatrix} 1 & 2 \\ 1 & 1 \end{vmatrix} = -1$. Ce choix de mineur dicte une organisation de variables u en deux sous-familles de variables :
1. *les variables principales* : elles correspondent aux deux premières colonnes du système ;les variables x et y.
2. *les variables paramètres* : ce sont toutes les autres variables. Dans l'exemple, il s'agit de la variable z.

Ce simple choix donne une information importante :si l'ensemble $\mathcal{S}$ des solutions est non vide (ce qui correspondrait à un système impossible), alors c'est un *sous-espace affine* de $\mathbb{R}^3$ dont la dimension est égale au nombre de *paramètres*, en l'occurence 1.

Résoudre le système signifie exprimer les variables principales en fonction des variables paramètres.

Dans l'exemple, c'est très facile. On passe la variable paramètre z du côté droit de l'équation :

$$\begin{aligned} x + 2y &= 1 - 2z \\ x + y &= -z \end{aligned}$$

et on résout "comme on le peut" le système obtenue en les inconnues principales. Ici, si on suit la méthode "en cascade", on résout la première équation par rapport à l'une des inconnues principales (par exemple y), on reporte la solution dans la deuxième équation, et on en déduit x (*détail des calculs laissés en exercice*). On obtient :

$$x = -1, \; y = 1 - z,$$

ce qu'on écrit :

$$\mathcal{S} = \left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} : z \in \mathbb{R} \right\}.$$

L'ensemble $\mathcal{S}$ des solutions apparaît donc comme le *graphe* $\Gamma_\varphi \subset \mathbb{R}^3$ de la fonction affine $\varphi\colon \mathbb{R} \to \mathbb{R}^2$, $z \mapsto \varphi(z) = (-1, 1 - z, z)$.

Remarque 9.1.2 (interprétation géométrique). L'ensemble $\mathcal{S}$ de l'exemple ci-dessus est l'ensemble $F^{-1}(0)$, où F est l'application affine définie par $F(x, y, z) = (x + 2y + 2z - 1, x + y + z)$. A cette application est associée une matrice $A \in \mathcal{M}_{2,3}(\mathbb{R})$ de rang 2. On sait alors *a priori* que $\mathcal{S}$ est soit l'ensemble vide, soit un *sous-espace affine* de $\mathbb{R}^3$ de dimension $3 - 2 = 1$ (si $\mathcal{S} \neq \emptyset$, $\dim(\mathcal{S})$=dimension de l'espace de départ − rang de la matrice associée à F), c'est à dire une droite affine : c'est en l'occurence la droite affine passant par le point $(-1, 1, 0)$ et de vecteur directeur $\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$.

On peut également voir $\mathcal{S}$ comme l'intersections de deux sous-espaces affines $f_1^{-1}(0) \cap f_2^{-1}(0)$, où $f_1(x, y, z) = x + 2y + 2z - 1$ et $f_2(x, y, z) = x + y + z$. Il s'agit de deux plans affines de $\mathbb{R}^3$, non parallèles, à cause de l'existence d'un mineur 2×2 non nul. Leur intersection est donc une droite affine. De façon générale, un espace affine donné par une seule équation s'appelle un *hyperplan*. Un sous-espace affine (général) est une intersection d'hyperplans.

Exemple 9.1.3. Il s'agit de résoudre l'équation $f(x, y, z) = x - 2y + z = 3$. La matrice associée à cette équation est $A = (1, -2, 1)$. Elle est de rang 1, car la taille maximale d'un mineur non nul est 1×1. Choisissons par exemple le mineur $\frac{\partial f}{\partial z}(x, y, z) = 1$. Il y a donc un unique *variable principale* z, et deux *variables paramètres* x et y. L'ensemble $\mathcal{S}$ des solutions est donc un espace affine de dimension 2. On résout l'équation en exprimant z en fonction de x et y :

$$z = 3 - x + 2y.$$

L'ensemble $\mathcal{S}$ des solutions de l'équation est donc le graphe $\Gamma_\varphi \subset \mathbb{R}^3$ de la fonction $\varphi\colon \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto \varphi(x, y) = 3 - x + 2y$. On a :

$$\mathcal{S} = \left\{ \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} + x \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \right\}.$$

C'est le plan affine passant par le point $(0, 0, 3)$ et de vecteurs directeurs $\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ et $\begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$.

On peut résumer la discussion précédente en un principe : *résoudre un système d'équations signifie représenter l'ensemble des solutions de ce système comme le graphe d'une fonction.* Comme on va le voir, ce principe valide en algèbre linéaire reste vrai en Calcul Différentiel. A une différence notable près : alors que les ensembles solutions des systèmes d'équation linéaires sont des ensembles "rectilignes", décrits comme des graphes de fonctions affines définies sur $\mathbb{R}^k$ tout entier, les ensembles solutions d'équations données par des applications différentiables sont "courbés", et vont apparaître comme des graphes de fonctions différentiables définies localement, c'est à dire au voisinage d'un point.

9.2 De l'inversion locale aux fonctions implicites

Dans l'énoncé suivant, on pose $p = m + q$, et on note $(x, y) = (x_1, \ldots, x_m, y_1, \ldots, y_q) \in \mathbb{R}^m \times \mathbb{R}^q$. Cela représente le découpage en *variables paramètres* $(x_1, \ldots, x_m)$ et variables principales $(y_1, \ldots, y_q)$.

Théorème 9.2.1. *Soit $f\colon U \subseteq \mathbb{R}^{m+q} \to \mathbb{R}^q$ une application de classe $\mathcal{C}^1$. On considère un point $(a,b) \in U$ tel que $f(a,b) = 0$. Supposons que la matrice*

$$\begin{pmatrix} \dfrac{\partial f_1}{\partial y_1}(a,b) & \cdots & \dfrac{\partial f_1}{\partial y_q}(a,b) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_q}{\partial y_1}(a,b) & \cdots & \dfrac{\partial f_q}{\partial y_q}(a,b) \end{pmatrix}$$

soit inversible. Alors il existe un voisinage V de a dans $\mathbb{R}^m$, un voisinage W de b dans $\mathbb{R}^q$ et une application $\varphi\colon V \to W$ tels que, pour tout $(x,y) \in V \times W$, on a :

$$f(x,y) = 0 \Longleftrightarrow y = \varphi(x).$$

Remarque 9.2.2. L'hypothèse sur f dit que $\mathrm{d}_{(a,b)}f \in \mathcal{L}(\mathbb{R}^m \times \mathbb{R}^q, \mathbb{R}^q)$ est surjective (en effet sa matrice jacobienne contient un mineur non nul de taille q).

Démonstration. Ce théorème se démontre en appliquant le théorème d'inversion locale à l'application $g\colon \mathbb{R}^m \times \mathbb{R}^q \to \mathbb{R}^m \times \mathbb{R}^q$ définie par $g(x,y) = (x, f(x,y))$. En particulier, on a $g(a,b) = (a,0) \in \mathbb{R}^m \times \mathbb{R}^q$. La matrice jacobienne de g en un point $(x,y) \in \mathbb{R}^m \times \mathbb{R}^q$ est

$$J_{(x,y)}g = \begin{pmatrix} \mathrm{Id}_m & 0 \\ B(x,y) & A(x,y) \end{pmatrix},$$

où

$$A(x,y) = \begin{pmatrix} \dfrac{\partial f_1}{\partial y_1}(x,y) & \cdots & \dfrac{\partial f_1}{\partial y_q}(x,y) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_q}{\partial y_1}(x,y) & \cdots & \dfrac{\partial f_q}{\partial y_q}(x,y) \end{pmatrix} = \frac{\partial f}{\partial y}(x,y),\ B(x,y) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(x,y) & \cdots & \dfrac{\partial f_1}{\partial x_m}(x,y) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_q}{\partial x_1}(x,y) & \cdots & \dfrac{\partial f_q}{\partial x_m}(x,y) \end{pmatrix} = \frac{\partial f}{\partial x}(x,y).$$

Attention, les notations $\dfrac{\partial f}{\partial x}(x,y)$ et $\dfrac{\partial f}{\partial y}(x,y)$ sont des notations raccourcies pour les matrices jacobiennes $B(x,y)$ et $A(x,y)$.

