Calcul Différentiel

Julia Matos

Université d'Evry Val-d'Essonne
2012/2013

Table des matières

1 L'espace $\mathbb{R}^n$ - Rappels

1.1 Généralités

Nous énonçons ce qu'il faut savoir sur les espaces de dimension finie.

Soit E un espace vectoriel réel ou complexe (sur le corps $\mathbb{K}$ où $\mathbb{K} = \mathbb{R}$ ou $\mathbb{K} = \mathbb{C}$).

Définition 1.1 *Une norme sur l'espace vectoriel E est une application $N : E \to \mathbb{R}^+$ vérifiant :*

1. *$N(x) \geq 0$, $\forall x \in E$ et $N(x) = 0 \iff x = 0$ (positivité stricte).*
2. *$N(\lambda x) = |\lambda| N(x)$, $\forall x \in E$ et $\forall \lambda \in \mathbb{K}$ (homogénéité).*
3. *$N(x + y) \leq N(x) + N(y)$, $\forall x, y \in E$ (inégalité triangulaire).*

Un espace vectoriel E doté d'une norme est dit un espace vectoriel normé (e.v.n. en abrégé).

Si $x \in E$, le nombre $N(x)$ s'appelle la norme de x et se note usuellement par $N(x) = \|x\|$.

On a les propriétées suivantes :

1. Pour tout $x \in E$, $\|-x\| = \|x\|$.
2. Pour tous $x, y \in E$, $|\|x\| - \|y\|| \leq \|x\| - \|y\|$.

Exemples :

1. Soit $E = \mathbb{R}^n$, $n \geq 1$ et $\mathbb{K} = \mathbb{R}$. Alors
$$\|x\|_2 = \sqrt{x_1^2 + \ldots + x_n^2}, \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n,$$
est la norme associé au produit scalaire canonique de $\mathbb{R}^n$ qui est définit par $(x|y) = \sum_{i=1}^n x_i y_i$. Cette norme $\|\cdot\|_2$ s'appelle la norme euclidienne. Dans le cas $n = 2$, c'est la norme correspondant au théorème de Pythagore.
2. Il y a d'autres normes sur $\mathbb{R}^n$, dites normes usuelles. Pour tout $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, on pose
$$\|x\|_1 = \sum_{i=1}^n |x_i|, \quad \|x\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$
Les applications $\|\cdot\|_1$ et $\|\cdot\|_\infty$ ainsi définies sont aussi des normes sur $\mathbb{R}^n$ et s'appellent respectivement la norme ℓ^1 et la norme du max ou norme infinie.

Définition 1.2 *Soient N_1 et N_2 deux normes sur E. On dit que N_1 et N_2 sont normes équivalentes s'il existe $\alpha, \beta > 0$ telles que*
$$\alpha N_2(x) \leq N_1(x) \leq \beta N_2(x), \ \ \forall x \in E.$$

Exemple : Les normes sur $\mathbb{R}^n$ définies dans les exemples précédents sont toutes équivalentes. En effet,
$$\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 \leq n\|x\|_\infty, \quad \forall x \in \mathbb{R}^n.$$

Théorème 1.1 *Toutes les normes d'un espace vectoriel de dimension finie sont équivalentes.*

La notion de norme permet de définir la distance entre deux points d'un e.v.n. E.

Définition 1.3 *Soit $(E, \|\cdot\|)$ un e.v.n.. On définit l'application $d : E \times E \to \mathbb{R}^+$ par*

$$d(x, y) = \|x - y\|.$$

Cette application s'appelle distance sur E associée à la norme $\|\cdot\|$. Elle vérifie les propriétés suivantes :

1. *$d(x, y) \geq 0$, $\forall x, y \in E$ et $d(x, y) = 0 \iff x = y$.*
2. *$d(x, y) = d(y, x)$, $\forall x, y \in E$ (symétrie).*
3. *$d(x, y) \leq d(x, z) + d(z, y)$, $\forall x, y, z \in E$ (inégalité triangulaire).*

Étant donné $x \in E$ et $r > 0$, la boule ouverte de centre x et rayon r est définie par :

$$B(x, r) = \{y \in E : \|x - y\| < r\}.$$

La boule fermée de centre x et rayon $r \geq 0$ est définie par :

$$\bar{B}(x, r) = \{y \in E : \|x - y\| \leq r\}.$$

On appelle *boule unité* de E la boule de centre 0 et rayon 1 : $B(0, 1)$ et *sphère unité* à l'ensemble $S = \{x \in E : \|x\| = 1\}$.

Quelques notions de topologie :

1. Une partie $A \subset E$ est dite bornée s'il existe une boule qui la contient (c'est-à-dire, il existe $x \in E$ et $r > 0$ tels que $A \subset B(x, r)$).
2. Une application $f : X \to E$ dans un e.v.n. est dite bornée si son image $f(X)$ est une partie bornée de E.
3. Une partie A est un voisinage d'un point $a \in E$ s'il existe une boule ouverte centrée en a contenue dans A.

   Toute intersection finie de voisinages de a est un voisinage de a. Tout ensemble qui contient un voisinage de a est encore voisinage de a.
4. Une partie $O \subset E$ est dite ouverte si, pour tout $x \in O$, O est voisinage de x.

   L'ensemble vide et E sont ouverts. Toute boule ouverte est un ouvert.
5. Toute partie ouverte au sens d'une norme l'est au sens de toute autre norme équivalente.
6. On dira qu'une partie est fermée si son complémentaire est ouvert.
7. L'adhérence d'une partie A de E, notée $\bar{A}$, est l'intersection de tous les fermés contenant A. Il s'agit du plus petit fermé (au sens de l'inclusion) contenant A.
8. L'intérieur d'une partie A de E, notée $\mathrm{int}A$, est la réunion de tous les ouverts contenus dans A. C'est le plus grand ouvert contenu dans A.
9. Étant donnée deux points $a, b \in E$, on définit le segment reliant a à b comme l'ensemble

$$[a, b] = \{(1 - t)a + tb : t \in [0, 1]\}.$$

Dans un espace de dimension finie, on n'a donc pas besoin de préciser avec quelle norme on travaille. Souvent, on choisira la norme adaptée au problème posé. Tout espace vectoriel de dimension est isomorphe à $\mathbb{R}^n$, et il suffit donc de savoir ce qui se passe dans $\mathbb{R}^n$.

Définition 1.4 *Soit $(x_k)_{k\in\mathbb{N}}$ est une suite d'éléments de $(E, \|\cdot\|)$. On dit que $(x_k)_{k\in\mathbb{N}}$ converge vers $x \in E$ si*

$$\lim_{k\to+\infty} \|x_k - x\| = 0,$$

et on note $\lim_{k\to+\infty} x_k = x$.

Remarques.

1. On peut montrer que cette notion de limite ne dépend pas de la norme choisie (parmi des normes équivalentes).
2. Si $E = \mathbb{R}^n$ et $x = (x_1, \cdots, x_n)$, $x_k = (x_{k,1}, \cdots, x_{k,n})$, alors

$$x_k \longrightarrow x \iff \forall 1 \leq i \leq n, \quad x_{k,i} \longrightarrow x_i.$$

Définition 1.5 *Une suite $(x_k)_{k\in\mathbb{N}}$ de E est dite de Cauchy si $\|x_{k+m} - x_k\| \to 0$ lorsque $k, m \to +\infty$.*

Définition 1.6 *Soit E un espace vectoriel normé. On dit que E est un espace complet si toute suite de Cauchy dans E y est convergente.*

Théorème 1.2 (Théorème de Bolzano-Weierstrass) *Dans un espace vectoriel normé de dimension finie, toute suite bornée contient une sous-suite convergente.*

On peut aussi dire que toute suite bornée admet au moins une valeur d'adhérence. Autre formulation du Théorème de Bolzano-Weierstrass :

Dans un espace vectoriel normé de dimension finie, une partie est compacte si et seulement si elle est fermée et bornée.

Propriétés :

1. La boule unité d'un espace vectoriel normé est compacte si et seulement si cet espace est de dimension finie.
2. Tout espace vectoriel de dimension finie est complet.

1.2 Fonctions à plusieurs variables

On regarde les fonctions de $\mathbb{R}^n$ dans $\mathbb{R}$ ou dans $\mathbb{R}^p$.

1.2.1 Limite et continuité

Définition 1.7 *Soit A une partie de $\mathbb{R}^n$ et $f : A \to \mathbb{R}^p$. On dit que f tend vers l quand x tend vers a lorsque :*

$$\forall \varepsilon > 0, \quad \exists \delta > 0 \text{ tel que } \|x - a\|_{\mathbb{R}^n} < \delta \implies \|f(x) - l\|_{\mathbb{R}^p} < \varepsilon.$$

On note $\lim\limits_{x\to a} f(x) = l$.

Exemple 1. La fonction $f(x, y) = \dfrac{x^2y + y^2 \sin x}{x^2 + y^2}$ tend vers 0 quand (x, y) tend vers $(0, 0)$ dans $\mathbb{R}^2$.

Exemple 2. La fonction $f(x, y) = \dfrac{xy}{x^2 + y^2}$ n'a pas de limite quand (x, y) tend vers $(0, 0)$ dans $\mathbb{R}^2$.

Définition 1.8 *Soit A une partie de $\mathbb{R}^n$ et $f : A \to \mathbb{R}^p$. On dit que f est continue en $a \in A$ si $\lim\limits_{x \to a} f(x) = f(a)$.*

Si f est continue en tout point $a \in A$, on dit que f est continue sur A.

La fonction f est continue en a si et seulement si pour toute suite (x_k) convergeant vers a, la suite image $f(x_k)$ converge vers $f(a)$.

1.2.2 Théorèmes sur les fonctions continues

Théorème 1.3 *L'image d'un compact par une application continue est encore un compact.*

Théorème 1.4 *Toute fonction continue sur un compact K de $\mathbb{R}^n$ à valeurs réelles atteint son minimum absolu et son maximum absolu sur K.*

Le théorème suivant donne une caractérisation des fonctions continues. Il sert surtout à montrer qu'une partie de $\mathbb{R}^n$ est ouverte ou fermé.

Théorème 1.5

1. *Une fonction est continue si et seulement si l'image réciproque de tout ouvert est un ouvert.*
2. *Une fonction est continue si et seulement si l'image réciproque de tout fermé est un fermé.*

2 Applications différentielles

Ce chapitre est consacré à la définition de la différentielle et à ses propriétés élémentaires.

2.1 Notations et définitions

Définition 2.1 *Soient $x \in \mathbb{R}^n$ et $f : \mathbb{R}^n \to \mathbb{R}^p$. On dira qu'une fonction $f(x)$ est un $o(x)$ ("petit o de x") si*

$$f(0) = 0 \quad \textit{et} \quad \lim_{x \to 0_{\mathbb{R}^n}} \frac{f(x)}{\|x\|_{\mathbb{R}^n}} = 0_{\mathbb{R}^p},$$

c'est-à-dire, pour tout $\varepsilon > 0$ il existe $r > 0$ tel que

$$\|x\|_{\mathbb{R}^n} \leq r \implies \|f(x)\|_{\mathbb{R}^p} \leq \varepsilon \|x\|_{\mathbb{R}^n},$$

ou encore que $f(x)$ est de la forme :

$$f(x) = \|x\|_{\mathbb{R}^n} \varepsilon(x) \quad \textit{avec} \quad \lim_{x \to 0_{\mathbb{R}^n}} \varepsilon(x) = 0_{\mathbb{R}^p}.$$

Exemple : Les fonctions $f(x) = \|x\|^2_{\mathbb{R}^n}$ et $g(x) = \|x\|^{1+\alpha}_{\mathbb{R}^n}$ avec $\alpha > 0$ sont des $o(x)$ (à valeurs réelles).

Rappel : Une fonction $f : \mathbb{R} \to \mathbb{R}$ est différentiable en $a \in \mathbb{R}$ s'il existe $f'(a) \in \mathbb{R}$ tel que

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

Cette définition n'a pas de sens pour une application $f : \mathbb{R}^n \to \mathbb{R}^m$, avec $n > 1$ (on ne sait pas diviser par un vecteur $h \in \mathbb{R}^n$).

Le calcul différentiel est un outil d'analyse qui permet d'approcher certaines fonctions par des fonctions plus simples, par exemple des fonctions affines.

Si f est dérivable en $a \in \mathbb{R}$,

$$f(x) = f(a) + f'(a)(x-a) + o(x-a) = \text{polynôme du premier degré} + o(x-a).$$

Si f est suffisamment dérivable, on peut encore mieux l'approcher par un polynôme de degré supérieur : celui de son développement limité.

La généralisation à $\mathbb{R}^n$ se fait en définissant la différentiabilité de f en un point a, comme la possibilité d'approcher $f(x)$ au voisinage de a par une expression affine :

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

On désignera $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ l'espace des applications linéaires de $(\mathbb{R}^n, \|\cdot\|_{\mathbb{R}^n})$ vers $(\mathbb{R}^p, \|\cdot\|_{\mathbb{R}^p})$. Cet espace $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ est lui-même un espace normé pour la norme définie par

$$\|L\| = \sup_{\|x\|_{\mathbb{R}^n} \leq 1} \|L(x)\|_{\mathbb{R}^p} = \sup_{\|x\|_{\mathbb{R}^n} = 1} \|L(x)\|_{\mathbb{R}^p} = \sup_{\|x\|_{\mathbb{R}^n} \neq 0} \frac{\|L(x)\|_{\mathbb{R}^p}}{\|x\|_{\mathbb{R}^n}}.$$

Notons que

$$\forall x \in \mathbb{R}^n, \quad \|L(x)\|_{\mathbb{R}^p} \leq \|L\| \|x\|_{\mathbb{R}^n}.$$

On vérifie que la norme de L est égale à la borne inférieure des réels $C \geq 0$ tels que

$$\forall x \in \mathbb{R}^n, \quad \|L(x)\|_{\mathbb{R}^p} \leq C\|x\|_{\mathbb{R}^n}.$$

De plus, si pour un tel C il existe $x_0 \neq 0$ tel que $\|L(x_0)\|_{\mathbb{R}^p} = C\|x_0\|_{\mathbb{R}^n}$, alors $\|L\| = C$.

On remarque que toute application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$ est une application continue. Le résultat suivant est plus général.

Proposition 2.1 *Soient E et F des espaces vectoriels normés et $L : E \to F$ une application linéaire. Si E est de dimension finie, alors L est une application continue.*

Preuve. Soit $(e_1, \ldots, e_n)$ une base de E. Pour tout $x \in E$, on a

$$\|L(x)\|_F = \|\sum_{i=1}^{n} x_i L(e_i)\|_F \leq \sum_{i=1}^{n} |x_i| \|L(e_i)\|_F \leq CN(x),$$

où $C = \max\{\|L(e_i)\|_F : 1 \leq i \leq n\}$ et $N(x) = \sum_{i=1}^{n} |x_i|$. N définit une norme sur E et d'autre part, puisque E est de dimension finie, toutes les normes sur E sont équivalentes. Alors, il existe $\beta > 0$ tel que $N(x) \leq \alpha\|x\|_E$, pour tout $x \in E$. Donc, $\|L(x)\|_F \leq \alpha C\|x\|_E$, pour tout $x \in E$ et par conséquent, pour tous $x, y \in E$,

$$\|L(x) - L(y)\|_F = \|L(x - y)\|_F \leq \alpha C\|x - y\|_E.$$

On conclut que L est lipschitzienne sur E et donc elle est continue sur E.

Soit Ω un ouvert de $\mathbb{R}^n$ et $f : \Omega \to \mathbb{R}^p$.