On voit que toutes les dérivées partielles de g sont continues, et que le déterminant de $J_{(x,y)}g$ égale le déterminant de $B(x,y)$, qui est supposé non nul au point (a,b). D'après le Théorème d'inversion locale, il existe un voisinage $U' \subseteq U$ de (a,b) tel que $g\colon U' \to g(U')$ soit un difféomorphisme de classe $\mathcal{C}^1$. Quitte à réduire U', on peut supposer qu'il est de la forme $V \times W$ où V est un voisinage de a dans $\mathbb{R}^m$ et W un voisinage de b dans $\mathbb{R}^q$. Par conséquent $g(U')$ est de la forme $V_1 \times W_1$ ou V_1 est un voisinage ouvert de a et W_1 est un voisinage ouvert de $f(a,b) = 0$. L'application g admet un inverse de classe $\mathcal{C}^1$. Il est de la forme :

$$g^{-1}\colon\ (x,y) \longmapsto (x, h(x,y))$$

avec $h\colon V \times W_1 \to \mathbb{R}^m \times \mathbb{R}^q$ une fonction de classe $\mathcal{C}^1$. On note que :

$$f(x,y) = 0 \Longleftrightarrow g(x,y) = (x,0) \Longleftrightarrow (x,y) = g^{-1}(x,0) \Longleftrightarrow y = h(x,0).$$

Donc l'ensemble des points (x,y) tels que $f(x,y) = 0$ est l'ensemble $g^{-1}(x,0)$ avec $x \in V$, c'est à dire l'ensemble de points $(x, h(x,0))$. La fonction φ demandée est donc $\varphi\colon x \mapsto h(x,0)$. C'est bien une application de classe $\mathcal{C}^1$. □

Remarque 9.2.3. On pourrait également démontrer le théorème d'inversion locale à partir du théorème des fonctions implicites.

Proposition 9.2.4. *Avec les notations du théorème précédent (et de sa preuve), la matrice jacobienne de φ en tout point $x \in V$ est donnée par :*

$$J_x\varphi = -\left(\frac{\partial f}{\partial y}(x,\varphi(x))\right)^{-1} \times \left(\frac{\partial f}{\partial y}(x,\varphi(x))\right).$$

Démonstration. On rappelle que, pour tout $x \in V$, $\varphi(x)$ est la deuxième composante de $g^{-1}(x,0) = (x,\varphi(x))$. On calcule dans un premier temps la différentielle de la composée : on a $\varphi = g^{-1} \circ L$, où $L\colon \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^q$ est l'application linéaire $x \mapsto (x,0)$. En particulier, $\mathrm{d}_x L = L$ pour tout $x \in \mathbb{R}^m$. On a donc, pour $u \in \mathbb{R}^m$:

$$\begin{aligned}\mathrm{d}_x\varphi(u) &= \mathrm{d}_x\left(g^{-1} \circ L\right)(u)\\ &\mathrm{d}_{(x,0)}\left(g^{-1}\right) \circ \left(\mathrm{d}_{g^{-1}(x)} L\right)(u)\\ &= \mathrm{d}_{(x,0)}\left(g^{-1}\right)(L(u))\\ &= \left(\mathrm{d}_{g^{-1}(x,0)} g\right)^{-1}(u,0)\\ &= \left(\mathrm{d}_{(x,\varphi(x))} g\right)^{-1}(u,0).\end{aligned}$$

Or, pour $c \in \mathbb{R}^m \times \mathbb{R}^q$:

$$\mathrm{d}_c g(u,v) = (w,0) \iff \begin{pmatrix} \mathrm{Id}_m & 0 \\ \dfrac{\partial f}{\partial x}(c) & \dfrac{\partial f}{\partial y}(c) \end{pmatrix} \cdot \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} w \\ 0 \end{pmatrix} \iff u = w \text{ et } \frac{\partial f}{\partial x}(c)(u) + \frac{\partial f}{\partial y}(c)(v) = 0,$$

donc on trouve $v = -\left(\dfrac{\partial f}{\partial y}(c)\right)^{-1} \cdot \dfrac{\partial f}{\partial x}(c)(w)$. On conclut en appliquant ce résultat au point $c = (x,\varphi(x))$. □

Quatrième partie

Sous-variétés de $\mathbb{R}^n$

Les ensembles décrits par l'algèbre linéaire - c'est à dire à l'aide des équations linéaires ou affines - sont les ***sous-espaces vectoriels ou affines*** des espaces $\mathbb{R}^n$. On peut tous les présenter comme des intersections finies d'***hyperplans*** vectoriels ou affines.

Le propos de ce cours est de résumer les propriétés des les ensembles décrits par la ***géométrie différentielle***, c'est à dire décrits à l'aide d'équations données par des fonctions différentiables. Les deux résultats principaux des chapitres précédents - le Théorème des fonctions implicites et le Théorème d'inversion locale - sont les outils essentiels de cette étude.

10 Deux types d'exemples

10.1 Le cas linéaire

Les exemples traités dans la Section 9.1 sont des cas particuliers d'un fait général. Tous les sous-espaces affines (non vides) de $\mathbb{R}^n$, $n \in \mathbb{N}$, sont le graphe d'une fonction affine définie sur un certain espace $\mathbb{R}^k$. Ce sont alors des espaces affines de dimension k. Ils constituent l'exemple le plus simple de *sous-variétés de dimension* k.

10.2 La courbe de Viviani

On considère dans $\mathbb{R}^3$ l'intersection V de la sphère unité $S = \left\{(x,y,z) \in \mathbb{R}^3 \colon x^2+y^2+z^2 = 1\right\}$ et du cylindre vertical C au-dessus du cercle de cenre $\left(\frac{1}{2},0,0\right)$ et de rayon $\frac{1}{2}$. L'équation de C est donc : $\left(x-\frac{1}{2}\right)^2 + y^2 = \frac{1}{4}$, c'est à dire $x^2 - x + y^2 = 0$. L'ensemble V est donc défini par

$$x^2+y^2+z^2 = 1, \quad x^2 - x + y^2 = 0.$$

Afin de simplifier ce système, on remplace x^2+y^2 par x dans la première équation, et considérer que C est défini par $\left\{x+z^2-1 = 0, x^2-x+y^2 = 0\right\}$. On introduit $f\colon \mathbb{R}^3 \to \mathbb{R}^2$ défine par

$$f(x,y,z) = \left(x+z^2-1, x^2-x+y^2\right),$$

et donc $V = f^{-1}(0)$.

Etudions en quels points de V on peut appliquer le Théorème des fonctions implicites à l'application f.

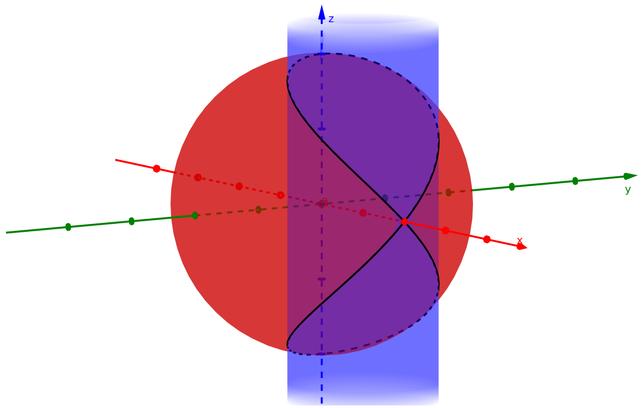

FIGURE 10.2.1 – Courbe de Viviani

La matrice jacobienne de f au point $a = (x, y, z)$ est

$$J_a f = \begin{pmatrix} 1 & 0 & 2z \\ 2x - 1 & 2y & 0 \end{pmatrix}.$$

Les trois mineurs 2×2 de cette matrice sont, au signe près, $2y$, $2z\,(2x - 1)$ et $4yz$. S'ils s'annulent tous les trois, alors $y = 0$. D'après les équations de V, les seuls points de V en lequel $y = 0$ sont $(1, 0, 0)$ et $(0, 0, \pm 1)$. Or, aux points $(0, 0, \pm 1)$ le deuxième mineur ne s'annule pas.

En conclusion, au voisinage de tout point $a \in V$ différent de $(1, 0, 0)$, le Théorème des fonctions implicites affirme qu'on peut représenter V comme le graphe d'une fonction différentiable d'une variable. Nous dirons que $V \setminus \{(1, 0, 0)\}$ est une *sous-variété différentiable de dimension* 1 de $\mathbb{R}^3$, (plus simplement, une *courbe* de $\mathbb{R}^3$).

Exercice 10.2.1. Montrer qu'on paramétrer cette courbe par

$$x\,(t) = \cos^2 t, \quad y\,(t) = \cos\,(t) \sin\,(t)\,, \quad z\,(t) = \sin t, \quad t \in \left]-\pi, \pi\right[.$$

On note que deux branches se coupent au point (1,0,0), qui correspondent à $t = 0$ et $t \to \pm\pi$.

11 Sous-variétés définies par des équations

11.1 Sous-variétés, coordonnées rectifiantes et paramétrages

Définition 11.1.1. Un ensemble $M \subseteq \mathbb{R}^n$ est ***une sous-variété de dimension*** p au point $a \in \mathcal{M}$ s'il existe un voisinage ouvert U de a dans $\mathbb{R}^n$ et une application $f = (f_1, \ldots, f_{n-p}) : U \to \mathbb{R}^{n-p}$ de classe $\mathcal{C}^1$ telle que :
1. $\mathrm{d}_a f \in \mathcal{L}\,(\mathbb{R}^n, \mathbb{R}^{n-p})$ est de rang $n - p$ (c'est à dire $\mathrm{d}_a f$ est *surjective*) ;
2. $M \cap U = f^{-1}\,(0)$.

On dit que les composantes $f_1, \ldots, f_{n-p}$ de f sont ***indépendantes*** au point a. On dit que M est une sous-variété de si c'est une sous-variété en chacun de ses points.

Exemple 11.1.2. Soit $f \colon U \subseteq \mathbb{R}^p \to \mathbb{R}^m$ de classe $\mathcal{C}^1$. Alors le ***graphe*** $\Gamma_f = \{(x, f\,(x)) : x \in U\} \subset \mathbb{R}^p \times \mathbb{R}^m$ ***de*** f est une sous-variété de dimension p de $\mathbb{R}^{p+m}$.

Proposition 11.1.3. *Un sous-ensemble $M \subseteq \mathbb{R}^n$ est une sous-variété de dimension p de $\mathbb{R}^n$ si et seulement si, pour tout $a \in M$ il existe un difféomorphisme φ entre un voisinage U de a dans $\mathbb{R}^n$ et un voisinage V de $0 \in \mathbb{R}^n$ tel que :*
1. $\varphi\,(a) = 0$;
2. $\varphi\,(M \cap U) = \{(y_1, \ldots, y_n) \in V : y_{p+1} = \cdots = y_n = 0\}$.

*Dans ce cas φ est appelée **coordonnée rectifiante** (ou **coordonnée**, ou **carte**) de M au point a.*

Démonstration. Soit $a = (a_1, \ldots, a_n) \in M$. On suppose qu'une telle coordonnée φ existe. On a alors $M \cap U = f^{-1}(0)$ avec $f = (\varphi_{p+1}, \ldots, \varphi_n) : U \to \mathbb{R}^{n-p}$. De plus, $\mathrm{d}_a f \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^{n-p})$ est bien surjective car $\mathrm{d}_a \varphi$ l'est.

Réciproquement, supposons que M est une sous-variété de dimension p au point $a = (a_1, \ldots, a_n)$. Il existe alors une fonction $f = (f_1, \ldots, f_{n-p}) : U \subseteq \mathbb{R}^n \to \mathbb{R}^{n-p}$ de classe $\mathcal{C}^1$ définie sur un voisinage U de a telle que $U \cap M = f^{-1}(0)$, et telle que la différentielle $\mathrm{d}_a f$ soit surjective. Donc de la matrice jacobienne $J_a f \in \mathcal{M}(n-p, n)$ on peut extraire une matrice inversible de taille $n-p$. Quitte a munir $\mathbb{R}^n$ d'un système de coordonnées $(x_1, \ldots, x_p, y_{p+1}, \ldots, y_n)$ obtenu par une permutation convenable des coordonnées , on peut supposer que cette matrice inversible est obtenue en prenant les $n-p$ lignes et et les $n-p$ dernières colonnes de $J_a f$, c'est à dire :

$$\frac{\partial f}{\partial y}(a) = \begin{pmatrix} \dfrac{\partial f_1}{\partial y_{p+1}}(a) & \cdots & \dfrac{\partial f_1}{\partial y_n}(a) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_{n-p}}{\partial y_n}(a) & \cdots & \dfrac{\partial f_{n-p}}{\partial y_n}(a) \end{pmatrix}.$$

L'application $g \colon (x, y) \mapsto (x, f(x, y))$ introduite dans la preuve du Théorème des Fonctions Implicites est un difféomorphisme local au point $a = (a_1, \ldots, a_p, a_{p+1}, \ldots, a_n)$. C'est une coordonnée rectifiante, car $g(M \cap U) = \{(y_1, \ldots, y_p, 0, \ldots, 0) \in g(U)\}$. $\square$

Exemple 11.1.4. On considère le cercle unité $\mathbb{S}^1 = \left\{(x, y) \in \mathbb{R}^2 \colon x^2 + y^2 - 1\right\}$, qui est une sous-variété de dimension 1 de $\mathbb{R}^2$. On se place au point $a = (0, 1) \in \mathbb{S}^1$. ON pose $f \colon (x, y) \mapsto x^2 + y^2 - 1$. On a $\frac{\partial f}{\partial x}(a) = 0$ et $\frac{\partial f}{\partial y}(a) = 2 \neq 0$. On se place donc sur un ouvert $U \subseteq \mathbb{R}^2$ qui contient le point a sur lequel $\frac{\partial f}{\partial y} \neq 0$. Par exemple $U = \left\{(x, y) \in \mathbb{R}^2 \colon y > 0\right\}$. On *complète* f en un système de coordonnées (difféomorphisme) sur U en posant $g(x, y) = (x, f(x, y))$. Pour être précis, on détermine $g(U) \subseteq \mathbb{R}^2$. On cherche donc les éléments $(X, Y) \in \mathbb{R}^2$ pour lesquels le système

$$\begin{aligned} x &= X \\ x^2 + y^2 - 1 &= Y \end{aligned}$$

admet une solution (forcément unique) $(x, y) \in U$. On a bien sûr $x = X$. Puis on a l'équation de degré 2 en y : $y^2 - \left(Y - \left(X^2 - 1\right)\right) = 0$. Cette équation n'a de solutions dans U que si $Y > X^2 - 1$. On pose donc $V = \left\{(X, Y) \in \mathbb{R}^2 \colon Y > X^2 - 1\right\}$. L'image d'une verticale $\{x = x_0\}$ est la verticale $\{X = x_0\}$ et l'image d'une horizontale $\{y = y_0\}$, paramétrée par $\left\{(t, y_0) \in \mathbb{R}^2, t \in \mathbb{R}\right\}$, est la courbe paramétrée $\left\{\left(t, t^2 + y_0 - 1\right) : t \in \mathbb{R}\right\}$, c'est à dire la parabole d'équation $Y = X^2 - 1 + y_0$.

Le résultat précédent permet de donner une deuxième définition, équivalente, de sous-variété :

Proposition 11.1.5. *Un sous-ensemble M de $\mathbb{R}^n$ est une sous-variété de dimension p de $\mathbb{R}^n$ si et seulement si, pour tout $a \in M$, il existe un voisinage $\mathcal{U}$ de a dans $\mathbb{R}^n$, un voisinage W de $0 \in \mathbb{R}^p$ et une application $\gamma \colon W \to \mathcal{U} \cap M$ de classe $\mathcal{C}^1$ telle que :*
1. $\gamma(0) = a$;
2. γ *soit une bijection de W sur $U \cap M$;*
3. $\mathrm{d}_y \gamma$ *soit injective en tout point $y \in W$.*

*L'application γ est appelée **paramétrisation locale** de M au voisinage de $a \in M$.*

Remarque 11.1.6. La propriété 3. équivaut à dire que la matrice jacobienne $J_y \gamma$ est de rang p, c'est à dire par continuité, que $J_0 \gamma$ est de rang p.

Démonstration. a) si M est une sous-variété de dimension p de $\mathbb{R}^n$, et $a \in M$, alors M admet une coordonnée rectifiante au voisinage de a : il existe un voisinage U de a dans $\mathbb{R}^n$, un voisinage V de $0 \in \mathbb{R}^p \times \mathbb{R}^{n-p}$ et un difféomorphisme $\varphi \colon U \to V$ tel que $\varphi(U \cap M) = V \cap \left(\mathbb{R}^p \times \{0\}^{n-p}\right)$. On note $\psi = \varphi^{-1}$, et on pose $\gamma(x_1, \ldots, x_p) = \psi(x_1, \ldots, x_p, 0)$. La matrice $J_0 \psi$ est inversible, donc la matrice $\left(\dfrac{\partial \psi_j}{\partial x_i}(0)\right)_{\substack{i = 1, \ldots, p \\ j = 1, \ldots, n}}$, constituée des p premières colonnes de $J_0 \psi$ est faite de p vecteurs linéairement indépendants, donc est de rang p. Or cette matrice est précisément la matrice jacobienne $J_0 \gamma$. Par continuité cette propriété reste vraie pour y au voisinage de 0.

b) Quitte à translater, on peut supposer que $a = 0$. On a $\gamma \colon u = (u_1, \ldots, u_p) \in \mathbb{R}^p \mapsto \gamma(u) = (\gamma_1(u), \ldots, \gamma_n(u)) \in \mathbb{R}^n$. Grâce à l'hypothèse, on peut suppose que les p premières colonnes de J_0 sont indépendantes, c'est à dire que l'application $g\colon u \mapsto (\gamma_1(u), \ldots, \gamma_p(u))$ est un difféomorphisme local à l'origine. On note h l'inverse local de g. Ainsi, quitte à restreindre les ouverts de définition, les points de $M \cap U$ sont caractérisés par les $n-p$ équations

$$x_{p+1} - \gamma_{p+1}(g(x_1, \ldots, x_p)) = 0, \ldots, x_n = \gamma_n(g(x_1, \ldots, x_p)) = 0,$$

qui forment bien un système de rang $n-p$. ☐

Définition 11.1.7 (Application différentiable entre sous-variétés)**.** Soit $M \subseteq \mathbb{R}^m$ une sous-variété de dimension p et soit $N \subseteq \mathbb{R}^n$ une sous-variété de dimension q. Une application $f : M \to N$ est dite de ***classe*** $\mathcal{C}^1$ si pour tout $a \in M$, tout paramétrage local γ de M en a défini sur un voisinage U de 0 dans $\mathbb{R}^p$ et tout paramétrage local $\tilde{\gamma}$ de N en $f(a)$ défini sur un voisinage local V de 0 dans $\mathbb{R}^q$ tels que $f(\gamma(U)) \subseteq \tilde{\gamma}(V)$, l'application :

$$\tilde{\gamma}^{-1} \circ f \circ \gamma\colon U \to V$$

est de classe $\mathcal{C}^1$.