Définition 2.2 *On dit que f est différentiable au point $a \in \Omega$, s'il existe une application lnéaire $L : \mathbb{R}^n \to \mathbb{R}^p$ telle que*

$$\lim_{h \to 0} \frac{\|f(a + h) - f(a) - L(h)\|_{\mathbb{R}^p}}{\|h\|_{\mathbb{R}^n}} = 0.$$

Plus précisément :

$$\forall \varepsilon > 0, \quad \exists r > 0 : \|h\|_{\mathbb{R}^n} \leq r \implies \|f(a + h) - f(a) - L(h)\|_{\mathbb{R}^p} \leq \varepsilon\|h\|_{\mathbb{R}^n},$$

ou encore de manière équivalente :

$$f(a + h) = f(a) + L(h) + o(h), \quad \textit{pour tout } h \textit{ suffisamment petit.}$$

Proposition 2.2

1. *Si L existe, elle est unique. On l'appelle la différentielle de f en a et on note*
$$Df(a) \quad \textit{ou} \quad df(a) \quad \textit{ou même parfois} \quad f'(a).$$
2. *Si f est différentiable en a, alors f est continue en a.*
3. *Localité : f est différentiable en a si et seulement si la restriction de f à tout ouvert $\mathcal{U}$ de Ω tel que $a \in \mathcal{U}$ est différentiable en a.*
4. *On ne change pas la différentiabilité et la différentielle en remplaçant la norme par des normes équivalentes.*

Preuve :

1. Si L_1 et L_2 vérifient la définition, alors pour $\varepsilon > 0$ fixé, il existe $r > 0$ tel que, si $\|h\|_{\mathbb{R}^n} \leq r$,
$$\|f(a+h) - f(a) - L_1(h)\|_{\mathbb{R}^p} \leq \varepsilon \|h\|_{\mathbb{R}^n} \quad \text{et} \quad \|f(a+h) - f(a) - L_2(h)\|_{\mathbb{R}^p} \leq \varepsilon \|h\|_{\mathbb{R}^n}.$$
Alors,
$$\|L_1(h) - L_2(h)\|_{\mathbb{R}^p} \leq 2\varepsilon \|h\|_{\mathbb{R}^n},$$
pour tout h suffisamment petit. Donc, $\|L_1 - L_2\| \leq 2\varepsilon$. Puisque $\varepsilon > 0$ est arbitraire, on conclut que $L_1 = L_2$.
2. Si f est différentiable en a, alors
$$\|f(a+h) - f(a)\|_{\mathbb{R}^p} \leq \|L\| \|h\|_{\mathbb{R}^n} + o(h),$$
et donc $\lim_{h \to 0} f(a+h) = f(a)$.
3. Évident par la définition.
4. L'égalité $f(a+h) - f(a) - L(h) = o(h)$ ne dépend pas de la norme (équivalente) choisie.

Remarques :

1. Si elle existe, $Df(a)$ est une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$. La différentielle $Df(a)$ est répresentée par une matrice à p lignes et n colonnes appellée la *matrice jacobienne* de f en a.
2. Si $n = p = 1$, les matrices n'ayant pas de parenthèses, l'application linéaire $Df(a)$ s'identifie au nombre dérivée
$$Df(a)(1) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} = f'(a).$$
3. Au voisinage de a, l'application f se comporte a peu près comme l'application $x \mapsto f(a) + Df(a)(x-a)$ (somme d'une constante et d'une application linéaire), pour laquelle on peut utiliser les outils d'algèbre linéaire : calcul matriciel, rang, ...

Pour montrer qu'une fonction est différentiable en un point, on utilise la plupart du temps la condition suffisante du paragraphe suivant, avec les dérivées partielles. Quand la fonction est compliquée (souvent à l'origine), il faut revenir à la définition.

Exemple : Soit $n = p = 2$. On considère sur $\mathbb{R}^2$ la norme ℓ^1 :
$$\|(x, y)\| = |x| + |y|, \quad (x, y) \in \mathbb{R}^2.$$
Soit f définie sur $\mathbb{R}^2$ par : $f(x, y) = (x + y, xy)$. Si $a = (a_1, a_2) \in \mathbb{R}^2$ et $h = (h_1, h_2) \in \mathbb{R}^2$, on a
$$f(a+h) - f(a) = (h_1 + h_2, a_2 h_1 + a_1 h_2) + (0, h_1 h_2).$$
Puisque $\|(0, h_1 h_2)\| = |h_1 h_2| \leq (|h_1| + |h_2|)^2 = \|h\|^2$, alors $(0, h_1 h_2) = o(h)$. Donc, f est différentiable en a et
$$Df(a)(h) = (h_1 + h_2, a_2 h_1 + a_1 h_2) = \begin{pmatrix} 1 & 1 \\ a_2 & a_1 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}.$$
La matrice jacobienne de f en a est $\begin{pmatrix} 1 & 1 \\ a_2 & a_1 \end{pmatrix}$.

Définition 2.3 *On dit que f est différentiable sur Ω si, pour tout $x \in \Omega$, f est différentiable en x. Alors, $Df : \Omega \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ s'appelle la différentielle de f.*

Définition 2.4 *On dit que f est C^0 (ou de classe C^0) si elle est continue dans Ω (et on pose $D^0 f = f$).*

On dit que f est continûment différentiable sur Ω, ou C^1, si Df est continue dans Ω.

Remarquer que si f est différentiable sur Ω, pour tout $x \in \Omega$, $Df(x)$ est continue, alors que Df peut l'être ou pas.

Définition 2.5 *Si Df est elle même différentiable sur Ω, on dit que f est deux fois différentiable et, l'on note $D(Df) = D^2 f$. C'est un élément de $\mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p))$ qu'on appelle la différentielle seconde de f. En définissant par récurrence la notation*

$$D^k f = D(D^{k-1} f),$$

on dit que f est n fois différentiable ($n \geq 1$) si $D^{n-1} f$ est différentiable. On appelle différentielle d'ordre n de f à $D^n f$.

Définition 2.6 *On dit que f est C^n (ou de classe C^n) si $D^n f$ est continue. Elle est dite C^∞ (ou de classe C^∞) si $D^n f$ existe pour tout entier $n \geq 0$.*

2.2 Propriétés

Linéairité de la différentielle

Soit Ω un ouvert de $\mathbb{R}^n$, $f : \Omega \to \mathbb{R}^p$ et $g = \Omega \to \mathbb{R}^p$ différentiables en $a \in \Omega$ (resp. C^k sur Ω) alors, pour tous $\lambda, \mu \in \mathbb{R}$, $\lambda f + \mu g$ est différentiable en a (resp. C^k sur Ω) et

$$D(\lambda f + \mu g)(a) = \lambda Df(a) + \mu Dg(a).$$

Différentielle d'une constante

Soit $f : \Omega \to \mathbb{R}^p$ une fonction constante, c'est-à-dire

$$\forall x \in \Omega, \quad f(x) = y_0 \in \mathbb{R}^p.$$

Alors f est C^∞ et $Df(x) = 0$ (application identiquement nulle), pour tout $x \in \Omega$.

En effet, $f(x+h) - f(x) = 0 = o(h)$.

Différentielle d'une application linéaire

Soit $\varphi \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$. Alors φ est différentiable dans $\mathbb{R}^n$ (et même de classe C^∞) et $D\varphi(x) = \varphi$, pour tout $x \in \mathbb{R}^n$.

En effet, $\varphi(x+h) - \varphi(x) - \varphi(h) = 0 = o(h)$. Ainsi, $D\varphi : E \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ est une application constante, donc elle est différentiable et sa différentielle est l'application identiquement nulle.

Différentielle d'une application bilinéaire

Soit $b : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ une application bilinéaire, c'est-à-dire b est linéaire dans les variables x et y. Alors, b est C^∞ et

$$Db(a)(h) = b(h_1, a_2) + b(a_1, h_2), \quad a = (a_1, a_2), \quad h = (h_1, h_2).$$

Preuve : Soit $a = (a_1, a_2) \in \mathbb{R}^n \times \mathbb{R}^m$ fixé. On a $b(a+h) - b(a) = b(h_1, a_2) + b(a_1, h_2) + b(h_1, h_2)$. L'application $(h_1, h_2) \in \mathbb{R}^n \times \mathbb{R}^m \mapsto b(h_1, a_2) + b(a_1, h_2)$ est linéaire. Alors, pour conclure que $Df(a)(h) = b(h_1, a_2) + b(a_1, h_2)$, il suffit de montrer que $b(h_1, h_2) = o(h)$ où $h = (h_1, h_2)$. En effet, puisque b est bilinéaire de $\mathbb{R}^n \times \mathbb{R}^m$ dans $\mathbb{R}^p$, il existe $C > 0$ telle que

$$\|b(h_1, h_2)\|_{\mathbb{R}^p} \leq C\|h_1\|_{\mathbb{R}^n}\|h_2\|_{\mathbb{R}^m}, \quad \forall h = (h_1, h_2) \in \mathbb{R}^n \times \mathbb{R}^m.$$

Alors,

$$\|b(h_1, h_2)\|_{\mathbb{R}^p} \leq C\|h\|^2, \text{ avec } \|h\| = \max(\|h_1\|_{\mathbb{R}^n}, \|h_2\|_{\mathbb{R}^m}).$$

On remarque facilement que l'application $a \mapsto Db(a)$ est linéaire. D'autre part,

$$\begin{array}{rcl} \|Db(a)(h_1, h_2)\|_{\mathbb{R}^p} & = & \|b(h_1, a_2) + b(a_1, h_2)\|_{\mathbb{R}^p} \\ & \leq & C(\|h_1\|_{\mathbb{R}^n}, \|a_2\|_{\mathbb{R}^m}, + \|a_1\|_{\mathbb{R}^n}, \|h_2\|_{\mathbb{R}^m},) \\ & \leq & C(\|a_1\|_{\mathbb{R}^n}, + \|a_2\|_{\mathbb{R}^m},) \max(\|h_1\|_{\mathbb{R}^n}, , \|h_2\|_{\mathbb{R}^m},). \end{array}$$

Alors $\|Db(a)\| \leq 2C\|a\|$ où $C = \|b\|$. D'où, Db comme application linéaire en $a \in E_1 \times E_2$ est continue et $\|Db\| \leq 2C = 2\|b\|$. Donc, $D^2(b) = D(Db)$ est constante et $D^3(b) = 0$.

Différentielle d'une application composée

Théorème 2.1 *Soient Ω ouvert de $\mathbb{R}^n$ et $\mathcal{U}$ ouvert de $\mathbb{R}^m$, $f : \Omega \to \mathcal{U}$ et $g : \mathcal{U} \to \mathbb{R}^p$. On suppose f différentiable en $a \in \Omega$ et g différentiable en $f(a)$, alors $g \circ f$ est différentiable en a et*

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

Si de plus, f et g sont C^1, respectivement sur Ω et $\mathcal{U}$, alors $g \circ f$ est C^1 sur Ω.

Preuve : L'application $Dg(f(a)) \circ Df(a)$ est linéaire. On veut montrer que

$$g(f(a+h)) - g(f(a)) - Dg(f(a))[Df(a)(h)] = o(h).$$

Par hypothèse,

$$g(f(a+h)) - g(f(a)) = Dg(f(a))[f(a+h) - f(a)] + o_1(f(a+h) - f(a)),$$

avec $\lim_{h \to 0} \frac{o_1(h)}{\|h\|} = 0$. De plus,

$$f(a+h) - f(a) = Df(a)(h) + o_2(h), \quad \text{avec} \quad \lim_{h \to 0} \frac{o_2(h)}{\|h\|} = 0.$$

Donc,

$$g(f(a+h)) - g(f(a)) = Dg(f(a))[Df(a)(h) + o_2(h)] + o_1(f(a+h) - f(a)).$$

Alors, par la linéairité de $Dg(f(a))$, on obtient

$$g(f(a+h)) - g(f(a)) - Dg(f(a))[Df(a)(h)] = Dg(f(a))[o_2(h)] + o_1(f(a+h) - f(a)).$$

On a, par la continuité de $Dg(f(a))$,

$$\|Dg(f(a))[o_2(h)]\| \leq \|Dg(f(a))\| \|o_2(h)\|,$$

alors $Dg(f(a))[o_2(h)] = o(h)$.

Étant donné $\varepsilon > 0$, il existe $r > 0$ tel que

$$\|h\| \leq r \implies \|f(a+h) - f(a)\| = \|Df(a)(h) + o_2(h)\| \leq (\|Df(a)\| + \varepsilon)\|h\|.$$

Alors, pour $\|h\| \leq r$,

$$\frac{\|f(a+h) - f(a)\|}{\|h\|} \leq \|Df(a)\| + \varepsilon.$$

Par la continuité de f en a, $f(a+h) - f(a) \to 0$ si $h \to 0$. Donc $o_1(f(a+h) - f(a)) = o(h)$. D'où le résultat.

Supposons f, g de classe C^1, alors $x \mapsto (Dg(f(x)), Df(x))$ est continue. L'opérateur de composition $B = \mathcal{L}(\mathbb{R}^m, \mathbb{R}^p) \times \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ défini par $B(u, v) = u \circ v$ est bilinéaire continu. On peut écrire

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

qui est donc continue par composition.

Proposition 2.3 *Si f, g sont k fois différentiables (resp. C^k) alors $g \circ f$ est k fois différentiable (resp. C^k).*

Preuve : Par récurrence sur k, en utilisant la règle de la différentielle des fonctions composées.

2.3 Vitesse et dérivée directionnelle

Soit I un intervalle de $\mathbb{R}$ et une courbe dans $\mathbb{R}^p$, c'est-à-dire une application continue $c : I \to \mathbb{R}^p$.

Si c est différentiable en $t_0 \in I$, il existe $L : \mathbb{R} \to \mathbb{R}^p$ linéaire telle que

$$c(t_0 + t) - c(t_0) = L(t) + o(t) = tL(1) + o(t),$$

d'où

$$\lim_{t \to 0} \frac{c(t_0 + t) - c(t_0)}{t} = L(1) \in \mathbb{R}^p.$$

On notera $c'(t_0)$ cette limite et on l'appelera *dérivée* ou *vitesse* de c en t_0. Si $p = 1$, on retrouve le nombre dérivée usuel.

Réciproquement si

$$c'(t_0) = \lim_{t \to 0} \frac{c(t_0 + t) - c(t_0)}{t}$$

existe alors c est différentiable en t_0 et l'on a $Dc(t_0)(h) = hc'(t_0)$.

Pour les courbes on ne distingue pas $c'(t_0)$ et $Dc(t_0)$ identifiée à sa valeur en 1.

Exemple : Soient $u, x_0 \in \mathbb{R}^n$ fixés et $c(t) = x_0 + tu$. La vitesse de la courbe c est $c'(t) = u$. Si c est différentiable à valeurs dans l'ouvert $\Omega \subset \mathbb{R}^n$ et $f : \Omega \to \mathbb{R}^p$ est différentiable sur Ω, alors $f \circ c$ est différentiable et $(f \circ c)'(t) = Df(c(t))(c'(t))$.

On dit que la courbe est transportée par une fonction et que la vitesse est transportée par sa différentielle.

Définition 2.7 *On dit que f admet une dérivée en a dans la direction h, si la courbe $t \mapsto f(a+th)$ est dérivable en $t = 0$. La vitesse en 0 est appelée dérivée directionnelle en a dans la direction h et notée $f'(a;h)$ ou $\frac{\partial f}{\partial h}(a)$. On a donc*

$$f'(a;h) = \lim_{t \to 0} \frac{f(a+th) - f(a)}{t}.$$

Théorème 2.2 *Si f est différentiable en $a \in \Omega$ alors, pour tout $h \in \mathbb{R}^n$, $f'(a;h)$ existe et*

$$f'(a;h) = Df(a)(h).$$

Preuve : La courbe $t \mapsto c(t) = a + th$ est dérivable avec $c'(t) = h$ et $c(0) = a$. Par composition, $t \mapsto f(a+th)$ est dérivale en 0 et

$$f'(a;h) = (f \circ c)'(0) = Df(c(0))(c'(0)) = Df(a)(h).$$

La réciproque de ce théorème est fausse.

Exemple : Soit $A = \{(x,y) \in \mathbb{R}^2 : x > 0 \ \text{ et } \ 0 < y < x^2\}$ (ouvert de $\mathbb{R}^2$) et χ_A la fonction caractéristique de A. La dérivée directionnelle en $(0,0)$ est nulle dans toutes les directions, mais χ_A nest pas différentiable en $(0,0)$ car elle est discontinue en $(0,0)$.