Remarque 11.1.8. Cette définition ne dépend pas du choix des paramétrages γ et $\tilde{\gamma}$.

11.2 Espace tangent à une sous-variété

Proposition 11.2.1. *Soient $M \subseteq \mathbb{R}^n$ une sous-variété de dimension p et $a \in M$. On suppose que sur un voisinage U de a dans $\mathbb{R}^n$, M est définie par l'équation $f(x) = 0$ où $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^{n-p}$ est de classe $\mathcal{C}^1$ et de rang $n-p$ sur U. On considère également un paramétrage local $\gamma\colon V \subseteq \mathbb{R}^p \to U \cap M \subseteq \mathbb{R}^n$ de M au point a, avec $\gamma(0) = a$. Alors :*

$$\ker \mathrm{d}_a f = \mathrm{Im} \mathrm{d}_0 \gamma.$$

Démonstration. On remarque que l'application $f \circ \gamma$ est identiquement nulle au voisinage de 0 dans V, donc $\mathrm{d}_a f \circ \mathrm{d}_0 \gamma = 0$. On en déduit l'inclusion $\mathrm{Im}\mathrm{d}_0\gamma \subseteq \ker \mathrm{d}_a f$. Puisque $\mathrm{d}_0\gamma$ est injective, on a $\dim \mathrm{Im}\mathrm{d}_0\gamma = p$. D'autre part, puisque $\mathrm{d}_a f \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^{n-p})$ est de rang $n-p$, on déduit du théorème du rang que $\dim \ker \mathrm{d}_0 f = p$. Ces deux espaces sont donc bien égaux. ☐

Définition 11.2.2 (Espace tangent)**.** Avec les notations de la définition précédente, l'***espace tangent à*** M ***au point*** a est le sous-espace affine de dimension p :

$$T_a M = a + \ker \mathrm{d}_a f = a + \mathrm{Im}\mathrm{d}_0\gamma.$$

Remarque 11.2.3. En particulier, si M est de codimension 1 et si $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ fournit l'équation locale $f(x) = 0$ de M au point a, on note que, pour tout $h \in \mathbb{R}^n$:

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

On en déduit que $T_a M$ est l'*orthogonal* de $\mathrm{grad} f(a)$ au point a.

Cinquième partie

Différentielles d'ordre supérieur

La recherche des maxima et minima des fonctions, qui se pose déjà pour les fonctions d'une variable réelle, se pose également pour les fonctions différentiables de plusieurs variables. Nous traitons dans ce chapitre deux problèmes principaux :

1. La recherche des extrema d'une fonction différentiable sur un ouvert de $\mathbb{R}^n$;
2. La recherche des extrema d'une fonction différentielle restreinte à une sous-variété lisse de $\mathbb{R}^n$.

12 Différentielles et dérivées d'ordre supérieur - Formules de Taylor

12.1 Différentielles d'ordre supérieur

Définition 12.1.1. Soit $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}$.
1. La fonction f est deux fois différentiable en un point $a \in U$ si f est différentiable au voisinage de a et si l'application $\mathrm{d}f$ est différentiable au point a. On note $\mathrm{d}_a^2 f = \mathrm{d}_a(\mathrm{d}f)$ la ***différentielle seconde*** de f au point a.
2. On dit que la fonction f est de classe $\mathcal{C}^2$ sur U si l'application $\mathrm{d}f$ est définie et est de classe $\mathcal{C}^1$ sur U.

Remarque 12.1.2.
1. Il faut bien comprendre les domaines et espaces d'arrivée des différentielles secondes. L'application $\mathrm{d}f$ est définie sur un voisinage V de a contenu dans $U \subseteq \mathbb{R}^n$, est est à valeurs dans $\mathcal{L}(E, \mathbb{R})$. Donc $\mathrm{d}_a^2 f = \mathrm{d}_a(\mathrm{d}f) \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}))$. Donc $\mathrm{d}_a^2 f$ s'applique linéairement à des éléments $u \in \mathbb{R}^n$, et $\mathrm{d}_a^2 f(u)$ s'applique linéairement à des éléments $v \in \mathbb{R}^n$, en étant à valeurs réelles.

$$\begin{array}{ccccc} \mathbb{R}^n & \xrightarrow{\mathrm{d}_a^2 f} & \mathcal{L}(\mathbb{R}^n, \mathbb{R}) & \longrightarrow & \mathbb{R} \\ u & \longmapsto & \mathrm{d}_a^2 f(u) : v & \longmapsto & \mathrm{d}_a^2 f(u)(v) = (\mathrm{d}_a(\mathrm{d}f)(u))(v) \end{array}$$

On peut donc identifier $\mathrm{d}_a^2 f$ à une *forme bilinéaire* en les deux arguments u et v. On notera donc désormais $\mathrm{d}_a^2 f(a) \in \mathcal{L}_2(\mathbb{R}^n, \mathbb{R})$ (l'espace des formes bilinéaires sur $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$), et $\mathrm{d}_a^2 f(u, v)$ au lieu de $\mathrm{d}_a^2 f(u)(v)$.
2. Il s'avère que la forme bilinéaire $\mathrm{d}_a^2 f$ est *symétrique*. On peut donc lui associer la *forme quadratique* $h \in \mathbb{R}^n \mapsto \mathrm{d}_a^2 f(h, h)$. Pour simplifier, nous noterons $\mathrm{d}_a^2 f(h)$ plutôt que $\mathrm{d}_a^2 f(h, h)$.

Définition 12.1.3. Soit $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}$.
1. Par récurrence sur $r \in \mathbb{N}$, on dit que f est de ***classe $\mathcal{C}^r$ sur*** U si f est différentiable sur U st sa différentielle $\mathrm{d}f$ est de classe $\mathcal{C}^{r-1}$ sur U. On note $\mathrm{d}_a^r f$ la différentielle de f en un point a de U.
2. On dit que f est de ***classe $\mathcal{C}^\infty$ sur*** U si f est de classe $\mathcal{C}^r$ pour tout $r \in \mathbb{N}$.

Remarque 12.1.4. A nouveau, pour tout $a \in U$, on peut identifier $\mathrm{d}_a^r f$ à une *forme r-linéaire symétrique* sur $\mathbb{R}^n$. Pour $h \in \mathbb{R}^n$Nous noterons $\mathrm{d}_a^r f(h)$ au lieu de $\mathrm{d}_a^r f(h, \ldots, h)$.

12.2 Dérivées d'ordre supérieur et fonctions de classe $\mathcal{C}^k$

On suppose $\mathbb{R}^n$ muni de la base canonique $(\mathrm{e}) = (\mathrm{e}_1, \ldots, \mathrm{e}_n)$. On sait que la différentielle d'une fonction s'exprime d'une fonction $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}$ en un point $a \in U$ s'exprime à l'aide des dérivées partielles de cette fonction, en l'appliquant aux éléments de la base (e) :

$$\frac{\partial f}{\partial x_j}(a) = \mathrm{d}_a f(\mathrm{e}_j), \; j = 1, \ldots, n.$$

De même, les différentielles d'ordre supérieur de f en un point $a \in U$ peuvent s'exprime à l'aide des *dérivées d'ordre supérieur*. Par exemple :

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

On a :

Proposition 12.2.1. *Une fonction $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ est* **de classe $\mathcal{C}^2$ sur** *U si f est de classe $\mathcal{C}^1$ sur U et si les dérivées partielles $\dfrac{\partial f}{\partial x_i}$, $i = 1, \ldots, n$, sont de classe $\mathcal{C}^1$ sur U.*

Le théorème suivant, permet de simplifier considérablement les notations des dérivées partielles.

Théorème 12.2.2 (Schwartz)**.** *Soit $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$ dont les dérivées partielles sont différentiables sur U (mais pas nécessairement de classe $\mathcal{C}^1$). Alors :*

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}, \; \textit{pour tous } i, j = 1, \ldots, n.$$

Démonstration. Soit $a \in U$. On note e_1 et e_2 les deux premiers vecteurs de la base canonique de $\mathbb{R}^n$. On pose :

$$A(t) = f(a + t\mathrm{e}_1 + t\mathrm{e}_2) - f(a + t\mathrm{e}_1) - f(a + t\mathrm{e}_j) + f(a).$$

Pour t fixé au voisinage de 0, on remarque que $A(t) = g(t) - g(0)$ où $g\colon s \mapsto f(a + s\mathrm{e}_1 + t\mathrm{e}_2) - f(a + s\mathrm{e}_1)$. D'après le théorème des accroissements finis, on a $A(t) = tg'(c_t)$ avec $c_t \in]0, t[$. Or :

$$\begin{aligned}
g'(c_t) &= \frac{\partial f}{\partial x_1}(a + c_t\mathrm{e}_1 + t\mathrm{e}_2) - \frac{\partial f}{\partial x_1}(a + c_t\mathrm{e}_1) \\
&= \left(\frac{\partial f}{\partial x_1}(a + c_t\mathrm{e}_1 + t\mathrm{e}_2) - \frac{\partial f}{\partial x_1}(a)\right) - \left(\frac{\partial f}{\partial x_1}(a + c_t\mathrm{e}_1) - \frac{\partial f}{\partial x_1}(a)\right) \\
&= \left(\frac{\partial^2 f}{\partial x_1^2}(a) \cdot c_t + \frac{\partial^2 f}{\partial x_1 \partial x_2}(a) \cdot t + o\left(\sqrt{c_t^2 + t^2}\right)\right) - \left(\frac{\partial^2 f}{\partial x_1^2}(a) \cdot c_t + o(c)\right).
\end{aligned}$$

Puisque $c_t \in [0, t]$, il vient :

$$A(t) = \frac{\partial^2 f}{\partial x_1 \partial x_2}(a) \cdot t^2 + o\left(t^2\right).$$

D'autre part, en écrivant

$$A(t) = h(t) - h(0) \text{ avec } h(s) = f(a + t\mathrm{e}_1 + s\mathrm{e}_2) - f(a + s\mathrm{e}_2),$$

on a de même !

$$A(t) = \frac{\partial^2 f}{\partial x_2 \partial x_1}(a) \cdot t^2 + o\left(t^2\right).$$

On conclut en divisant par t^2 et en faisant tendre t vers 0. □

On peut ainsi formuler à l'aide des dérivées partielles d'ordre supérieur les notions sur les différentielles d'ordre supérieur :

Proposition 12.2.3. *Soit $k \geq 1$. Une fonction $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ est de* **classe $\mathcal{C}^k$** *sur U si :*

1. *f est de classe $\mathcal{C}^1$ sur U,*
2. *les dérivées partielles $\frac{\partial f}{\partial x_i}$, $i = 1, \ldots, n$, sont de classe $\mathcal{C}^{k-1}$ sur U.*

Naturellement, toutes ces notions portant sur des *fonctions* restent valides pour des *applications :*

Définition 12.2.4. Une application $f\colon U \subset \mathbb{R}^n \to \mathbb{R}^m$ est dite de classe $\mathcal{C}^k$ si toutes ses composantes sont de classe $\mathcal{C}^k$.
Les ***dérivées partielles d'ordre k de*** f sont les dérivées partielles d'ordre $k - 1$ des fonctions $\frac{\partial f}{\partial x_i}$, $i = 1, \ldots, n$.

Proposition 12.2.5. *La fonction f est* **de classe $\mathcal{C}^\infty$** *si elle est de classe $\mathcal{C}^k$ pour tout entier $k > 0$.*

Remarque 12.2.6. On déduit immédiatement du Théorème de Schwarz la conséquence suivante. Si f est de classe $\mathcal{C}^k$ sur U, les dérivées partielles d'ordre inférieur à k ne dépendent pas de l'ordre dans lequel elles sont calculées. Nous noterons :

$$\text{si } \alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n \text{ et } |\alpha| = \alpha_1 + \cdots + \alpha_n, \quad \frac{\partial^{|\alpha|} f}{\partial x^\alpha} = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}.$$

Exemple 12.2.7. On vérifiera sur l'exemple suivant que, dans le théorème de Schwarz, l'hypothèse de *différentiabilité* des dérivées partielles $\frac{\partial f}{\partial x_j}$ ne peut pas être remplacée par la simple *existence* des dérivées partielles d'ordre 2. Il s'agit de la fonction

$$f\colon \mathbb{R}^2 \to \mathbb{R}, \quad (x, y) \longmapsto \begin{cases} xy\dfrac{x^2 - y^2}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0), \\ 0 & \text{si } (x, y) = (0, 0). \end{cases}$$

Nous laissons en exercice la preuve du résultat suivant :

Proposition 12.2.8. *La somme et le produit de deux fonctions de classe $\mathcal{C}^k$ sur un ouvert $U \subset \mathbb{R}^n$ sont également de classe $\mathcal{C}^k$. La composée d'applications de classe $\mathcal{C}^k$ est de classe $\mathcal{C}^k$. Enfin, si $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ est une fonction de classe $\mathcal{C}^k$ qui ne s'annule pas sur U, alors son inverse $1/f$ est de classe $\mathcal{C}^k$.*

Exercice 12.2.9. On peut montrer que si les données des énoncés des théorèmes des fonctions implicites ou d'inversion locale sont de classe $\mathcal{C}^k$, alors les applications obtenues (la solution des équations pour le théorème des fonctions implicites, ou le difféomorphisme réciproque pour le théorème d'inversion locale) sont également de classe $\mathcal{C}^k$.

13 Formule de Taylor-Young

13.1 Notations préalables

Notation 13.1.1. Si $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$ et $h = (h_1, \dots, h_n) \in \mathbb{R}^n$, on note

$$\alpha! = \alpha_1! \cdots \alpha_n! \text{ et } h^\alpha = h_1^{\alpha_1} \cdots h_n^{\alpha_n}.$$

A l'aide de ces notations, on peut exprimer l'action des différentielles d'ordre supérieur à l'aide des dérivées partielles d'ordre supérieur.

Proposition 13.1.2. *Si $f \colon U \subset \mathbb{R}^n \to \mathbb{R}$ est différentiable d'ordre m au point $a \in U$, alors :*

$$\mathrm{d}_a^m f \colon h = (h_1, \dots, h_n) \in \mathbb{R}^n \longmapsto \sum_{\substack{\alpha \in \mathbb{N}^n \\ |\alpha| = m}} \frac{m!}{\alpha!} \frac{\partial^{|\alpha|} f}{\partial x^\alpha}(a) \cdot h^\alpha.$$

Remarque 13.1.3. En particulier, on retrouve le fait que $\mathrm{d}_a^1 f(h) = \mathrm{d}_a f(h) = \sum_{i=1}^m \dfrac{\partial f}{\partial x_i}(a)\, h_i$.

Notation 13.1.4. De façon générale $\mathrm{d}_a^m f$ est un ***polynôme homogène*** de degré m en la variable h. C'est le polynôme obtenu en développant la quantité :

$$\left(\sum_{i=1}^n \frac{\partial f}{\partial x_i}(a)\, h_i \right)^{[m]} = \left(\mathrm{d}_a^1 f(h) \right)^{[m]}.$$

comme une "puissance symbolique" d'un multinôme (ce développement est rendu possible grâce au théorème de Schwarz, qui assure la "commutativité" des dérivées partielles). On voit parfois noté $\mathrm{d}_a^m f = (\mathrm{d}_a f)^{[m]}$.

Exemple 13.1.5. Si $f \colon U \subset \mathbb{R}^2 \to \mathbb{R}$ est de classe $\mathcal{C}^2$, on aura :

$$\begin{aligned}
\mathrm{d}_a^1 f(u,v) &= \frac{\partial f}{\partial x}(a)\, u + \frac{\partial f}{\partial y}(a)\, v \text{ et} \\
\mathrm{d}_a^2 f(u,v) &= \left(\frac{\partial f}{\partial x}(a)\, u + \frac{\partial f}{\partial y}(a)\, v \right)^{[2]} = \frac{\partial^2 f}{\partial x^2}(a)\, u^2 + 2\frac{\partial^2 f}{\partial x \partial y}(a)\, uv + \frac{\partial^2 f}{\partial y^2}(a)\, v^2 \\
\mathrm{d}_a^3 f(u,v) &= \left(\frac{\partial f}{\partial x}(a)\, u + \frac{\partial f}{\partial y}(a)\, v \right)^{[3]} = \frac{\partial^3 f}{\partial x^3}(a)\, u^3 + 3\frac{\partial^3 f}{\partial x^2 \partial y}(a)\, u^2 v + 3\frac{\partial^3 f}{\partial x \partial^2 y}(a)\, uv^2 + \frac{\partial^3 f}{\partial y^3}(a)\, v^3 \\
&\dots
\end{aligned}$$

Remarque 13.1.6. On rappelle que la différentielle seconde $\mathrm{d}_a^2 f \colon \mathbb{R}^n \to \mathbb{R}$ est une forme bilinéaire symétrique, grâce au Théorème de Schwarz. On lui associe une *forme quadratique* (c'est à dire un polynôme homogène de degré 2), également notée $\mathrm{d}_a^2 f$. La matrice $H_a f$ de cette forme quadratique dans la base canonique :

$$H_a f = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2}(a) & \dfrac{\partial^2 f}{\partial x_1 \partial x_2}(a) & \cdots \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_2}(a) & \dfrac{\partial^2 f}{\partial x_2^2}(a) & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

est appelée la ***matrice hessienne*** de f au point a.