2.4 Dérivées partielles

On note π_i la projection de $\mathbb{R}^n$ sur sa i-ième composante $\mathbb{R}$,

$$\pi_i(x_1, \ldots, x_n) = x_i, \quad 1 \leq i \leq n.$$

L'application π_i est linéaire continue donc, pour tout $x \in \mathbb{R}^n$, $D\pi_i(x) = \pi_i$. On note $c_i : \mathbb{R} \to \mathbb{R}^n$ l'injection définie par $c_i(x) = (0, \ldots, 0, x, 0, \ldots, 0)$ (où x est dans le i-ième rang). L'application c_i est linéaire et $D_{c_i}(x) = c_i$, pour tout $x \in \mathbb{R}$. On a

$$\forall 1 \leq i \leq n, \quad \pi_i \circ c_i = \mathrm{id}_{\mathbb{R}}, \quad \sum_{i=1}^n c_i \circ \pi_i = \mathrm{id}_{\mathbb{R}^n}.$$

Soit $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^p$ Pour $x \in \Omega$, on note $f(x) = (f_1(x), \ldots, f_p(x))$ où $f_i = \pi_i \circ f$, $1 \leq i \leq p$.

Théorème 2.3 *L'application f est différentiable en $a \in \Omega$ si et seulement si, pour tout $1 \leq i \leq p$, f_i est différentiable en a. Alors*

$$Df(a)(h) = (Df_1(a)(h), \ldots, Df_n(a)(h)), \quad h \in \mathbb{R}^n.$$

De plus, f est C^1 sur Ω si et seulement si, pour tout $1 \leq i \leq p$, f_i est C^1 sur Ω.

Preuve : Par la composition, si f est différentiable en $a \in \Omega$, pour tout $1 \leq i \leq p$, f_i l'est aussi (resp. C^1 si f l'est) et

$$Df_i(x) = \pi_i \circ Df(x), \quad \forall x \in \Omega,$$

c'est-à-dire, pour tout $h \in E$, $Df_i(x)(h)$ est la i-ième composante de $Df(x)(h)$.

Réciproquement si, pour tout $1 \leq i \leq p$, f_i est différentiable, comme

$$f = \sum_{i=1}^{p} c_i \circ \pi_i \circ f = \sum_{i=1}^{p} c_i \circ f_i,$$

alors f est différentiable (car somme de fonctions différentiables).

Soit Ω un ouvert de $\mathbb{R}^n$ et $f : \Omega \to \mathbb{R}$. On fixe $a = (a_1, \ldots, a_n) \in \Omega$.

Définition 2.8 *On dit que f admet la i-ième dérivée partielle en a si f admet la dérivée en a dans la direction e_i, où $e_1, \ldots, e_n$ désignent les éléments de la base canonique de $\mathbb{R}^n$. On note $\dfrac{\partial f}{\partial x_i}(a) = f'(a; e_i)$.*

Remarque : On a

$$\frac{\partial f}{\partial x_i}(a) = \lim_{h \to 0} \frac{f(a + he_i) - f(a)}{h}.$$

Théorème 2.4 *Si f est différentiable en $a \in \Omega$ alors, pour tout $1 \leq i \leq n$, $\frac{\partial f}{\partial x_i}(a)$ existe et l'on a*

$$Df(a)(h) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a) h_i, \quad \forall h = (h_1, \ldots, h_n) \in \mathbb{R}^n.$$

Alors,

$$Df(a)(h) =< \nabla f(a), h >$$

où $\nabla f(a) = (\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a))$ est le gradient de f au point a et $< \cdot, \cdot >$ désigne le produit scalaire de $\mathbb{R}^n$. De plus, si f est C^1, alors les applications $x \mapsto \frac{\partial f}{\partial x_i}(x)$ sont continues sur Ω.

Preuve : Par le théorème 2.2, les dérivées partielles premières de f existent. D'autre part, par la linéarité de $Df(a)$,

$$Df(a)(h) = \sum_{i=1}^{n} Df(a)(0, \ldots, h_i, \ldots, 0) = \sum_{i=1}^{n} h_i Df(a)(e_i) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a) h_i.$$

Si f est C^1, alors l'application $\Omega \to \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ définie par $x \mapsto Df(x)$ est continue. Alors, en composant avec l'opérateur de composition des opérateurs linéaires, on trouve que $x \mapsto \frac{\partial f}{\partial x_i}(x) = Df(x)(e_i)$ est continue sur Ω.

Remarque : Dans la section suivante, on montre que

$$f \text{ est } C^1 \iff \frac{\partial f}{\partial x_i} \text{ existent et son continues }, \ \forall 1 \leq i \leq n.$$

Exemple : La fonction $f(x, y) = \frac{xy}{x^2+y^2}$ si $(x, y) \neq 0$ et $f(0, 0) = 0$ admet des dérivées partielles sur $\mathbb{R}^2$ mais n'est pas différentiable (ni même continue) en $(0, 0)$ et, donc elle n'est pas de classe C^1 dans $\mathbb{R}^2$.

Soient Ω ouvert de $\mathbb{R}^n$ et $f : \Omega \to \mathbb{R}^p$ $(n, p \geq 1)$. On écrit $x = (x_1, \ldots, x_n)$ et $f(x) = (f_1(x), \ldots, f_p(x))$. On sait que f est différentiable sur Ω si et seulement si f_i est différentiable sur Ω, pour tout $1 \leq i \leq p$ et, dans ce cas, les $\frac{\partial f_i}{\partial x_j}$ existent, pour tout $1 \leq j \leq n$.

Dans les bases canoniques de $\mathbb{R}^n$ et $\mathbb{R}^p$, $Df(a)$ est répresenté par la matrice dans $M_{p\times n}(\mathbb{R})$ (matrice à p lignes et n colonnes) dont les coordonnées de la j-ième colonne sont les coordonnées du vecteur

$$Df(a)(e_j) = (Df_1(a)(e_j), \ldots, Df_p(e_j)),$$

pour $1 \leq j \leq n$, où $(e_1, \ldots, e_n)$ est la base canonique de $\mathbb{R}^n$. La i-ième coordonnée de ce vecteur est $Df_i(a)(e_j)$. Mais,

$$Df_i(a)(e_j) = \frac{\partial f_i}{\partial x_j}(a).$$

D'où la définition suivante :

Définition 2.9 *La matrice de $Df(a)$, appelée matrice jacobienne de f en a, s'écrit*

$$J(f)(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \ldots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & & \vdots \\ \frac{\partial f_p}{\partial x_1}(a) & \ldots & \frac{\partial f_p}{\partial x_n}(a) \end{pmatrix}.$$

De plus, pour tout $h \in \mathbb{R}^n$,

$$Df(a)(h) = J(f)(a)h, \quad \text{où } h \text{ s'identifie avec son vecteur colonne.}$$

La règle de la différentielle d'une application composée s'écrit par rapport aux matrices jacobiennes de la forme suivante.

Proposition 2.4 *Soient Ω ouvert de $\mathbb{R}^n$ et $\mathcal{U}$ ouvert de $\mathbb{R}^m$, $f : \Omega \to \mathcal{U}$ et $g : \mathcal{U} \to \mathbb{R}^p$. On suppose f différentiable en $a \in \Omega$ et g différentiable en $f(a)$, alors $g \circ f$ est différentiable en a et*

$$J(g \circ f)(a) = J(g)(f(a))J(f)(a).$$

2.5 Théorèmes de la moyenne

Cette section est consacré à la généralisation du théorème classique des accroissements finis.

Rappel : Soit f une fonction à valeurs réelles définie et continue sur un segment $[a, b]$ et dérivable dans $]a, b[$. Le théorème classique des accroissements finis dit qu'il existe $c \in]a, b[$ tel que

$$f(b) - f(a) = f'(c)(b - a).$$

Ce résultat ne peut pas être généralisé avec cet énoncé. Par exemple, si $f : \mathbb{R} \to \mathbb{C}$ est définie par $f(t) = e^{it}$, on a $f(2\pi) - f(0) = 0$ et, pour tout $c \in \mathbb{R}$, $f'(c) = ie^{it} \neq 0$.

Par contre, la conséquence du théorème des accroissements finis :

$$|f(b) - f(a)| \leq (b - a) \sup_{x \in]a,b[} |f'(x)|,$$

sera généralisée.

Théorème 2.5 (Théorème de la moyenne) *Soit $f : [a,b] \to \mathbb{R}^p$ et $g : [a,b] \to \mathbb{R}$. On suppose f et g continues sur $[a,b]$ et dérivables sur $]a,b[$ et telles que*

$$\forall x \in [a,b], \quad \|f'(x)\|_{\mathbb{R}^p} \leq g'(x).$$

Alors,

$$\forall x \in [a,b], \quad \|f(x) - f(a)\|_{\mathbb{R}^p} \leq g(x) - g(a).$$

Preuve : On va montrer que pour tout $\varepsilon > 0$,

$$\|f(x) - f(a)\| \leq g(x) - g(a) + \varepsilon(x-a) + \varepsilon. \tag{1}$$

Notons $A = \{y \in [a,b] : \forall x \in [a,y],\ x \text{ vérifie } (1)\}$. Puisque $a \in A$, $A \neq \emptyset$. D'autre part, si $y \in A$, par la définition $[a,y] \in A$.

Notons $c = \sup(A)$. On a $c \leq b$. Par la continuité de f et g, il existe $\delta > 0$ tel que

$$a \leq x \leq a + \delta \implies \|f(x) - f(a)\| < \frac{\varepsilon}{2} \text{ et } |g(x) - g(a)| < \frac{\varepsilon}{2},$$

alors :

$$\|f(x) - f(a)\| < \frac{\varepsilon}{2} \leq g(x) - g(a) + \varepsilon,$$

et donc, x vérifie (1). D'où $[a, a+\delta] \subset A$ et alors $c > a$. En faisant x tendre vers c par des valeurs inférieures, on obtient (par la continuité de f et g),

$$\|f(c) - f(a)\| \leq g(c) - g(a) + \varepsilon(c-a) + \varepsilon.$$

Donc $A = [a,c]$.

Supposons par contradiction que $c < b$. Alors, f et g sont dérivables en c et il existe $\gamma > 0$ tel que $c + \gamma < b$. Au voisinage de c, on a

$$f(x) - f(c) = f'(c)(x-c) + \varepsilon_1(x)(x-c) \text{ et } g(x) - g(c) = g'(c)(x-c) + \varepsilon_2(x)(x-c),$$

avec

$$\lim_{x\to c} \varepsilon_1(x) = 0_{\mathbb{R}^p} \quad \text{et} \quad \lim_{x\to c} \varepsilon_2(x) = 0.$$

Donc, on peut choisir γ suffisamment petit tel que

$$\|\varepsilon_1(x)\| < \frac{\varepsilon}{2} \quad \text{et} \quad |\varepsilon_2(x)| < \frac{\varepsilon}{2}.$$

Alors, pour tout $x \in [c, c+\gamma]$, on a

$$\|f(x) - f(c)\| \leq \|f'(c)\|(x-c) + \frac{\varepsilon}{2}(x-c) \text{ et } g'(c)(x-c) \leq g(x) - g(c) + \frac{\varepsilon}{2}(x-c).$$

En utilisant l'hypothèse, on obtient

$$\begin{aligned}
\|f(x) - f(c)\| &\leq \|f'(c)\|(x-c) + \frac{\varepsilon}{2}(x-c) \\
&\leq g'(c)(x-c) + \frac{\varepsilon}{2}(x-c) \\
&\leq g(x) - g(c) + \varepsilon(x-c).
\end{aligned}$$

Finalement, pour tout $x \in [c, c+\gamma]$, on a

$$\|f(x) - f(a)\| \leq \|f(x) - f(c)\| + \|f(c) - f(a)\| \leq g(x) - g(a) + \varepsilon(x-a) + \varepsilon.$$

Donc, $[c, c+\gamma] \subset A$, ce qui contradit la définition de c. D'où $c = b$.

Corollaire 2.1 *Soit $f : [a,b] \to \mathbb{R}^p$ continue sur $[a,b]$ et dérivable sur $]a,b[$. Supposons qu'il existe $M > 0$ tel que*

$$\forall x \in [a,b], \quad \|f'(x)\| \leq M.$$

Alors,

$$\|f(b) - f(a)\| \leq M(b-a).$$

Preuve : Il suffit de prendre $g(x) = Mx$ et d'appliquer le théorème de la moyenne.

Ensuite, on suppose que f est définie sur un ouvert Ω d'un espace vectoriel normé qui n'est plus necéssairement $\mathbb{R}$.

Si $a, b \in \mathbb{R}^n$, on définit le segment reliant a à b comme l'ensemble

$$[a,b] = \{(1-t)a + tb : t \in [0,1]\}.$$

Corollaire 2.2 *Soit Ω ouvert de $\mathbb{R}^n$ et $f : \Omega \to \mathbb{R}^p$. On suppose $a, b \in \Omega$ tels que $[a,b] \subset \Omega$. Alors,*

$$\|f(b) - f(a)\|_{\mathbb{R}^p} \leq \|b-a\|_{\mathbb{R}^n} \sup_{x \in [a,b]} \|Df(x)\|.$$

On l'appelle l'inégalité de la moyenne.

Preuve : On peut supposer que $M = \sup_{x\in[a,b]} \|Df(x)\| < +\infty$ (cas contraire le membre de droite de l'inégalité vaut $+\infty$ et l'inégalité est évidente). On définit $c : [0,1] \to [a,b] \subset \Omega$ par

$$c(t) = (1-t)a + tb.$$

Alors, l'application $f \circ c : [0,1] \to \mathbb{R}^p$ est différentiable sur $[0,1]$ et

$$(f \circ c)'(t) = Df(c(t))(c'(t)) = Df(c(t))(b-a).$$

De plus,

$$\|(f \circ c)'(t)\|_{\mathbb{R}^p} \leq \sup_{x\in[a,b]} \|Df(x)(b-a)\|_{\mathbb{R}^p} \leq \|b-a\|_{\mathbb{R}^n} \sup_{x\in[a,b]} \|Df(x)\| = M\|b-a\|_{\mathbb{R}^n}.$$

D'où, par le corollaire 2.1,

$$\|f(b) - f(a)\|_{\mathbb{R}^p} = \|(f \circ c)(1) - (f \circ c)(0)\|_{\mathbb{R}^p} \leq M\|b-a\|_{\mathbb{R}^n}.$$

Définition 2.10 *On dit qu'un sous-ensemble Ω d'un espace vectoriel normé E est convexe si, pour tous $a, b \in \Omega$, le segment $[a,b]$ est dans Ω.*

Corollaire 2.3 (Inégalité des accroissements finis pour les convexes) *Soit Ω ouvert convexe de $\mathbb{R}^n$ et $f : \Omega \to \mathbb{R}^p$ différentiable. Si $\|Df(x)\| \leq M$, pour tout $x \in \Omega$, alors*

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

Pour les applications différentiables, être lipschitzienne est équivalent à avoir sa différentielle bornée.

Proposition 2.5 *Soit f une application différentiable sur Ω ouvert convexe. Alors,*

$$f \text{ est lipschitzienne} \iff Df \text{ est borné sur } \Omega.$$

Preuve : La condition suffisante est immédiate d'après le corollaire 2.3.

Condition nécessaire : supposons

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

Pour $y = x + th$, avec $t > 0$ et $h \in \mathbb{R}^n$, on a

$$\frac{\|f(x + th) - f(x)\|}{t} \leq M\frac{t\|h\|}{t} = M\|h\|.$$

Donc,

$$\|Df(x)(h)\| = \lim_{t \to 0^+} \frac{\|f(x + th) - f(x)\|}{t} \leq M\|h\|.$$

D'où, $\sup_{x \in \Omega} \|Df(x)\| \leq M$.

Conséquence : Si Df s'annule sur un ouvert convexe $\mathcal{U}$ on a $\|f(x) - f(y)\| = 0$, pour tous $x, y \in \mathcal{U}$, donc f est constante sur $\mathcal{U}$. Les boules ouvertes sont convexes donc, si $Df = 0$ sur un ouvert Ω, alors f est constante sur une boule au voisinage de tout point de Ω. Plus précisemment, f est constante sur les composantes connexes de Ω.

Théorème 2.6 *Soient Ω ouvert connexe de $\mathbb{R}^n$ et $f : \Omega \to \mathbb{R}^p$ différentiable. Si $Df(x) = 0$ pour tout $x \in \Omega$, alors f est constante sur Ω.*

Preuve : On fixe $a \in \Omega$. Soit $A = \{x \in \Omega : f(x) = f(a)\}$. Par la continuité de f, A est un fermé de Ω. Pour $x \in A$, $x \in \Omega$ alors il existe $r > 0$ tel que $B = B(x, r) \subset \Omega$. L'ensemble B est convexe, alors par le corollaire 2.3, f est constante sur B, c'est-à-dire

$$f(y) = f(x) = f(a), \quad \forall y \in B.$$

Donc, $B \subset A$. On conclut que A est une ensemble ouvert de Ω. Comme $A \neq \emptyset$ et Ω est connexe, $A = \Omega$.