Si $f \colon U \subseteq \mathbb{R}^n \to \mathbb{R}$ est de classe $\mathcal{C}^2$, nous avons donc :

1. la ***matrice jacobienne*** qui représente la différentielle (première) $\mathrm{d}_a f$ de f un point, qui est une matrice à 1 ligne et n colonnes. On rappelle que $\mathrm{d}_a f$ est une *application linéaire* de $\mathbb{R}^n$ dans $\mathbb{R}$.
2. La ***matrice hessienne***, qui représente la différentielle seconde $\mathrm{d}_a^2 f$ de f en un point, qui est une matrice symétrique carrée de taille n. On rappelle que $\mathrm{d}_a^2 f$ est une *forme bilinéaire* sur $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$.

13.2 Développements de Taylor

Définition 13.2.1. Si $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ est une fonction de classe $\mathcal{C}^k$, et $a = (a_1, \ldots, a_n) \in U$, alors le polynôme

$$T_a^k f(h) = \sum_{m=0}^{k} \frac{1}{m!} \mathrm{d}_a^m f(h)$$

est le $k^{\text{ème}}$ ***polynôme de Taylor de*** f ***au point*** a (en la variable h).

Exemple 13.2.2. Si f est une fonction des variables (x, y), si $a = (x_0, y_0)$ et $h = (k, l)$, alors :

$$T_a^2 f(h) = f(a) + \frac{\partial f}{\partial x}(a)\, h + \frac{\partial f}{\partial x}(a)\, l + \frac{1}{2}\left(\frac{\partial^2 f}{\partial x^2}(a)\, h^2 + 2\frac{\partial^2 f}{\partial x \partial y}(a)\, hl + \frac{\partial^2 f}{\partial y^2}(a)\, l^2\right).$$

Nous pouvons donc maintenant énoncer la formule de Taylor-Toung :

Théorème 13.2.3. **(Formule de Taylor-Young)**. *Soient* $f : U \subset \mathbb{R}^n \to \mathbb{R}$ *une fonction de classe* $\mathcal{C}^r$ *et* a *un point de* U*. Alors ;*

$$f(a+h) = T_a^r f(h) + \mathrm{o}\left(\|h\|^r\right) \text{ quand } h \to 0 \in \mathbb{R}^n.$$

Démonstration. On démontre ce résultat par récurrence sur r.

Si $r = 0$, la continuité de f au point a dit précisément que $f(a+h) = f(a) + o(1)$ (on rappelle que $o(1)$ symbolise $\varepsilon(h)$, une fonction de limite nulle en 0).

Si $r > 0$, supposons le théorème démontré pour $r-1$. On fixe $h \in U$ tel que le segment $[a, h]$ soit contenu dans U. On considère la fonction $g\colon\ [0,1] \subset \mathbb{R} \to U$ définie par :

$$g(t) = f(a+th) - \sum_{k=1}^{r} \frac{t^k}{k!} \mathrm{d}_a^k f(h).$$

On a, pour tout $t \in [0,1]$:

$$\begin{aligned} g'(t) &= \mathrm{d}_{a+th} f(h) - \sum_{k=1}^{r} \frac{t^{k-1}}{(k-1)!} \mathrm{d}_a^k f(h) \\ &= \left(\mathrm{d}_{a+th} f - \sum_{k=0}^{r-1} \frac{t^k}{k!} \mathrm{d}_a^{k-1}(\mathrm{d}f)(h)\right)(h). \end{aligned}$$

On applique à $\mathrm{d}f\colon U \subseteq \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ la formule de Taylor-Young à l'ordre $r-1$: il existe une boule ouverte B centrée en $0 \in \mathbb{R}^n$ sur laquelle :

$$\mathrm{d}f(a+h') = \mathrm{d}f(a) + \sum_{j=1}^{r-1} \frac{1}{j!} \mathrm{d}_a^j(\mathrm{d}f)(h) + \|h'\|^{r-1} \varepsilon(h')$$

où ε est une fonction de limite nulle en 0. Pour $h \in B$ et $t \in [0,1]$, en appliquant cette formule à $h' = th$, on obtient :

$$|g'(t)| \leq \varepsilon(th)\, t^{r-1} \|h\|^r.$$

Puisque

$$g(1) - g(0) = f(a+h) - f(a) + \sum_{k=1}^{r} \frac{1}{k!} \mathrm{d}_a^k f(h),$$

on conclut en appliquant l'inégalité des accroissements finis (Théorème 7.2.2) sur V. □

14 Points critiques d'une fonction

Nous utilisons la formule de Taylor-Young pour étudier les extrema locaux éventuels d'une fonction différentiable.

14.1 Définitions et premières propriétés

Définition 14.1.1. Un point $a \in U \subset \mathbb{R}^n$ est un ***maximum*** (*resp.* ***minimum***) ***local*** d'une fonction $f\colon U \to \mathbb{R}$ s'il existe un voisinage V de a inclus dans U tel que, pour tout $x \in V$, on a $f(x) \leq f(a)$ (*resp.* $f(x) \geq f(a)$). Dans les deux cas, le points a est appelé ***extremum local*** de f.

La recherche des extrema locaux d'une fonction différentiable f est guidée par la connaissance de ses *points critiques* :

Définition 14.1.2. Un point $a \in U \subset \mathbb{R}^n$ est un ***point critique*** d'une fonction différentiable $f\colon U \to \mathbb{R}$ si sa différentielle au point a est nulle (c'est à dire si $\dfrac{\partial f}{\partial x_i}(a) = 0$ pour $i = 1, \ldots, n$).

Proposition 14.1.3. *Si $a \in U \subset \mathbb{R}^n$ est un extremum local de la fonction différentiable $f\colon U \to \mathbb{R}^n$, alors a est un point critique de f.*

Démonstration. On désigne par $(\mathrm{e}_1, \ldots, \mathrm{e}_n)$ la base canonique de $\mathbb{R}^n$. Puisque a est un extremum local de f, les fonctions $g_i : t \mapsto f(a + t\mathrm{e}_i)$, $i = 1, \ldots, n$, admettent un extremum local en $t = 0$. Donc leur dérivées en 0, qui sont les dérivées partielles de f au point a, sont toutes nulles. ☐

Remarque 14.1.4. Il est bien connu que la réciproque de cette proposition est fausse, ne serait-ce que pour les fonctions d'une variable réelle. Un exemple classique en deux variables est la fonction $f\colon (x, y) \mapsto x^2 - y^2$, pour laquelle l'origine est un point critique sans être un extremum local.

14.2 Classification des points critiques d'une fonction

Elle se base sur la formule de Taylor-Young à l'ordre 2.

Proposition 14.2.1. (**Condition nécessaire d'extremum local**) *Soit $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ une fonction de classe $\mathcal{C}^2$ et soit a un point critique de f. Alors :*

1. *Si f admet en a un minimum local, alors la forme quadratique $\mathrm{d}^2_a f$ est positive.*
2. *Si f admet en a un maximum local, alors la forme quadratique $\mathrm{d}^2_a f$ est négative.*

Démonstration. Supposons que f admette un minimum local au point a. Alors pour tout $h \in \mathbb{R}^n$ et $t \in \mathbb{R}$ suffisament petit, on a

$$f(a + th) \geq f(a).$$

On veut démontrer que $\mathrm{d}^2_a f(h) \geq 0$. Supposonse que $\mathrm{d}^2_a f(h)$ soit non nul. Il résulte de la formule de Taylor-Young à l'ordre 2 que

$$f(a + th) = f(a) + t\mathrm{d}_a f(h) + \frac{t^2}{2}\mathrm{d}^2_a f(h) + \mathrm{o}\left(t^2\right) = f(a) + \frac{t^2}{2}\mathrm{d}^2_a f(h) + \mathrm{o}\left(t^2\right).$$

Donc, pour t suffisamment petit, le signe de $f(a + th) - f(a)$ est celui de $\mathrm{d}^2_a f(h)$, ce qui montre que la forme $\mathrm{d}^2_a f(h)$ est positive.

On raisonne de façon analogue si f admet un maximum local au point a. ☐

Remarque 14.2.2. Si la forme quadratique $\mathrm{d}^2_a f$ n'est ni positive, ni négative (et qu'elle a donc une ***signature*** (p, q) avec $p > 0$ et $q > 0$), le point a n'est ni un maximum, ni un minimum de f. On dit que a est un ***point de selle*** (ou un ***col***) de f.