2.5.1 Approximation linéaire et quadratique

Théorème 2.7 *Soient Ω ouvert de $\mathbb{R}^n$, $f : \Omega \to \mathbb{R}^p$ différentiable et $a, b \in \Omega$ tels que $[a, b] \in \Omega$. Alors,*

$$\|f(b) - f(a) - Df(a)(b - a)\| \leq \|b - a\| \sup_{x \in [a,b]} \|Df(x) - Df(a)\|.$$

En particulier, pour tout h tel que $[a, a + h] \in \Omega$,

$$\|f(a + h) - f(a) - Df(a)(h)\| \leq \|h\| \sup_{x \in [a,a+h]} \|Df(x) - Df(a)\|.$$

Preuve : On applique le corollaire 2.2 à $\varphi(x) = f(x) - Df(a)(x)$.

Corollaire 2.4 *Si f est deux fois différentiable, alors*

$$\|f(b) - f(a) - Df(a)(b - a)\| \leq \|b - a\|^2 \sup_{x \in [a,b]} \|D^2 f(x)\|.$$

Preuve : On peut appliquer l'inégalité de la moyenne à Df sur le segment $[a, x]$,

$$\|Df(x) - Df(a)\| \leq \|x - a\| \sup_{y \in [a,x]} \|D^2 f(y)\|,$$

d'où

$$\sup_{x \in [a,b]} \|Df(x) - Df(a)\| \leq \|b - a\| \sup_{y \in [a,b]} \|D^2 f(y)\|.$$

Par le théorème 2.7, on obtient le résultat.

2.5.2 Différentiabilité et dérivées partielles

Théorème 2.8 *Soit Ω ouvert de $\mathbb{R}^n$. Alors $f : \Omega \to \mathbb{R}^p$ est de classe C^1 si et seulement si, pour tout $1 \leq i \leq n$, les dérivées partielles $\frac{\partial f}{\partial x_i}$ existent et sont continues.*

Preuve : La condition nécessaire a été montré dans le théorème 2.4.

Condition suffisante : Soit $a \in \Omega$. Si f est différentiable en a, on sait que

$$Df(a)(h) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a) h_i, \quad h = (h_1, \ldots, h_n) \in \mathbb{R}^n.$$

Donc, f est de classe C^1 en a, si on montre que $Df(a)$ existe. On définit $g : \Omega \to \mathbb{R}^p$ par

$$g(x) = f(x) - f(a) - \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a)(x_i - a_i).$$

Alors, $\frac{\partial g}{\partial x_j}(x) = \frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(a)$. Soit $\varepsilon > 0$ fixé. Puisque les dérivées partielles de f sont continues en a, il existe $r > 0$ tel que, pour tout $1 \leq j \leq n$,

$$\left\| \frac{\partial g}{\partial x_j}(x) \right\| = \left\| \frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(a) \right\| < \varepsilon, \quad \forall x \in B(a, r).$$

D'autre part,

$$g(x) = g(x) - g(a) = \sum_{k=1}^{n} \left(g(x_1, \ldots, x_k, a_{k+1}, \ldots, a_n) - g(x_1, \ldots, x_{k-1}, a_k, \ldots, a_n) \right).$$

Alors, par le théorème des accroissements finis pour les convexes, pour tout $x \in B(a, r)$,

$$\begin{aligned} \|g(x)\| &\leq \sum_{k=1}^{n} \|g(x_1, \ldots, x_{k-1}, x_k, a_{k+1}, \ldots, a_n) - g(x_1, \ldots, x_{k-1}, a_k, a_{k+1}, \ldots, a_n)\| \\ &\leq \sum_{k=1}^{n} \varepsilon |x_k - a_k| \leq n\varepsilon \|x - a\|, \text{ où } \|x - a\| = \max_{1 \leq k \leq n} |x_k - a_k|. \end{aligned}$$

Comme ε est arbitraire, on conclut que $g(x) = o(x - a)$. Donc,

$$f(x) = f(a) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a)(x_i - a_i) + o(x - a).$$

Finalement, $Df(a)$ existe et, pour tout $h \in \mathbb{R}^n$,

$$Df(a)(h) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a) h_i.$$

3 Différentielles d'ordre supérieur

Ce chapitre est consacré aux différentielles d'ordre supérieur d'une application différentiable et aux régles de calcul les concernant. Les résultats principaux sont : le théorème de Schwarz sur la symétrie des différentielles d'ordre supérieur, la formule de Taylor et les applications à l'étude des points critiques pour les fonctions à valeurs réelles.

3.1 Théorème de Schwarz

On commence par un résultat préliminaire algébrique :

Proposition 3.1 *Soient E, F, G trois e.v.n.. On note $\mathcal{L}(E,F;G)$ l'espace des applications bilinéaires continues de $E\times F$ dans G. Alors, $\mathcal{L}(E,\mathcal{L}(F,G))$ est canoniquement isométrique à $\mathcal{L}(E,F;G)$.*

Preuve : Pour $L\in\mathcal{L}(E,\mathcal{L}(F,G))$, on définit $\Phi(L)\in\mathcal{L}(E,F;G)$ par $\Phi(L)(x,y)=L(x)(y)$ et, pour $B\in\mathcal{L}(E,F;G)$ on définit $\Psi(B)\in\mathcal{L}(E,\mathcal{L}(F,G))$ par $\Psi(B)(x)(y)=B(x,y)$. On montre facilement que Φ et Ψ sont deux isomorphismes (applications linéaires bijectives) réciproques l'une de l'autre.

Par la continuité de $L\in\mathcal{L}(E,\mathcal{L}(F,G))$, on a

$$\|\Phi(L)(x,y)\|=\|L(x)(y)\|\leq\|L(x)\|\|y\|\leq\|L\|\|x\|\|y\|,\quad\forall(x,y)\in E\times F.$$

Alors, $\|\Phi(L)\|\leq\|L\|$. Donc, Φ est continue et $\|\Phi\|\leq 1$. D'autre part, par la continuité de $B\in\mathcal{L}(E,F;G)$, on a, pour tout $x\in E$,

$$\|\Psi(B)(x)(y)\|=\|B(x,y)\|\leq\|B\|\|x\|\|y\|,\quad\forall y\in F\iff\|\Psi(B)(x)\|\leq\|B\|\|x\|.$$

Alors, $\|\Psi(B)\|\leq\|B\|$. Donc, Ψ est continue et $\|\Psi\|\leq 1$. Finalement, on obtient

$$\|L\|=\|\Psi\circ\Phi(L)\|\leq\|\Phi(L)\|\leq\|L\|.$$

D'où, $\|\Phi(L)\|=\|L\|$, c'est-à-dire Φ est une isométrie.

Ce résultat est facilement généralisable. Un application linéaire continue :

$$L\in\mathcal{L}(E_1,\mathcal{L}(E_2,\ldots,\mathcal{L}(E_n,F))\ldots)$$

s'identifie à une application n-linéaire continue de $E_1\times\ldots\times E_n$ dans F, en écrivant :

$$L(h_1,\ldots,h_n)=[L(h_1)(h_2)\ldots](h_n).$$

Soit Ω ouvert de $\mathbb{R}^n$ et $f:\Omega\to\mathbb{R}^p$ une application m_0 fois différentiable sur Ω ($m_0\geq 2$). En utilisant l'identification faite dans la proposition 3.1, la différentielle seconde de f en a est alors une application bilinéaire continue de $\mathbb{R}^n\times\mathbb{R}^n$ dans $\mathbb{R}^p$, définie par :

$$D^2f(a)(h,k)=D^2f(a)(h)(k),\quad h,k\in\mathbb{R}^n.$$

De plus, la différentielle seconde de f est une application :

$$D^2f:\Omega\to\mathcal{L}(\mathbb{R}^n,\mathbb{R}^n;\mathbb{R}^p).$$

Plus généralement, les différentielles d'ordre $m \leq m_0$ de f au point a sont des applications m-linéaires continues :

$$D^m f(a) : \mathbb{R}^n \times \ldots \mathbb{R}^n \to \mathbb{R}^p.$$

Le théorème suivant montre que les dérivées secondes sont des applications bilinéaires symétriques.

Théorème 3.1 (Théorème de Schwarz) *Soit Ω ouvert de $\mathbb{R}^n$ et $f : \Omega \to \mathbb{R}^p$ deux fois différentiable. Pour tout $a \in \Omega$, l'application bilinéaire $D^2 f(a)$ est symétrique :*

$$D^2 f(a)(h,k) = D^2 f(a)(k,h), \quad \forall h, k \in \mathbb{R}^n.$$

Preuve : On va approcher $D^2 f(a)(h,k)$ par une expression symétrique en (h,k).

Soient $h, k \in \mathbb{R}^n$ de norme suffisamment petite tels que, pour tout $t \in [0,1]$, $a+th+k \in \Omega$. On pose

$$g(t) = f(a+th+k) - f(a+th), \quad t \in [0,1].$$

Alors,

$$g(1) - g(0) = f(a+h+k) - f(a+h) - f(a+k) + f(a)$$

est une espression symétrique en (h,k) et

$$g'(t) = [Df(a+th+k) - Df(a+th)](h).$$

Par le théorème de la moyenne, on a

$$\begin{aligned} \|g(1) - g(0) - D^2 f(a)(k)(h)\| &\leq \|g(1) - g(0) - g'(0)\| + \|g'(0) - D^2 f(a)(k)(h)\| \\ &\leq \sup_{t \in [0,1]} \|g'(t) - g'(0)\| + \|g'(0) - D^2 f(a)(k)(h)\| \end{aligned} \qquad (2)$$

Soit $\varepsilon > 0$ fixé. Par la définition de la différentielle, il existe $r > 0$ tel que pour $\|(h,k)\| < r$ on a

$$\|[Df(a+th+k) - Df(a) - D^2 f(a)(th+k)](h)\| \leq \varepsilon \|th+k\| \|h\| \leq \varepsilon \|h\| (\|h\| + \|k\|)$$

et

$$\|[Df(a+th) - Df(a) - D^2 f(a)(th)](h)\| \leq \varepsilon \|th\| \|h\| \leq \varepsilon \|h\|^2.$$

Alors, en ajoutant ces inégalités et en utilisant la linéarité de $D^2 f(a)$, on obtient

$$\|[Df(a+th+k) - Df(a+th) - D^2 f(a)(k)](h)\| \leq 2\varepsilon \|h\| (\|h\| + \|k\|),$$

c'est-à-dire

$$\|g'(t) - D^2 f(a)(k,h)\| \leq 2\varepsilon \|h\| (\|h\| + \|k\|).$$

En particulier, pour $t = 0$, on a

$$\|g'(0) - D^2 f(a)(k,h)\| \leq 2\varepsilon \|h\| (\|h\| + \|k\|).$$

D'où,

$$\|g'(t) - g'(0)\| \leq 4\varepsilon \|h\| (\|h\| + \|k\|), \forall t \in [0,1] \implies \sup_{t \in [0,1]} \|g'(t) - g'(0)\| \leq 4\varepsilon \|h\| (\|h\| + \|k\|).$$

D'après (2), on a

$$\|g(1) - g(0) - D^2 f(a)(k,h)\| \le 6\varepsilon\|h\|(\|h\| + \|k\|).$$

En permutant h et k et en utilisant la symétrie de $g(1) - g(0)$ par rapport (h,k), on a aussi

$$\|g(1) - g(0) - D^2 f(a)(h,k)\| \le 6\varepsilon\|k\|(\|h\| + \|k\|).$$

D'où :

$$\|D^2 f(a)(k,h) - D^2 f(a)(h,k)\| \le 6\varepsilon(\|h\| + \|k\|)^2. \tag{3}$$

Cette inégalité étant valable pour $\|(h,k)\| < r$. Soit $(h,k) \in \mathbb{R}^n \times \mathbb{R}^n$ quelconque. On prend $\lambda > 0$ tel que $\|(\lambda h, \lambda k)\| < r$. Alors,

$$\|D^2 f(a)(\lambda k, \lambda h) - D^2 f(a)(\lambda h, \lambda k)\| \le 6\varepsilon(\|\lambda h\| + \|\lambda k\|)^2$$

$$\iff \lambda^2\|D^2 f(a)(k,h) - D^2 f(a)(h,k)\| \le \lambda^2 6\varepsilon(\|h\| + \|k\|)^2.$$

Donc, l'inégalité (3) reste valable pour tout (h,k) et, en particulier, pour $\|h\| \le 1$ et $\|k\| \le 1$, on a

$$\|D^2 f(a)(k,h) - D^2 f(a)(h,k)\| \le 24\varepsilon.$$

Ainsi l'application bilinéaire $(h,k) \mapsto D^2 f(a)(k,h) - D^2 f(a)(h,k)$ est de norme inférieure à 24ε, pour tout $\varepsilon > 0$, donc nulle.

Si f est une fonction définie sur un ouvert Ω de $\mathbb{R}^n$ et m fois différentiable, $m \ge 2$, alors les $\frac{\partial f}{\partial x_i}$ admettent des différentielles partielles que l'on note $\dfrac{\partial^2 f}{dx_j dx_i}$ et ainsi de suite.

L'expression de la différentielle comme la somme des dérivées partielles se généralise au second ordre de la façon suivante.

Proposition 3.2 *Soit Ω un ouvert de $\mathbb{R}^n$ et $f : \Omega \to \mathbb{R}$ deux fois différentiable. Pour tout $a \in \Omega$ et tous $(h_1, \ldots, h_n), (k_1, \ldots, k_n) \in \mathbb{R}^n$, on a*

$$D^2 f(a)[(h_1, \ldots, h_n), (k_1, \ldots, k_n)] = \sum_{i,j=1}^{n} \frac{\partial^2 f}{dx_i dx_j}(a) h_i k_j.$$

Preuve : On a :

$$\begin{aligned}
D^2 f(a)[(h_1, \ldots, h_n), (k_1, \ldots, k_n)] &= D\left(x \mapsto Df(x)(k_1, \ldots, k_n)\right)(a)(h_1, \ldots, h_n) \\
&= \sum_{i=1}^{n} \frac{\partial\left(x \mapsto Df(x)(k_1, \ldots, k_n)\right)}{\partial x_i}(a) h_i \\
&= \sum_{i=1}^{n} \frac{\partial\left(x \mapsto \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x) k_j\right)}{\partial x_i}(a) h_i \\
&= \sum_{i,j=1}^{n} \frac{\partial\left(x \mapsto \frac{\partial f}{\partial x_j}(x)\right)}{\partial x_i}(a) k_j h_i \\
&= \sum_{i,j=1}^{n} \frac{\partial^2 f}{dx_i dx_j}(a) h_i k_j.
\end{aligned}$$

Le corollaire suivant est conséquence du théorème de Schwarz et de la proposition précédente.

Corollaire 3.1 *Si f est deux fois différentiable, on peut permuter l'ordre des dérivées partielles : pour tout $x \in \Omega$, pour tous $1 \leq i, j \leq n$,*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x).$$

Alors, si les $\frac{\partial^2 f}{\partial x_i \partial x_j}$ existent et sont continues, on peut permuter l'ordre des dérivations.

En effet, la continuité des dérivées partielles secondes entraîne le caractère C^2 de f et donc la symétrie. Alors que la simple existence des $\frac{\partial^2 f}{\partial x_i \partial x_j}$ ne permet pas de conclure que f est deux fois différentiable.

Si $f : \Omega \to \mathbb{R}$ est deux fois différentiable sur Ω ouvert de $\mathbb{R}^n$ et $a \in \Omega$, $D^2 f(a)$ est une forme bilinéaire symétrique qui est matriciellement représentée par les valeurs prises sur les couples de vecteurs (e_i, e_j) de la base canonique.

Définition 3.1 *La matrice symétrique de $D^2 f(a)$, appelée hessienne de f au point a, s'écrit :*

$$H(f)(a) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(a) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) \\ \vdots & \frac{\partial^2 f}{\partial x_i \partial x_j}(a) & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(a) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n}(a) \end{pmatrix}.$$

Pour tous $h, k \in \mathbb{R}^n$, on a

$$D^2 f(a)(h, k) = {}^t h H(f)(a) k,$$

où h et k sont identifiés avec ces vecteurs colonnes.

3.2 Formule de Taylor

Dans cette formule interviennent les termes successifs

$$Df(a)(h), D^2 f(a)(h, h), D^3 f(a)(h, h, h), \dots$$

On introduit alors la notation abrégée : pour $h \in \mathbb{R}^n$, on note

$$(h)^m = (h, h, \dots, h) \in (\mathbb{R}^n)^m, \quad m \geq 1.$$

Lemme 3.1 *Soit $\varphi : (\mathbb{R}^n)^m \to \mathbb{R}^p$ un application m-linéaire continue symétrique et $\Phi : \mathbb{R}^n \to \mathbb{R}^p$ définie par $\Phi(h) = \varphi(h, h, \dots, h) = \varphi((h)^m)$. On a*

$$D\Phi(h)(k) = m\varphi((h)^{m-1}, k).$$

Preuve : Par la différentielle de la composée de φ et de l'application linéaire $h \in \mathbb{R}^n \mapsto (h)^m \in (\mathbb{R}^n)^m$ et, en utilisant la symétrie de φ, on obtient

$$D\Phi(h)(k) = \varphi(k, h, \dots, h) + \varphi(h, k, h, \dots, h) + \dots + \varphi(h, \dots, h, k) = m\varphi((h)^{m-1}, k).$$

Théorème 3.2 (Formule de Taylor-Young) *Soit $f : \Omega \to \mathbb{R}$ une fonction m fois différentiable au voisinage de $a \in \Omega$. Alors, pour h suffisamment petit, on a :*

$$f(a+h) = f(a) + Df(a)(h) + \ldots + \frac{1}{m!} D^m f(a)((h)^m) + o(\|h\|^m).$$

Preuve : Pour $m = 1$, la formule se réduit à la définition de la différentielle. On raisonne alors par récurrence sur $m \geq 1$. On suppose la relaction vraie jusqu'à $m-1$. On pose, pour h suffisamment petit,

$$\varphi(h) = f(a+h) - f(a) - Df(a)(h) - \frac{1}{2} D^2 f(a)(h,h) - \ldots - \frac{1}{m!} D^m f(a)((h)^m).$$

On calcule la différentielle de φ en utilisant le lemme précédent : pour h suffisamment petit et $k \in \mathbb{R}^n$,

$$D\varphi(h)(k) = Df(a+h)(k) - Df(a)(k) - D^2 f(a)(h,k) - \ldots - \frac{1}{(m-1)!} D^m f(a)((h)^{(m-1)}, k).$$

Par l'hypothèse de récurrence appliquée à Df, on a $D\varphi(h) = o(\|h\|^{m-1})$. Alors, pour $\varepsilon > 0$ fixé, il existe $\delta > 0$ tel que

$$\|h\| < \delta \implies \|D\varphi(h)\| \leq \varepsilon \|h\|^{m-1}.$$

Par le théorème de la moyenne, on a

$$\|\varphi(h)\| = \|\varphi(h) - \varphi(0)\| \leq \varepsilon \|h\|^{m-1} \|h\| = \varepsilon \|h\|^m, \quad \forall h \in \mathbb{R}^n, \quad \|h\| < \delta.$$

Donc, $\varphi(h) = o(\|h\|^m)$. D'où, la formule de Taylor-Young d'ordre m.

La formule de Taylor-Young précise seulement le comportement d'une expression lorsque h tend vers 0, elle est donc purement locale au voisinage de 0.

Exemple : On considère $f : \mathbb{R}^2 \to \mathbb{R}$ définie par $f(x,y) = x\sin(y)$ et $(a,b) \in \mathbb{R}^2$. On a :

$$Df(a,b)(h,k) = \frac{\partial f}{\partial x}(a,b)h + \frac{\partial f}{\partial y}(a,b)k = \sin(b)h + a\cos(b)k,$$

$$\begin{aligned} D^2 f(a,b)[(h,k),(h,k)] &= \frac{\partial^2 f}{\partial x^2}(a,b)h^2 + \frac{\partial^2 f}{\partial x \partial y}(a,b)hk + \frac{\partial^2 f}{\partial y \partial x}(a,b)kh + \frac{\partial^2 f}{\partial y^2}(a,b)k^2 \\ &= \frac{\partial^2 f}{\partial x^2}(a,b)h^2 + 2\frac{\partial^2 f}{\partial x \partial y}(a,b)hk + \frac{\partial^2 f}{\partial y^2}(a,b)k^2 \\ &= 2\cos(b)hk - a\sin(b)k^2. \end{aligned}$$

Alors, la formule de Taylor à l'ordre 2 de f en (a,b) s'écrit :

$$(a+h)\sin(b+k) - a\sin(b) = \sin(b)h + a\cos(b)k + \frac{1}{2}(2\cos(b)hk - a\sin(b)k^2) + o(\|(h,k)\|^2).$$

Théorème 3.3 (Formule de Taylor avec reste intégral) *Soit Ω un ouvert de $\mathbb{R}^n$ et $f : \Omega \to \mathbb{R}$ une fonction de classe C^{m+1}. Si $a \in \Omega$ et $h \in \mathbb{R}^n$ tels que le segment $[a, a+h] \subset \Omega$ alors,*

$$f(a+h)-f(a) = Df(a)(h)+\ldots+\frac{1}{m!}D^m f(a)((h)^m)+\int_0^1 \frac{(1-t)^m}{m!}D^{m+1}f(a+th)((h)^{m+1})\,dt.$$

En particulier, pour tout compact convexe K contenu dans Ω, il existe $C > 0$ tel que, pour $a \in K$, $a+h \in K$, on a la **Formule de Taylor-Lagrange** *d'ordre m :*

$$\|f(a+h) - f(a) - Df(a)(h) - \ldots \frac{1}{m!}D^m f(a)((h)^m)\| \leq C\|h\|^{m+1}.$$

La formule de Taylor avec reste intégrale est la plus longue à écrire mais aussi la plus précise.

3.3 Extrema simples

On rappelle quelques définitions.

Définition 3.2 *Soient X une partie d'un espace normé E, $a \in X$ et $f : X \to \mathbb{R}$. On dit que a est un maximum relatif ou local (respectivement, minimum relatif) de f, s'il existe un voisinage $\mathcal{U}$ de a tel que*

$$\forall x \in \mathcal{U} \cap X, \quad f(x) \leq f(a) \quad (\textit{respectivement } f(a) \leq f(x)).$$

Si $\mathcal{U} \cap X = X$, le point est alors dit maximum ou minimum absolu ou global.

On dit que a est un maximum ou minimum strict si les inégalités précédentes sont strictes pour $x \neq a$. Un point qui est maximum ou minimum est un extremum.

Soit $f : \Omega \to \mathbb{R}^p$ une fonction différentiable sur Ω ouvert de $\mathbb{R}^n$.

Définition 3.3 *On dit que $a \in \Omega$ est un point critique de f si le rang de $Df(a)$ est inférieur à p. Dans ce cas, on dit que $f(a)$ est une valeur critique de f.*

Remarquer que

$$rang Df(a) = dim Df(a)(E).$$

Alors, le rang de $Df(a)$ n'est pas maximum si et seulement si $Df(a)$ n'est pas surjective. Si $p = 1$, le rang de $Df(a)$ est au plus un. Dans ce cas, a est critique si et seulement si $Df(a) = 0$.

Proposition 3.3 *Soit $f : \Omega \to \mathbb{R}$ différentiable. Si $a \in \Omega$ est un extremum de f, alors a est un point critique de f.*

Preuve : Supposons que $a \in \Omega$ est un extremum de f. Soit $h \in \mathbb{R}^n$ fixé. On considère $\varphi : \mathbb{R} \to \mathbb{R}^n$ définie par $\varphi(t) = a + th$. On a $\varphi(0) = a$ et $\varphi'(t) = h$. La fonction $g = f \circ \varphi$ à valeurs réelles est définie et différentiable sur l'ouvert de $\mathbb{R}$: $U = \{t \in \mathbb{R} : \varphi(t) \in \Omega\}$. Alors, le point $t = 0$ est un extremum de g et donc $g'(0) = Df(a)(h) = 0$. Comme h est arbitraire dans $\mathbb{R}^n$, on a $Df(a) = 0$.

La récriproque de cette proposition est évidement fausse. Un point peut être critique sans être un extremum. Exemple : si $f : \mathbb{R} \to \mathbb{R}$ est définie par $f(t) = t^3$, alors $Df(0) = 0$ mais 0 n'est pas un extremum relatif.

Proposition 3.4 *Soit $f : \Omega \to \mathbb{R}$ deux fois différentiable en $a \in \Omega$.*

1. *Si a est un minimum relatif de f, alors $Df(a) = 0$ et $D^2f(a)$ est une forme quadratique positive, c'est-à-dire*
$$\forall h \in \mathbb{R}^n, \quad D^2f(a)(h,h) \geq 0.$$
2. *Si a est un maximum relatif de f, alors $Df(a) = 0$ et $D^2f(a)$ est une forme quadratique positive, c'est-à-dire*
$$\forall h \in \mathbb{R}^n, \quad D^2f(a)(h,h) \leq 0.$$

Preuve : Comme dans la proposition précédente, on raisonne sur la fonction $g(t) = f(a+th)$.

Dans ce qui suit, on supppose $f : \Omega \to \mathbb{R}$ au moins de classe C^2. Soit $a \in \Omega$ un point critique de f. Au voisinage de a, la formule de Taylor-Young s'écrit :

$$\begin{aligned} f(x) - f(a) &= Df(a)(x-a) + \frac{1}{2}D^2f(a)(x-a, x-a) + \|x-a\|^2\varepsilon(x-a) \\ &= \frac{1}{2}D^2f(a)(x-a,x-a) + \|x-a\|^2\varepsilon(x-a). \end{aligned}$$

On donne des conditions suffisantes pour que a soit un extremum de f. On note par Q la forme quadratique définie par la forme bilinéaire symétrique $D^2f(a)$. Alors, pour $h \in \mathbb{R}^n$,

$$Q(h) = D^2f(a)(h,h).$$

Proposition 3.5 *Soit Ω un ouvert de $\mathbb{R}^n$, $f : \Omega \to \mathbb{R}$ de classe C^2 et $a \in \Omega$ un point critique de f.*

1. *S'il existe $\alpha > 0$ tel que*
$$\forall h \in \mathbb{R}^n, \quad Q(h) \geq \alpha\|h\|^2, \tag{4}$$
alors a est un minimum relatif strict de f.
2. *S'il existe $\beta > 0$ tel que*
$$\forall h \in \mathbb{R}^n, \quad Q(h) \leq -\beta\|h\|^2, \tag{5}$$
alors a est un maximum relatif strict de f.

Preuve :

1. Par la formule de Taylor-Young, pour $h = x - a$, on a
$$f(x) - f(a) = \frac{1}{2}Q(x-a) + \|x-a\|^2\varepsilon(x-a) \geq (\frac{\alpha}{2} + \varepsilon(x-a))\|x-a\|^2.$$
Ainsi, pour x suffisamment proche de a et $x \neq a$, le signe de α prédomine et l'on a $f(x) > f(a)$.
2. On raisonne de façon analogue. Alors, pour tout x suffisamment proche de a et $x \neq a$, on a $f(x) < f(a)$.

La forme quadratique Q associée à $D^2f(a)$ est alors caractérisée par la matrice hessienne de f en a :

$$H(f)(a) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(a) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) \\ \vdots & \frac{\partial^2 f}{\partial x_i \partial x_j}(a) & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(a) & \dots & \frac{\partial^2 f}{\partial x_n^2}(a) \end{pmatrix}.$$

avec

$$Q(h) = {}^t hH(f)(a)h = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(a)h_i h_j, \quad \text{où} \quad h = (h_1, \dots, h_n) \in \mathbb{R}^n.$$

La matrice $H(f)(a)$ est symétrique et donc orthogonalement diagonalisable. Deux cas se présentent :

1. Il n'y a pas de valeur propre nulle : Q est *régulière* et le point critique a est dit *non-dégénéré* ou *régulier*.
2. Il y a au moins une valeur propre nulle : Q est *non-régulière* et le point critique est dit *dégénéré* ou *singulier*. Il faut alors examiner les différentielles d'ordre supérieur pour décider de la nature de a.

Proposition 3.6 *Soit Ω ouvert de $\mathbb{R}^n$, $f : \Omega \to \mathbb{R}$ de classe C^2 et $a \in \Omega$ un point critique régulier. Soient $\lambda_1, \dots, \lambda_n$ les valeurs propres de $H(f)(a)$.*

1. *Si, pour tout $1 \leq i \leq n$, $\lambda_i > 0$, alors a est un minimum relatif strict de f.*
2. *Si, pour tout $1 \leq i \leq n$, $\lambda_i < 0$, alors a est un maximum relatif strict de f.*
3. *S'il existe $i, j \in \{1, \dots, n\}$ tels que $\lambda_i \lambda_j < 0$, alors a n'est pas un extremum. On dit que a est un point selle ou col.*

Preuve :

1. Q est une forme définie positive, c'est-à-dire
$$\forall h \in \mathbb{R}^n, \quad h \neq 0, \quad Q(h) > 0.$$
En particulier, Q est strictement positive sur la sphère unité qui est compacte dans $\mathbb{R}^n$. Alors, la forme quadratique Q (qui est continue sur $\mathbb{R}^n$) y atteint le minimum par une valeur $\alpha > 0$. Donc, on a (4) et on conclut par la proposition 3.5.
2. Q est une forme définie négative, c'est-à-dire
$$\forall h \in \mathbb{R}^n, \quad h \neq 0, \quad Q(h) < 0.$$
Ensuite, on raisonne de façon analogue pour montrer qu'il existe $\beta > 0$ tel que (5) est vérifiée et on applique la proposition 3.5 pour conclure.
3. Il existe des directions h telles que $Q(h)$ soit positif, et d'autres pour lequel il est négatif. Donc, a n'est pas un extremum.

Puisque le signe de la forme quadratique $Q(h) = D^2f(x)(h,h)$ dépend du signe des mineurs principaux de la matrice hessienne $H(f)(x)$, la proposition précédente peut-être énoncée d'une façon équivalente comme suit. On rappelle d'abord que, pour une matrice A carrée d'ordre n, la *sous-matrice principale* d'ordre k de A est la matrice extraite de A en éliminant les $n-k$ dernières lignes et les $n-k$ dernières colonnes. On appelle alors *mineur principal d'ordre* k de A le déterminant de la sous-matrice principale d'ordre k de A.

Proposition 3.7 *Soit Ω un ouvert de $\mathbb{R}^n$, $f : \Omega \to \mathbb{R}$ de classe C^2 et $a \in \Omega$ un point critique de f.*

1. *Si les n mineurs principaux de $H(f)(a)$ sont tous positifs, c'est-à-dire*

$$\frac{\partial^2 f}{\partial x_1^2}(a) > 0, \quad \begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2}(a) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(a) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(a) & \frac{\partial^2 f}{\partial x_2^2}(a) \end{vmatrix} > 0, \quad \begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2}(a) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(a) & \frac{\partial^2 f}{\partial x_1 \partial x_3}(a) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(a) & \frac{\partial^2 f}{\partial x_2^2}(a) & \frac{\partial^2 f}{\partial x_2 \partial x_3}(a) \\ \frac{\partial^2 f}{\partial x_3 \partial x_1}(a) & \frac{\partial^2 f}{\partial x_3 \partial x_2}(a) & \frac{\partial^2 f}{\partial x_3^2}(a) \end{vmatrix} > 0, \cdots$$

   alors a est un minimum relatif strict de f.