Proposition 14.2.3 (Existence d'un extremum local)**.** *Soit $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ une application de classe $\mathcal{C}^2$ et a un point critique de f. Alors :*

1. *Si la forme $\mathrm{d}^2_a f$ est **définie positive**, la fonction f admet un minimum local au point a.*
2. *Si la forme $\mathrm{d}^2_a f$ est **définie négative**, la fonction f admet un maximum local au point a.*

Démonstration. Supposons que $\mathrm{d}_a^2 f$ soit définie positive. Les valeurs propres de la matrice Hessienne $H_a f$ sont toutes strictement positives. Soit $\lambda > 0$ la plus petite de ces valeurs propres. Un résultat d'algèbre bilinéaire affirme que

$$\inf_{\|h\|=1} \mathrm{d}_a^2 f\,(h) = \lambda \quad (*)$$

Or, d'après la formule de Taylor-Young à l'ordre 2, on a

$$f\,(a+h) - f\,(a) = \frac{1}{2}\mathrm{d}_a^2 f\,(h) + \|h\|^2 \varepsilon\,(h) = \|h\|^2 \left(\frac{1}{2}\mathrm{d}_a^2 f\left(\frac{h}{\|h\|}\right) + \varepsilon\,(h)\right), \text{ avec } \varepsilon\,(h) \underset{h\to 0}{\longrightarrow} 0.$$

Donc, si h est suffisament petit, $|\varepsilon\,(h)| < \lambda/2$ et $f\,(a+h) - f\,(a)$ est positif : f admet un minimum local au point a. □

Remarque 14.2.4. (Démonstration de (*)) On note $q = \mathrm{d}_a^2 f$, et on considère une base orthonormée $(e_1, \ldots, e_n)$ de vecteurs propres. Quitte à renuméroter, on suppose que les valeurs propres de q sont $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Pour tout $x \in \mathbb{R}^n \setminus \{0\}$ on peut écrire $x = \sum_{i=1}^n x_i e_i$, et donc $q\,(x) = \sum_{i=1}^n \lambda_i x_i^2$. De plus $\|x\|^2 = \sum_{i=1}^n x_i^2$.

Exercice 14.2.5 (Cas particulier important, la dimension 2)**.** On note les variables x et y. On considère une fonction $f\colon U \subset \mathbb{R}^2 \to \mathbb{R}$ de classe $\mathcal{C}^2$ et $a \in U$ un point critique de f. On pose

$$r = \frac{\partial f^2}{\partial x^2}\,(a)\,,\ s = \frac{\partial^2 f}{\partial x \partial y}\,(a)\,,\ t = \frac{\partial^2 f}{\partial y^2}\,(a) \qquad \text{(notations de Monge).}$$

Alors :

1. Si $s^2 - rt > 0$, la fonction f admet un point de selle au point a.
2. Si $s^2 - rt < 0$:
   (a) si $r > 0$, la fonction f admet un minimum local au point a.
   (b) si $r < 0$, la fonction f admet un maximum local au point a.

Dans tous les autres cas, il faut une étude spécifique.

Si on travaille en dimension supérieure, le critère précédent se généralise comme suit :

Définition 14.2.6 (Rappels sur les mineurs)**.** Soit $A \in \mathcal{M}_n$ une matrice carrée. Si I et J sont deux sous-ensembles de cardinal k de $\{1, \ldots, n\}$, on désigne par $A_{I,J}$ le $k \times k$-mineur de A obtenu en prenant les indices de ligne dans I et les indices de colonnes dans J.

1. Un mineur de la forme $A_{I,I}$ est dit ***principal***.
2. Si $I = \{1, \ldots, k\}$, le mineur $A_{I,I}$ est appelé ***mineur principal dominant***.

Théorème 14.2.7. *On considère une forme quadratique $q\colon \mathbb{R}^n \to \mathbb{R}$ de matrice symétrique A. Alors :*

1. *La forme quadratique q est **définie positive** si et seulement si tous les mineurs principaux dominants de A sont strictement positifs.*
2. *La forme q est **positive*** si et seulement si tous ses mineurs principaux sont positifs ou nuls.

Remarque 14.2.8. Pour montre qu'une forme quadratique $q\colon \mathbb{R}^n \to \mathbb{R}$ est ***négative*** (*resp.* ***définie négative***) on montre que la forme $-q$ est ***positive*** (*resp.* ***définie positive***).

Si on travaille sur un ensemble fermé A, il faut faire la recherche des extrema de f dans l'intérieur de A à l'aide des résultats précédents, puis chercher les éventuels extrema de f au bord de A.

Lorsque la forme $\mathrm{d}_a^2 f$ est positive sans être positive, les résultats précédents ne permettent pas de conclure. On peut alors utiliser les développements de Taylor d'ordre supérieur.

Proposition 14.2.9. *Soit $f\colon U \subseteq \mathbb{R}^n \to \mathbb{R}$ une fonction de classe $\mathcal{C}^p$ et $a \in U$ tel que $f\,(a) = \mathrm{d}_a^1 f = \cdots = \mathrm{d}_a^{p-1} f = 0$ et $\mathrm{d}_a^p f \neq 0$.*

1. *Si f présente un minimum local (*resp. *maximum local) au point a, alors $\mathrm{d}_a^p f\,(h)$ est positif (*resp. *négatif) pour tout $h \in \mathbb{R}^n$, et donc p est pair et que*
2. *Si p est **pair** et que $\mathrm{d}_a^p f\,(h)$ soit strictement positif (*resp. *négatif) pour tout $h \in \mathbb{R}^n \setminus \{0\}$. Dans ce cas, a est un extremum strict de f.*

Démonstration. La preuve de cette proposition n'est qu'une généralisation immédiate de celle de la Proposition 14.2.3. Il résulte de la formule de Taylor-Young que

$$f(a+h) - f(a) = \frac{1}{p!}\mathrm{d}_a^p f(h) + \mathrm{o}(\|h\|^p). \tag{14.2.1}$$

1. Fixons $x \in \mathbb{R}^n$. La fonction $t \mapsto f(a+tx) - f(a)$ est alors définie au voisinage de $0 \in \mathbb{R}$. D'après l'équation (14.2.1) appliquée à $h = tx$, on a :

$$f(a+tx) - f(a) = \frac{t^p}{p!}\mathrm{d}_a^p f(x) + \mathrm{o}(t^p). \tag{14.2.2}$$

Puisque f possède un minimum local au sens large au point a, on a $f(a+tx) - f(a) \geq 0$ pour t assez petit. On déduit de (14.2.2) que p est pair et que $\mathrm{d}_a^p f(x) \geq 0$.

2. Avec ces hypothèses la fonction $h \mapsto \mathrm{d}_a^p f(h)$ est strictement positive sur la sphère unité $\mathcal{S}^1 = \{x \in \mathbb{R}^n : \|x\| = 1\}$. Puisque $\mathcal{S}^1$ est compacte, $\mathrm{d}_a^p f$ atteint sur $\mathcal{S}^1$ sont minimum, qui est un nombre $\alpha > 0$. Donc, pour tout $h \in \mathbb{R}^n \setminus \{0\}$, on a :

$$\mathrm{d}_a^p f\left(\frac{h}{\|h\|}\right) \geq \alpha, \text{ ou encore } \mathrm{d}_a^p f(h) \geq \alpha \|h\|^p.$$

Soit $0 < \varepsilon < \frac{\alpha}{2p!}$. Alors, si h est assez petit :

$$f(a+h) - f(a) \geq \frac{\alpha}{p!}\|u\|^p - \varepsilon\|u\|^p \geq \frac{(\alpha/2)}{p!}\|u\|^p.$$

Ainsi $f(a+h) - f(a) > 0$ pour h assez petit et non nul. □

Exemple 14.2.10. **1.** La fonction $(x, y) \mapsto x^2 + y^4$ admet un minimum strict à l'origine.

2. La fonction $(x, y) \mapsto x^2 + y^3$ n'admet pas d'extremum à l'origine (considérer la restriction de f à la droite $(0, v)$, $v \in \mathbb{R}$).

3. Soit $f : \mathbb{R}^2 \to \mathbb{R}$ définie par

$$f(x, y) = \lambda x^4 + y^4 + 4x^3 y + x^5 - y^5.$$

Pour étudier le signe de $P(x, y) = \lambda x^4 + y^4 + 4x^3 y$ on pose $y = tx$. On est ramené à l'étude des racines de $t^4 + 4t + \lambda$.

1. Pour $\lambda > 3$, ce polynôme n'a pas de racines réelles et $P(x, y) > 0$ pour $(x, y) \neq (0, 0)$: minimum strict à l'origine.
2. Pour $\lambda < 3$, ce polynôme change de signe, donc pas d'extremum à l'origine.
3. Pour $\lambda = 3$, on a $P(x, y) = (x+y)^2\left(3x^3 - 2xy + y^2\right)$, d'où $f(x, -x) = 2x^5$: donc f change de signe dans tout voisinage de l'origine, qui n'est donc pas un extremum.

Exemple 14.2.11. Pour $\lambda \geq 0$, on considère $f_\lambda(x, y) = x^3 + y^3 - 3\lambda xy$.

1. Si $\lambda = 0$, f_0 a $(0, 0)$ comme unique point critique, qui n'est ni minimum, ni maximum.
2. Si $\lambda > 0$, le point $(0, 0)$ est point critique (ni maximum, ni minimum), et $a = (\lambda, \lambda)$ également. On a $\mathrm{d}_a^2 f(h) > 0$ pour tout $h \in \mathbb{R}^2 \setminus \{0\}$, donc a est un minimum local. Mais il n'est pas global, car $f(0, x) = x^3$ varie de $-\infty$ à $+\infty$.

15 Extrema liés

15.1 Problème et exemples

Exemple 15.1.1. (*exemple tiré du livre Mathématiques pour la Licence - vol. 3*) On doit construire un hangar de côtés x, y, de hauteur z, dont le volume doit atteindre $1200m^3$. On impose la contrainte supplémentaire $x = 2z$. Sachant que le coût du mètre carré au sol est de 10 euros, celui des murs de 20 euros, et celui du toit de 40 euros, déterminer les dimensions x, y et z pour que le coût soit minimisé.

Exemple 15.1.2. Déterminer le point de la courbe d'équation $2x^2 + 13y^2 + 2xy - 4 = 0$ situé le plus près de l'origine

Problème général. On considère une sous-variété lisse $X \subset \mathbb{R}^n$ et une fonction $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^1$. On se demande comment trouver les extrema locaux de f *restreinte à* $X \cap U$.

15.2 Méthode de Lagrange

Les deux résultats suivants résument les méthodes permettant de répondre au problème ci-dessus.

Proposition 15.2.1. *Soit $X \subset \mathbb{R}^n$ une sous-variété lisse et $g\colon U \subseteq \mathbb{R}^n \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$. Si $a \in X \cap U$ est un extremum local de la restriction de f à X, alors*

$$T_a X \subset a + \ker (\mathrm{d}_a g)\,.$$

Démonstration. On considère un paramètrage local $\varphi\colon W \subset \mathbb{R}^p \to \mathbb{R}^n$ de $X \cap U$ tel que $\varphi(0) = a$. Si g restreinte à X admet un extremum local au point a, alors la fonction $g \circ \varphi$ admet un extremum local en $0 \in \mathbb{R}^p$. Par conséquent :

$$\mathrm{d}_0 (g \circ \varphi) = 0.$$

On sait que $T_a X$ est dirigé par les vecteurs $\dfrac{\partial \varphi}{\partial s_i}(0)$, $i = 1, \ldots, p$. On a donc :

$$\mathrm{d}_a g \left(\frac{\partial \varphi}{\partial s_i}(0) \right) = \mathrm{d}_a g \left(\mathrm{d}_0 \varphi \left(\mathrm{e}_i \right) \right) = \mathrm{d}_0 (g \circ \varphi)(\mathrm{e}_i) = 0, \quad i = 1, \ldots, s.$$

Ce qui prouve la proposition. □

Théorème 15.2.2. *Soient $f_1, \ldots, f_q\colon U \subset \mathbb{R}^n \to \mathbb{R}$ des fonctions indépendantes de classe $\mathcal{C}^1$, et $g\colon U \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$. Si la fonction f restreinte à la sous-variété X de $\mathbb{R}^n$ définie par les équations $f_1 = 0, \ldots, f_q = 0$ admet un extremum local au point $a \in X$, alors il existe des nombres réels $\lambda_1, \ldots, \lambda_q$ tels que*

$$\mathrm{d}_a g = \sum_{i=1}^{q} \lambda_i \mathrm{d}_a f_i.$$

*Les nombres $\lambda_1, \ldots, \lambda_q$ sont appelés **multiplicateurs de Lagrange**.*

Démonstration. On considère un paramétrage local $\varphi\colon W \subseteq \mathbb{R}^p \to \mathbb{R}^n$ de X au voisinage de a tel que $\varphi(0) = a$. D'après la proposition précédente, les vecteurs $\dfrac{\partial \varphi}{\partial x_i}(0)$ appartiennent tous au noyau de la forme linéaire $\mathrm{d}_a g$. Puisque ces vecteurs forment une base de l'espace vectoriel $\cap_{j=1}^{q} \ker \mathrm{d}_a f_j$, on en déduit que $\mathrm{d}_a g$ s'annule sur $\cap_{j=1}^{q} \ker \mathrm{d}_a f_j$. Donc $\mathrm{d}_a g$ appartient à l'espace $(T_a X)^\circ$, qui est le dual orthogonal de $T_a X$ (l'espace des formes linéaires qui s'annulent sur $T_a X$), dont les formes linéaires $\mathrm{d}_a f_j$ forment une base. Ainsi $\mathrm{d}_a g$ est une combinaison linaire de $\mathrm{d}_a f_1, \ldots, \mathrm{d}_a f_q$. □

15.3 Exercice

Exercice. Déterminer le (ou les) point de la courbe $\mathcal{C}$ d'équation $f(x, y) = x^2 - 2xy + y^2 - 3x - 3y - 3 = 0$ le plus proche de l'origine.

Solution. Il s'agit donc de chercher le maximum de la fonction $g\colon (x, y) \mapsto x^2 + y^2$ en restriction à la courbe $\mathcal{C}$. En vertu du Théorème 15.2.2, on cherche les points $a \in \mathcal{C}$ en lesquels il existe un nombre réel $\lambda \in \mathbb{R}$ tel que $\mathrm{d}_a g = \lambda \mathrm{d}_a f$, c'est-à-dire en lesquels les gradients de f et de g sont colinéaires. Or :

$$\mathrm{grad} f(x, y) = \begin{pmatrix} -2x - 2y - 3 \\ -2x + 2y - 3 \end{pmatrix}, \text{ et } \mathrm{grad} g(x, y) = 2 \begin{pmatrix} x \\ y \end{pmatrix}.$$

L'équation de colinéarité de ces deux vecteurs est donc :

$$\begin{aligned} y(-2x - 2y - 3) - x(-2x + 2y - 3) &= 0 \\ (y - x)(3 + 2x + 2y) &= 0. \end{aligned}$$

On cherche donc des points (x, y) vérifiant le système de deux équations, constitué de l'équation de la courbe $\mathcal{C}$ et de l'équation de colinéarité des gradients :

$$\begin{aligned} (x-y)^2 - 3\,(x+y) - 3 = 0 \\ (y-x)\,(3+2x+2y) = 0. \end{aligned}$$

Donc soit

(a) $y = x$. Cela donne $x + y = -1$ et donc $a = (x, y) = \left(-\frac{1}{2}, -\frac{1}{2}\right)$.

(b) $3+2x+2y = 0$, c'est-à-dire $x+y = -\frac{3}{2}$. En reportant dans l'équation de la courbe, cela donne $(x-y)^2 + \frac{9}{2} - 3 = 0$, soit $(x-y)^2 = -\frac{3}{2}$, ce qui est impossible.

L'unique point vérifiant les conditions de Lagrange est donc $a = \left(-\frac{1}{2}, -\frac{1}{2}\right)$.

On veut maintenant savoir la nature du point a : maximum local sur $\mathcal{C}$, minimum local sur $\mathcal{C}$, autres$\cdots$? On étudie donc le comportement de l'accroissement $g\,(a+h) - g\,(a)$, où $a + h \in \mathcal{C}$. On pose $h = (h_1, h_2)$. On utilise pour cela la formule de Taylor pour la fonctin g au point a à l'ordre 2 (sans reste car g est un polynôme de degré 2) :

$$\begin{aligned} g\,(a+h) &= g\,(a) + \frac{\partial g}{\partial x}\,(a)\,h_1 + \frac{\partial g}{\partial y}\,(a)\,h_2 + \frac{1}{2!}\left(\frac{\partial^2 g}{\partial x^2}\,(a)\,h_1 + 2\frac{\partial^2 g}{\partial x \partial y}\,(a)\,h_1 h_2 + \frac{\partial^2 f}{\partial y^2}\,(a)\right) \\ &= \frac{1}{2} - (h1 + h2) + h_1^2 + h_2^2. \end{aligned}$$

Or, puisque $a + h \in \mathcal{C}$, on a $\left(-\frac{1}{2} + h_2 + \frac{1}{2} - h_1\right)^2 - 3\left(-\frac{1}{2} + h_1 - \frac{1}{2} + h_2\right) - 3 = (h_1 - h_2)^2 - 3\,(h_1 + h_2) = 0$. Donc

$$g\,(a+h) = g\,(a) + \frac{1}{3}\,(h_1 - h_2)^2 + h_1^2 + h_2^2 = g\,(a) + \frac{1}{3}\left(h_1^2 + h_2^2 + (h_1 + h_2)^2\right).$$

Ce qui implique, puisque $\frac{1}{3}\left(h_1^2 + h_2^2 + (h_1 + h_2)^2\right) > 0$, que f admet un minimum global strict au point $\left(-\frac{1}{2}, -\frac{1}{2}\right)$ sur $\mathcal{C}$.