2. *Si les n mineurs principaux de $H(f)(a)$ alternent de signe, le premier étant négatif, c'est-à-dire*

$$\frac{\partial^2 f}{\partial x_1^2}(a) < 0, \quad \begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2}(a) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(a) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(a) & \frac{\partial^2 f}{\partial x_2^2}(a) \end{vmatrix} > 0, \quad \begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2}(a) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(a) & \frac{\partial^2 f}{\partial x_1 \partial x_3}(a) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(a) & \frac{\partial^2 f}{\partial x_2^2}(a) & \frac{\partial^2 f}{\partial x_2 \partial x_3}(a) \\ \frac{\partial^2 f}{\partial x_3 \partial x_1}(a) & \frac{\partial^2 f}{\partial x_3 \partial x_2}(a) & \frac{\partial^2 f}{\partial x_3^2}(a) \end{vmatrix} < 0, \cdots$$

   alors a est un maximum relatif strict de f.

Soit $f : \Omega \to \mathbb{R}$ une fonction C^2 sur un ouvert Ω de $\mathbb{R}^2$. Étant donné un point $a \in \Omega$, on note

$$r = \frac{\partial^2 f}{\partial x^2}(a), \quad s = \frac{\partial^2 f}{\partial x \partial y}(a), \quad t = \frac{\partial^2 f}{\partial y^2}(a)$$

(ce qu'on appelle les *notations de Monge*). Si a est un point critique de f, alors $Df(a) = 0$. Les propositions précendentes, nous permettent alors de considérer les cas suivants :

1. Si $rt - s^2 > 0$ et $r > 0$, alors $D^2 f(a)$ est définie positive. La fonction f admet en a un minimum relatif strict.
2. Si $rt - s^2 > 0$ et $r < 0$, alors $D^2 f(a)$ est définie négative. La fonction f admet en a un maximum relatif strict.
3. Si $rt - s^2 < 0$, $D^2 f(a)$ a une valeur propre positive et une autre négative. Alors, a n'est pas extremum de f.
4. Si $rt - s^2 = 0$, $D^2 f(a)$ est dégénérée. Le comportement de f au voisinage de a dépend des termes suivants de son développement de Taylor.

4 Les théorèmes d'inversion locale et des fonctions implicites

Dans ce chapitre, on aborde les premiers théorèmes d'existence. Le résultat central est le théorème d'inversion locale et le théorème des fonctions implicites en est une des conséquences importantes.

Dans ce qui suit, les espaces normés sont complets (donc, espaces de Banach).

4.1 Difféomorphismes et isomorphismes

Soient E et F e.v.n.. Si u est un isomorphisme continu de E vers F (c'est-à-dire, une application linéaire bijective et continue), a priori rien ne dit que l'isomorphisme réciproque soit continu. Pourtant, si les espaces vectoriels normés E et F sont complets, c'est bien le cas. On admet les énoncés suivants.

Théorème 4.1 (Théorème de Banach) *Soient E, F espaces de Banach. Si $u : E \to F$ est un isomorphisme continu alors l'isomorphisme réciproque u^{-1} est aussi continu.*

On désigne $\mathrm{Isom}(E, F)$ l'ensemble des isomorphismes bicontinus de E vers F. Cette notation et le théorème précédent font que si $u \in \mathrm{Isom}(E, F)$ alors $u^{-1} \in \mathrm{Isom}(F, E)$.

Notons J l'application de $\mathrm{Isom}(E, F)$ dans $\mathrm{Isom}(F, E)$ définie par $J(u) = u^{-1}$.

Théorème 4.2 *Soient E, F espaces de Banach. L'application $J : Isom(E, F) \to Isom(F, E)$ est C^∞. De plus, pour tous $u \in Isom(E, F)$ et $h \in \mathcal{L}(E, F)$, on a*

$$DJ(u).h = -u^{-1} \circ h \circ u^{-1}.$$

Définition 4.1 *Soient $U \subset E$ et $V \subset F$ ouverts dans des espaces de Banach E et F. On dit que $f : U \to V$ est un difféomorphisme si f est une bijection différentiable telle que f^{-1} soit également différentiable.*

Proposition 4.1 *Si $f : U \to V$ est un difféomorphisme alors, en tout point $x \in U$, sa différentielle est un isomorphisme vérifiant :*

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

Si, de plus, f est C^k alors f^{-1} l'est aussi.

Preuve : Si $f : U \to V$ est un difféomorphisme alors

$$f \circ f^{-1} = 1_V \quad \text{et} \quad f^{-1} \circ f = 1_U.$$

En dérivant en tous points $x \in U$ et $y = f(x) \in V$, on obtient :

$$Df(x) \circ D(f^{-1})(y) = 1_F \quad \text{et} \quad D(f^{-1})(y) \circ Df(x) = 1_E.$$

Donc, $Df(x)$ et $D(f^{-1})(y)$ sont deux isomorphismes réciproques l'un de l'autre. On a

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

On peut écrire également : $D(f^{-1}) = J \circ Df \circ f^{-1}$. Alors, si f est C^1, $D(f^{-1})$ sera continu. Par récurrence, f^{-1} sera C^k si f l'est.

L'existence d'un difféomorphisme entre U et V fait que les espaces E et F sont isomorphes. Donc, il ne peut pas exister de difféomorphisme d'un ouvert de $\mathbb{R}^n$ vers un ouvert de $\mathbb{R}^m$ lorsque $n \neq m$. Dans le cas $E = F = \mathbb{R}^n$, U et V ouverts de $\mathbb{R}^n$ et $f : U \to V$ un difféomorphisme, alors pour tout $x \in U$,

$$D(f^{-1})(f(x)) = [D(f)(x)]^{-1}.$$

Ce qu'on appelle habituellement un *changement de variables* est en fait un difféomorphisme. En dimension deux, le changement de variables le plus courant est celui des *coordonnées polaires* :

$$(r, \theta) \mapsto (r\cos\theta, r\sin\theta),$$

qui réalise un difféomorphisme de $]0, +\infty[\times]0, 2\pi[$ sur $\mathbb{R}^2 \setminus \{(x, 0) : x > 0\}$.

4.2 Le Théorème d'inversion locale

Préliminaires : Soit $f : \mathbb{R} \to \mathbb{R}$ une application C^1 telle que $f'(a) \neq 0$. Il existe donc un intervalle ouvert I, contenant a, où f' garde un signe constant (par exemple, positif). Ainsi, f est croissante sur I et est bijective de I sur l'intervalle ouvert $J = f(I)$. Si y est “assez proche” de $f(a)$, c'est-à-dire $y \in J$, l'équation $f(x) = y$ a donc une solution $x = f^{-1}(y)$. D'autre part, on sait que $f^{-1} : J \to I$ est aussi C^1 et que $(f^{-1})'(y) = \frac{1}{f'(x)}$. La solution $x = f^{-1}(y)$ est donc une fonction C^1 de y. On va généraliser ce résultat.

Définition 4.2 *Soit E espace de Banach et $A \subset E$. On dit que $f : A \to A$ est une contraction de A si f est k-lipschitzienne avec $k \in [0, 1[$, c'est-à-dire*

$$\forall x, y \in A, \quad \|f(x) - f(y)\| \leq k\|x - y\|.$$

Théorème 4.3 (Théorème du point fixe) *Soit E un espace de Banach et A une partie fermée non vide de A. Alors toute contraction f de A possède un unique point fixe.*

Preuve : Soit $f : A \to A$ une contraction de A.

Unicité du point fixe : Si x et y sont deux points fixes de f, alors

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

Comme $0 \leq k < 1$, on a $\|x - y\| = 0 \iff x = y$.

Existence du point fixe : On fixe $x_0 \in A$ et on définit la suite récurrente $x_{n+1} = f(x_n)$. On a donc

$$\|x_{n+1} - x_n\| \leq k\|x_n - x_{n-1}\|.$$

Alors, par récurrence sur $n \geq 1$, on obtient

$$\|x_{n+1} - x_n\| \leq k^n\|x_1 - x_0\|.$$

De l'inégalité triangulaire on déduit, pour tout $n > m \geq 1$,

$$\|x_n - x_m\| = \|\sum_{i=m}^{n-1}(x_{i+1} - x_i)\| \leq \sum_{i=m}^{n-1}\|x_{i+1} - x_i\| \leq \sum_{i=m}^{n-1} k^i\|x_1 - x_0\| \leq k^m\|x_1 - x_0\|.$$

L'hypothèse $0 \leq k < 1$ entraîne que $(x_n)_{n\geq 0}$ est une suite de Cauchy dans A. Alors (x_n) converge vers une limite notée x dans $\bar{A} = A$. Par la continuité de f, on a :

$$f(\lim_{n\to+\infty} x_n) = \lim_{n\to+\infty} f(x_n) = \lim_{n\to+\infty} x_{n+1} \implies f(x) = x.$$

On a vu dans la proposition 4.1 que la différentielle d'un difféomorphisme est un isomorphisme. La réciproque va être localement vraie.

Théorème 4.4 (Inversion locale) *Soit Ω un ouvert de $\mathbb{R}^n$ et $f : \Omega \to \mathbb{R}^n$ au moins C^1. Supposons $a \in \Omega$ tel que*

$$Df(a) \in Isom(\mathbb{R}^n, \mathbb{R}^n).$$

Alors, f est un difféomorphisme local en a, c'est-à-dire : il existe un voisinage ouvert U de a, $U \subset \mathbb{R}^n$ et un voisinage V de $f(a)$, tels que $f : U \to V$ soit un difféomorphisme. De plus, $(f_{|U})^{-1}$ a la même classe de différentiabilité que f.

Remarque : Ce théorème affirme que l'équation $f(x) = y$ admet une solution x unique, pourvu que y soit choisi "assez proche" de $b = f(a)$ et que x soit cherché "assez proche" de a.

La démonstration du théorème d'inversion locale va être fait en plusieurs étapes.

Lemme 4.1 *Soit $f : U \to V$ un homéomorphisme différentiable tel que, en $a \in U$, $Df(a) \in Isom(\mathbb{R}^n, \mathbb{R}^n)$. Alors f^{-1} est différentiable en $b = f(a)$.*

Preuve : On note $g = f^{-1}$, $b = f(a)$ et $L = Df(a)$. Puisque U est un ouvert de $\mathbb{R}^n$ et par la continuité de f, pour h proche de l'origine, $a + h \in U$ et $k = f(a + h) - f(a)$ est proche de l'origine. Remarquons que

$$k = f(a + h) - f(a) \iff h = g(b + k) - g(b).$$

Par la continuité de f et g,

$$k \quad \text{tend vers zéro} \iff h \quad \text{tend vers zéro}.$$

On a

$$k = f(a + h) - f(a) = L(h) + \|h\|\varepsilon(h).$$

Alors,

$$L^{-1}(k) = h + \|h\|L^{-1}(\varepsilon(h)).$$

Pour conclure, il suffit de montrer que $\|h\|L^{-1}(\varepsilon(h))$ est un $o(k)$. On a :

$$\begin{aligned} L^{-1}(k) = h + \|h\|L^{-1}(\varepsilon(h)) \quad &\implies \quad \|h\| \leq \|h\|\|L^{-1}(\varepsilon(h))\| + \|L^{-1}\|\|k\| \\ &\implies \quad \|h\| \leq \frac{\|L^{-1}\|}{1 - \|L^{-1}(\varepsilon(h))\|}\|k\|. \end{aligned}$$

Quad h et k tendent vers zéro, $\frac{\|L^{-1}\|}{1-\|L^{-1}(\varepsilon(h))\|} \leq M$ (pour une constante $M > 0$). Alors,

$$\|h\|\|L^{-1}(\varepsilon(h))\| \leq M\|k\|\|L^{-1}(\varepsilon(h))\|.$$

Comme $\|L^{-1}(\varepsilon(h))\|$ tend vers zéro avec k, on a donc $\|h\|L^{-1}(\varepsilon(h)) = o(k)$.

Corollaire 4.1 *On suppose les conditions du lemme précédent avec, de plus, f de classe C^1. Alors f est un difféomorphisme local en a.*

Preuve : L'ensemble $\mathrm{Isom}(\mathbb{R}^n, \mathbb{R}^n)$ est ouvert dans $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ (on admet ce résultat). Par la continuité de Df, il existe un voisinage ouvert $W \subset U$ de a tel que $Df(W) \subset \mathrm{Isom}(\mathbb{R}^n, \mathbb{R}^n)$. Donc, en tout point $x \in W$, f vérifie les conditions du lemme précédent, alors $f_{|W}$ est un difféomorphisme sur $f(W)$.

Pour démonstrer le théorème d'inversion, il suffit donc de montrer que ses hypothèeses entraînent que f est un homéomorphisme local en a. Notons que la proposition 4.1 assure que f et f^{-1} sont de la même classe de différentiabilité.

Preuve du théorème 4.4 : On peut supposer que $f(a) = a = 0$ et $Df(a) = \mathrm{Id}_{\mathbb{R}^n}$. En effet, il suffit de remplacer $f(x)$ par $h(x) = [Df(a)]^{-1}[f(a+x) - f(a)]$ et de remarquer que f est difféomorphisme local en a si et seulement si h est difféomorphisme local en 0.

Par la continuité de Df, il existe $r > 0$ tel que

$$\|x\| \leq r \implies \|Df(x) - \mathrm{Id}_{\mathbb{R}^n}\| < \frac{1}{2}.$$

En appliquant l'inégalité de la moyenne, on a

$$\forall x \in B(0_{\mathbb{R}^n}, r), \quad \|x - f(x)\| \leq \frac{\|x\|}{2}.$$

Pour tout $y \in B(0_{\mathbb{R}^n}, r/2)$, et pour $\|x\| \leq r$, on a

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

Alors, $y + x - f(x) \in B(0_{\mathbb{R}^n}, r)$. On peut donc définir :

$$\varphi : \bar{B}(0_{\mathbb{R}^n}, r) \longrightarrow B(0_{\mathbb{R}^n}, r) \subset \bar{B}(0_{\mathbb{R}^n}, r) \quad \text{par} \quad \varphi(x) = y + x - f(x).$$

On a $\|D\varphi(x)\| = \|\mathrm{Id}_{\mathbb{R}^n} - Df(x)\| < \frac{1}{2}$, pour tout $x \in \bar{B}(0_{\mathbb{R}^n}, r)$, donc φ est une contraction de la boule fermée $\bar{B}(0_{\mathbb{R}^n}, r)$. Comme $\mathbb{R}^n$ est un espace de Banach, par le théorème du point fixe, il existe un unique point $x \in \bar{B}(0_{\mathbb{R}^n}, r)$ tel que :

$$x = \varphi(x) = y + x - f(x) \iff f(x) = y.$$

Mais φ prend ses valeurs dans la boule ouverte. Alors, pour tout $y \in B(0_{\mathbb{R}^n}, r/2)$ il existe un unique $x \in B(0_{\mathbb{R}^n}, r)$ tel que $y = f(x)$. Donc, f réalise une bijection, notée g, de

$$U = f^{-1}(B(0_{\mathbb{R}^n}, r)) \cap B(0_{\mathbb{R}^n}, r) \quad \text{sur} \quad V = B(0_{\mathbb{R}^n}, r/2).$$

Il est clair que $g = f_{|U}$ est continue. De plus, pour tous $x, x' \in B(0_{\mathbb{R}^n}, r)$, on a

$$\|f(x) - f(x') - (x - x')\| = \|\varphi(x) - \varphi(x')\| \leq \frac{1}{2}\|x - x'\|,$$

ce qui implique que

$$\|x - x'\| \leq 2\|f(x) - f(x')\|.$$

Alors, $g^{-1} : V \to U$ est 2-lipschitzienne et donc continue.

Remarque : En dimension finie, il suffit que le déterminant de la matrice jacobienne, appelé *jacobien*, de f en a soit différent de zéro pour conclure que f est un difféomorphisme local.

Comme exemples de difféomorphismes locaux, on trouve les changements de variables. Les coordonnées polaires ou sphèriques :

$$(r,\theta) \in \mathbb{R}^* \times \mathbb{R} \mapsto (r\cos\theta, r\sin\theta) \in \mathbb{R}^2$$

$$(r,\theta,\varphi) \in \mathbb{R}^* \times \mathbb{R} \times \mathbb{R} \mapsto (r\cos\theta\sin\varphi, r\sin\theta\sin\varphi, r\cos\varphi) \in \mathbb{R}^3$$

sont des applications qui sont partout localement des difféomorphismes, mais ce ne sont pas des difféomorphismes globaux (elles ne sont pas injectives).

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

$$f(x,y) = (y\sin(x), yx^2).$$

La fonction f est C^∞ et on a

$$J(f)(x,y) = \begin{pmatrix} y\cos(x) & \sin(x) \\ 2xy & x^2 \end{pmatrix}.$$

Donc $\det J(f)(x,y) = yx(x\cos(x) - 2\sin(x))$. Si $y \neq 0$, $x \neq 0$ et $x\cos(x) \neq 2\sin(x)$, $Df(x,y)$ est inversible et f détermine un difféomorphisme C^∞ d'un voisinage de (x,y).

Corollaire 4.2 *Soit Ω un ouvert de $\mathbb{R}^n$. Si $f : \Omega \to f(\Omega)$ est de classe C^1, injective et, pour tout $x \in \Omega$, $Df(x) \in Isom(\mathbb{R}^n, \mathbb{R}^n)$, alors f est un difféomorphisme de Ω sur $f(\Omega)$.*

4.3 Le Théorème des fonctions implicites

On considère l'équation classique $f(x,y) = x^2 + y^2 - 1 = 0$ qui définit le cercle unité sur le plan $\mathbb{R}^2$. On sait expliciter la variable y en fonction de x, $y = \sqrt{1-x^2}$, mais cette écriture n'est pas équivalente à l'équation du départ. Mais si on se restreint à des ouverts, les deux équations sont équivalentes.

Comme son nom l'indique, le théorème des *fonctions implicites* donne des conditions suffisantes pour que, dans un équation du type $f(x,y) = 0$, on puisse expliciter une variable en fonction de l'autre.

On considère un système d'équations

$$\begin{cases} f_1(x_1,\dots,x_n,y_1,\dots,y_p) = 0 \\ f_2(x_1,\dots,x_n,y_1,\dots,y_p) = 0 \\ \vdots \\ f_p(x_1,\dots,x_n,y_1,\dots,y_p) = 0 \end{cases}$$

où les f_i sont des fonctions réelles de variables réelles. En notant $x = (x_1,\dots,x_n)$, $y = (y_1,\dots,y_p)$ et $f = (f_1,\dots,f_p)$, le système s'écrit :

$$f(x,y) = 0 \in \mathbb{R}^p.$$

On suppose Ω ouvert de $\mathbb{R}^n \times \mathbb{R}^p$ et $f : \Omega \to \mathbb{R}^p$ une fonction C^1. Soient $(a,b) \in \Omega$ et $D_x f(a,b) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$, $D_y f(a,b) \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^p)$ les différentielles partielles de f en (a,b) par rapport à x et y, respectivement, définies par

$$D_x f(a,b)(h) = J_x f(a,b)h,\ h \in \mathbb{R}^n \ \text{ et } \ D_y f(a,b)(k) = J_y f(a,b)k,\ k \in \mathbb{R}^p,$$

où

$$J_x f(a,b) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a,b) & \dots & \frac{\partial f_1}{\partial x_n}(a,b) \\ \vdots & & \vdots \\ \frac{\partial f_p}{\partial x_1}(a,b) & \dots & \frac{\partial f_p}{\partial x_n}(a,b) \end{pmatrix} \ \text{ et } \ J_y f(a,b) = \begin{pmatrix} \frac{\partial f_1}{\partial y_1}(a,b) & \dots & \frac{\partial f_1}{\partial y_p}(a,b) \\ \vdots & & \vdots \\ \frac{\partial f_p}{\partial y_1}(a,b) & \dots & \frac{\partial f_p}{\partial y_p}(a,b) \end{pmatrix}.$$

sont les matrices des dérivées partielles de f par rapport aux x_i $(1 \leq i \leq n)$ et aux y_j $(1 \leq j \leq p)$, respectivement, en (a,b).

Théorème 4.5 (Fonctions implicites) *Soit Ω ouvert de $\mathbb{R}^n \times \mathbb{R}^p$ et $f : \Omega \to \mathbb{R}^p$ une fonction C^1. On suppose que, pour $(a,b) \in \Omega$, $f(a,b) = 0$ et $D_y f(a,b) \in Isom(\mathbb{R}^p, \mathbb{R}^p)$. Alors, il existe :*
- *un voisinage ouvert U de (a,b) dans Ω,*
- *un voisinage ouvert V de a en $\mathbb{R}^n$,*
- *une application $\varphi : V \to \mathbb{R}^p$, de même classe de différentiabilité que f,*

tels que

$$(x,y) \in U \ \text{ et } \ f(x,y) = 0 \iff x \in V \ \text{ et } \ y = \varphi(x).$$

De plus,

$$D\varphi(a) = -[D_y f(a,b)]^{-1} \circ D_x f(a,b).$$

La fonction φ est dite fonction explicitante.

On suppose que les f_i sont de classe C^k $(k \geq 1)$ et que $\det(J_y f(a,b)) \neq 0$. Alors, par le théorème des fonctions implicites, il existe un fonction explicitante φ de classe C^k à valeurs dans $\mathbb{R}^p$ telle que $y = \varphi(x)$. En posant $\varphi = (\varphi_1, \dots, \varphi_p)$, $x = (x_1, \dots, x_n)$ et $y = (y_1, \dots, y_p)$ on peut écrire :

$$\begin{cases} y_1 = \varphi_1(x_1, \dots, x_n) \\ y_2 = \varphi_2(x_1, \dots, x_n) \\ \vdots \\ y_p = \varphi_p(x_1, \dots, x_n) \end{cases}$$

Cet énoncé admet une variante dans laquelle la condition initiale $f(a,b) = 0$ est remplacée par $f(a,b) = c$. Dans ce cas, la fonction explicitante dépend différentiablement de c.

Théorème 4.6 (Fonctions implicites avec paramètre) *Soit Ω ouvert de $\mathbb{R}^n \times \mathbb{R}^p$ et $f : \Omega \to \mathbb{R}^p$ une fonction C^1. On suppose que, pour $(a,b) \in \Omega$, $f(a,b) = c$ et $D_y f(a,b) \in Isom(\mathbb{R}^p, \mathbb{R}^p)$. Alors, il existe :*
- *un voisinage ouvert U de (a,b) dans Ω,*
- *un voisinage ouvert W de (a,c) en $\mathbb{R}^n \times \mathbb{R}^p$,*
- *une application $\varphi : W \to \mathbb{R}^p$, de même classe de différentiabilité que f,*

tels que

$$(x,y) \in U \ \text{ et } \ f(x,y) = z \iff (x,z) \in W \ \text{ et } \ y = \varphi(x,z).$$

Le premier théorème est en certain sens un cas particulier du second.

Preuve Théorème 4.6 : On va se ramener au théorème d'inversion locale. On définit $g(x,y) = (x, f(x,y)) \in \mathbb{R}^n \times \mathbb{R}^p$. On a

$$Dg(a,b)(h,k) = (h, Df(a,b)(h,k)) = (h, D_x f(a,b)(h) + D_y f(a,b)(k)).$$

Pour tous $(h',k') \in \mathbb{R}^n \times \mathbb{R}^p$, l'hypothèse $D_y f(a,b) \in \mathrm{Isom}(\mathbb{R}^p, \mathbb{R}^p)$ permet de résoudre l'équation $(h',k') = Dg(a,b)(h,k)$. Alors, $Dg(a,b)$ est un isomorphisme dont l'inverse s'écrit :

$$[Dg(a,b)]^{-1}(h',k') = (h', [D_y f(a,b)]^{-1}(k' - D_x f(a,b)(h'))).$$

Par le théorème d'inversion locale, il existe un voisinage U de (a,b) et un voisinage W de $(a, f(a,b)) = (a,c)$ sur lesquels g est un difféomorphisme, l'application réciproque de g prenant sur W la forme $(x,z) \mapsto (x, \varphi_1(x,z))$. Comme $\varphi_1 = \pi_2 \circ g^{-1}$, sa classe de différentiabilité est celle de g (et de f). Par la bijectivité de g, on a

$$(x,y) \in U, \quad f(x,y) = z \iff (x,z) \in W, \quad y = \varphi_1(x,z).$$

Preuve du Théorème 4.5 : On raisonne comme pour le Théorème 4.6. En particulier, pour $z = 0$, on a

$$(x,y) \in U, \quad f(x,y) = 0 \iff (x,z) \in W, \quad y = \varphi_1(x,0).$$

En notant $j : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^p$ l'injection $j(x) = (x,0)$, $\varphi = \varphi_1 \circ j$ et $V = j^{-1}(W)$ (qui est un ouvert de $\mathbb{R}^n$ contenant a), on obtient

$$(x,y) \in U, \quad f(x,y) = 0 \iff x \in V, \quad y = \varphi(x).$$

La classe de différentiabilité de φ est celle de f. De plus, on a

$$\forall x \in V, \quad f(x, \varphi(x)) = 0.$$

Alors, en différentiant en x, pour tout $h \in \mathbb{R}^n$,

$$Df(x,\varphi(x))(h, D\varphi(x)(h)) = D_x f(x,\varphi(x))(h) + D_y f(x,\varphi(x))(D\varphi(x)(h)) = 0.$$

En particulier, en $(x,\varphi(x)) = (a,b)$, on obtient pour tout $h \in \mathbb{R}^n$,

$$D\varphi(a)(h) = -[D_y f(a,b)]^{-1}(D_x f(a,b)(h)).$$

D'où, $D\varphi(a) = -[D_y f(a,b)]^{-1} \circ D_x f(a,b)$.

Remarque. L'existence de $D\varphi(x)$ sur V est donnée par le théorème 4.5. Il est inutile d'apprendre une formule, mais bien plus profitable de retenir une formule : différencier les relations qui vérifient les fonctions implicites.

On sait que, pour tout $x = (x_1, \ldots, x_n) \in V$ et tout $1 \leq j \leq p$,

$$f_j(x_1, \ldots, x_n, y_1, \ldots, y_p) = 0,$$

c'est-à-dire

$$f_j(x_1, \ldots, x_n, \varphi_1(x_1, \ldots, x_n), \ldots, \varphi_p(x_1, \ldots, x_n)) = 0,$$

Alors, pour tout $1 \leq i \leq n$, on a

$$\frac{\partial f_j}{\partial x_i}(x, \varphi(x)) + \sum_{k=1}^{p} \frac{\partial f_j}{\partial x_k}(x, \varphi(x)) \frac{\partial \varphi_k}{\partial x_i}(x) = 0.$$

Ce résultat peut s'écrire

$$D\varphi(x) = -[D_y f(x, \varphi(x))]^{-1} \circ D_x f(x, \varphi(x)).$$

Exemple 1. Soit dans $\mathbb{R}^2$ le cercle unité :

$$S^1 = \{(x, y) \in \mathbb{R}^2 : f(x, y) = 0\},$$

avec $f(x, y) = x^2 + y^2 - 1$. La fonction $f : \mathbb{R}^2 \to \mathbb{R}$ étant de classe C^∞, elle admet ses deux dérivées partielles

$$\frac{\partial f}{\partial x}(x, y) = 2x, \quad \frac{\partial f}{\partial y}(x, y) = 2y.$$

Alors, $D_y f(a, b)(k) = 2bk$ et $D_y f(a, b)$ est un isomorphisme si et seulement si $b \neq 0$. Donc, en un point $(a, b) \in S^1$, $D_y f(a, b)$ est un isomorphisme si et seulement si $(a, b) \neq (1, 0)$ et $(a, b) \neq (-1, 0)$. Soit donc $(a, b) \in S^1$ qui n'est pas sur l'axe des x. On a :

- f est C^∞ sur $\mathbb{R}^2$,
- $f(a, b) = 0$,
- $D_y f(a, b) \in Isom(\mathbb{R}, \mathbb{R})$,
- f est C^∞ alors $(x, y) \mapsto D_y f(x, y)$ est continue.

Le théorème des fonctions implicites assure alors qu'au voisinage de a, S^1 est le graphe d'une application C^∞, $x \mapsto y = \varphi(x)$. On sait de plus que

$$D\varphi(x) = -[D_y f(x, \varphi(x))]^{-1} \circ D_x f(x, \varphi(x)).$$

Par exemple, en $(a, b) \in S^1$ et $b > 0$ (alors $b = \sqrt{1 - a^2}$),

$$\varphi'(a) = -\frac{\frac{\partial f}{\partial x}(a, b)}{\frac{\partial f}{\partial y}(a, b)} = -\frac{2a}{2b} = -\frac{a}{\sqrt{1 - a^2}}.$$

Exemple 2. Soit $f : \mathbb{R}^3 \to \mathbb{R}^2$ définie par $f(t, x, y) = (t + tx - y - x^3, x + ty - y^3)$. La fonction f est de classe C^1 sur $\mathbb{R}^3$ et $f(0, 0, 0) = (0, 0)$. Pour tout $(t, x, y) \in \mathbb{R}^3$,

$$\frac{\partial f}{\partial x}(t, x, y) = (t - 3x^2, 1), \quad \frac{\partial f}{\partial y}(t, x, y) = (-1, t), \quad \frac{\partial f}{\partial t}(t, x, y) = (1 + x, y).$$

Alors, l'application $D_{x,y} f(0, 0, 0) : \mathbb{R}^2 \to \mathbb{R}^2$ définie par

$$D_{x,y} f(0,0,0)((a,b)) = \begin{pmatrix} \frac{\partial f}{\partial x}(0,0,0) & \frac{\partial f}{\partial y}(0,0,0) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = (-b, a),$$

est un isomorphisme. Par le théorème des fonctions implicites, il existe un voisinage ouvert I de 0 dans $\mathbb{R}$, un voisinage ouvert U de $(0, 0, 0)$ dans $\mathbb{R}^3$, deux fonctions de classe C^1 : $x : I \to \mathbb{R}$ et $y : I \to \mathbb{R}$ telles que $x(0) = 0$, $y(0) = 0$ et

$$(t, x, y) \in U \text{ et } f(t, x, y) = (0, 0) \iff t \in I \text{ et } (x, y) = (x(t), y(t)).$$

Donc, pour tout $t \in I$,

$$f(t, x(t), y(t)) = (0,0) \iff \left\{ \begin{array}{l} t + tx(t) - y(t) - x^3(t) = 0 \\ x(t) + ty(t) - y^3(t) = 0 \end{array} \right. .$$

Alors, en dérivant en t, on obtient pour tout $t \in I$,

$$\left\{ \begin{array}{l} 1 + x(t) + tx'(t) - y'(t) - 3x'(t)x^2(t) = 0 \\ x'(t) + y(t) + ty'(t) - 3y'(t)y^2(t) = 0 \end{array} \right. .$$

Pour $t = 0$, on a

$$\left\{ \begin{array}{l} 1 - y'(0) = 0 \\ x'(0) = 0 \end{array} \right. \iff \left\{ \begin{array}{l} x'(0) = 0 \\ y'(0) = 1 \end{array} \right. .$$

5 Sous-variétés dans $\mathbb{R}^n$

5.1 Sous-variétés

Quoi de commun à la parabole d'équation

$$y = x^2,$$

à l'ovale de Cassini

$$(x^2 + y^2)^2 = x^2 - y^2 + 1,$$

à l'ellipse

$$x = a\cos t, \quad y = b\sin t,$$

les trois appelés courbes de $\mathbb{R}^2$? Quoi de commun au paraboloïde hyperbolique

$$z = xy,$$

au cylindre

$$x^2 + y^2 - x = 0,$$

au tore

$$x = (2 + \cos\varphi)\cos\theta, \quad y = (2 + \cos\varphi)\sin\theta, \quad z = \sin\varphi,$$

appelés surfaces de $\mathbb{R}^3$? La réponse est dans l'aspect local de ces sous-ensembles de $\mathbb{R}^2$ (respectivement $\mathbb{R}^3$) qui est celui d'une droite (respectivement un plan) que l'on aurait déformé.

L'application

$$F : (x, y) \longmapsto (u, v) = (x, y - x^2),$$

d'inverse

$$F^{-1} : (u, v) \longmapsto (x, y) = (u, v + u^2),$$

est un difféomorphisme de classe C^∞ de $\mathbb{R}^2$ sur lui-même, qui transforme la parabole $y = x^2$ en la droite $v = 0$.

Il peut sembler facile d'aplatir une parabole... Pourtant, ce serait impossible pour les ensembles définis par $y = |x|$ ou par $x^3 - y^3 = 0$. La première courbe est lisse à l'origine, les autres pas, au sens de la définition suivante.

Définition 5.1 *Soient M un sous-ensemble de $\mathbb{R}^n$, $a \in M$. On dit que M est lisse en a de dimension $p \leq n$, s'il existe un voisinage ouvert U de $\mathbb{R}^n$ et un difféomorphisme $\varphi : U \longrightarrow V \subset \mathbb{R}^n$ tels que*

$$\varphi(U \cap M) = V \cap (\mathbb{R}^p \times \{0_{\mathbb{R}^{n-p}}\}).$$

On dit que M est une sous-variété de dimension $p \leq n$ de $\mathbb{R}^n$ si M est lisse (de dimension p) en chacun de ses points.

Si le difféomorphisme est de classe C^k, M est dite C^k-sous-variété.

Une sous-variété se ramène (localement) à une droite, ou un plan, etc., par simple changement de coordonnées. Par définition, les notions de "lisse en un point", "sous-variété", sont invariantes par difféomorphisme : si φ est un difféomorphisme de $\mathbb{R}^n$, M est lisse en a (de dimension p) si et seulement si $\varphi(M)$ est lisse en $\varphi(a)$ (de dimension p).

On appelle *courbe lisse*, *surface lisse*, *hypersurface lisse*, une sous-variété de dimension 1, resp. 2, resp. $n - 1$, de $\mathbb{R}^n$.

Pour vérifier que certains sous-ensembles de $\mathbb{R}^n$ sont des sous-variétés, nous utilisons essentiellement *l'équation d'une sous-variété*, notion précisée dans l'énoncé suivant.

Théorème 5.1 *Soit Ω un ouvert de $\mathbb{R}^n$ et $g : \Omega \to \mathbb{R}^k$ une application C^1 avec $k < n$. On suppose que, pour tout $x \in \Omega$, $Dg(x)$ est surjective (on dit que g est une submersion C^1 sur Ω). Alors, le sous-ensemble $M = g^{-1}(\{0_{\mathbb{R}^k}\})$ est une sous-variété de dimension $n-k$ de $\mathbb{R}^n$. On dit que $g(x) = 0$ est l'équation de la sous-variété M.*

Nous admettons qu'une sous-variété peut toujours être *globalement définie* par une équation.

Remarque : Pour montrer que $M = g^{-1}(\{0_{\mathbb{R}^k}\})$ est une sous-variété de dimension $n-k$ de $\mathbb{R}^n$, où Ω un ouvert de $\mathbb{R}^n$ et $g : \Omega \to \mathbb{R}^k$ une application C^1 avec $k < n$, il suffit de montrer que, pour tout $x \in M$, $Dg(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)$ est surjective.

Exemples :

1. La sphère unité de $\mathbb{R}^{n+1}$, notée S^n, est une sous-variété de dimension n définie par l'équation :
$$g(x_1, \ldots, x_{n+1}) = x_1^2 + \ldots + x_{n+1}^2 - 1 = 0.$$
La différentielle de g a pour jacobienne :
$$J_g(x) = \begin{pmatrix} 2x_1 & 2x_2 & \ldots & 2x_{n+1} \end{pmatrix}, \quad x = (x_1, \ldots, x_{n+1}),$$
qui s'annule seulement à l'origine, point n'appartenant pas à $S^n = g^{-1}(0)$. Toutes les autres sphères de $\mathbb{R}^{n+1}$ (de rayon > 0) se déduisent par un difféomorphisme global et sont donc également des sous-variétés de dimension n.
2. Les cylindres dans $\mathbb{R}^3$ dont le modèle a pour équation $x^2 + y^2 - 1 = 0$, sont des sous-variétes de dimension 2.

Les sous-variétés peuvent aussi être caractérisées comme les images d'applications particulières : les *imersions* $f : O \to \mathbb{R}^n$ (c'est-à-dire $f : O \to \mathbb{R}^n$ différentiable, avec O ouvert de $\mathbb{R}^p$ et $p < n$, et telle que, pour tout $x \in \Omega$, $Df(x)$ est injective), qui réalisent un homéomorphisme de Ω sur leur image $f(\Omega)$ munie de la topologie induite.

Théorème 5.2 *Une partie M de $\mathbb{R}^n$ est une sous-variété de dimension p si et seulement si, en tout point $x \in M$, il existe une immersion $h : O \to \mathcal{U}$ d'un ouvert de $\mathbb{R}^p$ dans un voisinage ouvert $\mathcal{U}$ de x telle que h réalise un homéomorphisme de O sur $h(O) = \mathcal{U} \cap M$. Une telle immersion est un système de coordonnées locales ou encore une paramétrisation locale de M.*

Remarque. L'hypothèse que h réalise un homéomorphisme sur son image, empêche une situation de point double de se produire.

Il faut généralement plusieurs systèmes de coordonnées locales pour recouvrir une sous-variété.

Par exemple, les *coordonnées sphériques* :
$$(\theta, \phi) \in]0, 2\pi[\times] - \frac{\pi}{2}, \frac{\pi}{2}[\mapsto (R\cos(\theta)\cos(\phi), R\sin(\theta)\cos(\phi), R\sin(\phi)),$$
définissent une paramétrisation pour la sphère de rayon R. Il faut quatre ouverts de définition différentes pour récouvrir la sphère. Également, $(x, y) \mapsto (x, y, \sqrt{1 - x^2 - y^2})$ est un système de coordonnées locales pour la sphère unité.

5.2 Espace tangent

Intuitivement, l'espace tangent à une sous-variété est formé des vitesses de courbes tracés sur la sous-variété.

Théorème 5.3 *Soit $M = g^{-1}(\{0_{\mathbb{R}^k}\})$ une sous-variété de dimension $n-k$, où $g : \Omega \to \mathbb{R}^k$ est une submersion C^1 sur un ouvert Ω de $\mathbb{R}^n$. Alors :*

1. *M est partout localement un graphe.*
2. *Pour tout $x \in M$, le noyau $\ker Dg(x)$ est formé des vitesses de courbes tracées sur M et passant par x.*

On remarque d'une part que $kerDg(x)$ est de dimension $n-k$, celle de M, et d'autre part qu'il ne dépend pas du choix de g, puique formé des vitesses des courbes sur M. Ce qui justifie la définition suivante.

Définition 5.2 *Étant donné une sous-variété M définie par une équation $g(x) = 0$ (où g satisfait les hypothèses du théorème précédent), on appelle espace tangent à M au point $x \in M$, le sous-espace vectoriel $\ker Dg(x)$. On le note $T_x(M)$. Le sous-espace affine tangent en x est défini par $x + \ker Dg(x)$.*

Remarquer que M est définie par

$$0 = g(x) = Dg(a)(x-a) + \dots,$$

(puisque $g(a) = 0$), et M est donc voisine du sous-espace d'équation

$$Dg(a)(x-a) = 0,$$

qui n'est autre que son espace affine tangent en a.

5.3 Surfaces de $\mathbb{R}^3$

Soit $f : \mathbb{R}^3 \to \mathbb{R}$ une fonction de classe C^1 telle que, pour tout (a,b,c) de l'ensemble S d'équation $f(x,y,z) = 0$,

$$\frac{\partial f}{\partial z}(a,b,c) \neq 0.$$

Alors S est une sous-variété de dimension 2 de $\mathbb{R}^3$ que l'on appelle *surface de* $\mathbb{R}^3$.

L'application F définie sur $\mathbb{R}^3$ par

$$F(x,y,z) = (x,y,f(x,y,z)) = (u,v,w),$$

est un difféomorphisme local qui transforme S en un plan $w = 0$.

En effet, l'application F est de classe C^1 sur $\mathbb{R}^3$ et

$$DF = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{pmatrix}.$$

Donc $DF(a,b,c)$ est inversible, et le théorème d'inversion locale montre que F est un difféomorphisme d'un voisinage V de (a,b,c) sur un voisinage W de $F(a,b,c) = (a,b,0)$. Alors, en notant $(u,v,w) = F(x,y,z)$,

$$(x,y,z) \in S \cap V \iff (u,v,w) \in W \text{ et } w = 0.$$

Le changement de variables F transforme S (au voisinage du point considéré) en le plan $w = 0$, donc S est lisse de dimension 2 en tout point de $S \cap V$.

Soit $(a,b,c) \in S$. Alors $(X,Y,Z) \in \mathbb{R}^3$ est tangent en (a,b,c) à S si et seulement si

$$\frac{\partial f}{\partial x}(a,b,c)X + \frac{\partial f}{\partial y}(a,b,c)Y + \frac{\partial f}{\partial z}(a,b,c)Z = 0.$$

C'est l'équation du *plan vectoriel tangent* à S en (a,b,c) (noyau de l'application linéaire $Df(a,b,c)$). L'hypothèse $\frac{\partial f}{\partial z}(a,b,c) \neq 0$ signifie que ce plan n'est pas vertical.

5.4 Courbes de $\mathbb{R}^3$

Soient $f, g : \mathbb{R}^3 \longrightarrow \mathbb{R}$ deux fonctions de classe C^1. Soit C l'ensemble défini par

$$f(x,y,z) = 0, \quad g(x,y,z) = 0.$$

On suppose que, pour tout $(a,b,c) \in C$,

$$\frac{\partial f}{\partial y}(a,b,c)\frac{\partial g}{\partial z}(a,b,c) - \frac{\partial f}{\partial z}(a,b,c)\frac{\partial g}{\partial y}(a,b,c) \neq 0. \tag{6}$$

Alors, C est une sous-variété de dimension 1 de $\mathbb{R}^3$ que l'on appelle *courbe de* $\mathbb{R}^3$.

L'application F définie sur $\mathbb{R}^3$ par

$$F(x,y,z) = (x, f(x,y,z), g(x,y,z)) = (u,v,w),$$

est un difféomorphisme local qui transforme C en la droite $v = w = 0$

En effet, F est de classe C^1 sur $\mathbb{R}^3$ et

$$DF = \begin{pmatrix} 1 & 0 & 0 \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} & \frac{\partial g}{\partial z} \end{pmatrix}.$$

Pour tout $(x,y,z) \in \mathbb{R}^3$,

$$\det DF(x,y,z) = \frac{\partial f}{\partial y}(x,y,z)\frac{\partial g}{\partial z}(x,y,z) - \frac{\partial f}{\partial z}(x,y,z)\frac{\partial g}{\partial y}(x,y,z).$$

D'après l'hypothèse (6), $DF(a,b,c)$ est inversible d'où, par inversion locale, F est un difféomorphisme entre un voisinage V de (a,b,c) et un voisinage W de $F(a,b,c) = (a,0,0)$. En notant $(u,v,w) = F(x,y,z)$, on a

$$(x,y,z) \in C \cap V \iff (u,v,w) \in W \text{ et } v = w = 0.$$

Soit $(a,b,c) \in C$. Alors $(X,Y,Z) \in \mathbb{R}^3$ est tangent en (a,b,c) à C si et seulement si

$$\left\{ \begin{array}{l} \frac{\partial f}{\partial x}(a,b,c)X + \frac{\partial f}{\partial y}(a,b,c)Y + \frac{\partial f}{\partial z}(a,b,c)Z = 0 \\ \frac{\partial g}{\partial x}(a,b,c)X + \frac{\partial g}{\partial y}(a,b,c)Y + \frac{\partial g}{\partial z}(a,b,c)Z = 0 \end{array} \right.$$

C'est l'équation de la *droite vectorielle tangente* en (a,b,c) à C, intersection des plans vectoriels tangents aux deux surfaces $f = 0$, $g = 0$. L'hypothèse (6) signifie que ces deux plans se coupent selon une droite non parallèle au plan YOZ (c'est-à-dire $X \neq 0$).

5.5 Extrema liés – Multiplicateurs de Lagrange

Des nombreuses questions peuvent nous conduire à rechercher les extremums d'une fonction de plusieurs variables, sachant que ces variables sont "liées" par certaines relations ou *sous contraintes*. Nous allons nous intéresser aux extrema de la restriction à une sous-variété d'une fonction différentiable.

Définition 5.3 *Soient Ω un ouvert de $\mathbb{R}^n$, $f : \Omega \to \mathbb{R}$ et A une partie de Ω. On dit que a est un maximum lié (respectivement, minimum lié) de f sur A, si a est un maximum (respectivement, minimum) relatif de $f_{|A}$.*

Un point qui est un maximum ou un minimum lié est un extremum lié de f sur A.

Dans la suite de ce paragraphe, on établit une condition nécessaire pour qu'en un point a de A la fontion $f_{|A}$ admette un minimum relatif ou un maximum relatif.

Rappel d'algèbre linéaire. Soient E, F et G espaces vectoriels et deux applications $u \in \mathcal{L}(E, F)$, $v \in \mathcal{L}(E, G)$, on a l'équivalence :

$$\ker v \subset \ker u \iff \exists\, w \in \mathcal{L}(G, F) \text{ telle que } u = w \circ v.$$

Il suffit de considérer un sous-espace vectoriel supplémentaire H de $v(E)$ et de définier w par $H = \ker w$ et, pour tout $x \in E$, $w(v(x)) = u(x)$. L'hypothèse $\ker v \subset \ker u$ rend la valeur de w indépendante du choix de x. Comme cas particulier, on considère $F = \mathbb{R}$, $G = \mathbb{R}^k$ avec $k \geq 1$. Alors, $v = (v_1, \ldots, v_k)$ et l'équivalence devient :

$$\ker v \subset \ker u \iff \exists\, (\lambda_1, \ldots, \lambda_k) \in \mathbb{R}^k \text{ tel que } u(x) = \lambda_1 v_1(x) + \ldots + \lambda_k v_k(x).$$

Les coefficients $\lambda_1, \ldots, \lambda_k$ sont appelés *multiplicateurs de Lagrange*.

Contraintes d'égalité.

Soit Ω ouvert de $\mathbb{R}^n$, $f : \Omega \to \mathbb{R}$ de classe C^1 et $g : \Omega \to \mathbb{R}^k$ différentiable, avec $k \leq n$. On note $g = (g_1, \ldots, g_k)$. On suppose que, pour tout $x \in M = g^{-1}(\{0\})$, $Dg(x)$ est surjective. Alors, M est sous-variété de dimension $n - k$ de $\mathbb{R}^n$.

Notons que $x \in M = g^{-1}(\{0\})$ si les k contraintes d'égalité sont vérifiées :

$$g_1(x) = 0, \ldots, g_k(x) = 0.$$

Théorème 5.4 *Soit la sous-variété définie par $M = g^{-1}(\{0\})$. Si $a \in M$ est un point d'extremum de $f_{|M}$, il existe alors k multiplicateurs de Lagrange tels que :*

$$Df(a) = \lambda_1 Dg_1(a) + \ldots + \lambda_k Dg_k(a).$$

Preuve : Si a est un extremum de $f_{|M}$, pour tout courbe c tracée sur M et passant par $c(0) = a$, $f \circ c$ admet 0 comme extremum. Donc, $c'(0) \in \ker Df(a)$. Mais le théorème 5.3 affirme que tout vecteur de $\ker Dg(a)$ est de la forme $c'(0)$, alors

$$\ker Dg(a) \subset \ker Df(a).$$

Il suffit d'appliquer le rappel d'algèbre linéaire pour conclure.

Il est utile d'introduire la *fonction de Lagrange* ou *Lagrangien*.

Définition 5.4 *On appelle fonction de Lagrange ou Lagrangien associée à la fonction f et à la sous-variété $M = g^{-1}(\{0\})$, la fonction définie par :*

$$\mathcal{L}(x,\lambda) = f(x) - (\lambda_1 g_1(x) + \ldots + \lambda_k g_k(x)),$$

$x = (x_1,\ldots,x_k) \in \mathbb{R}^k$, $\lambda = (\lambda_1,\ldots,\lambda_k)$.

Remarques :

1. Nous pouvons dire que la recherche des minima ou maxima liés de f, pour la liaison $g(x) = 0$ (contrainte d'égalité), se ramène à la recherche des minima ou maxima relatifs (au sens usuel) de la fonction de Lagrange $\mathcal{L} = f - (\lambda_1 g_1 + \ldots + \lambda_k g_k)$. Il faut remarquer que les multiplicateurs de Lagrange $\lambda_1,\ldots,\lambda_k$ ne sont pas connus d'avance. Ils seront déterminés au même temps que le point a, grâce à la résolution des équations :

$$\begin{cases} Df(a) - (\lambda_1 Dg_1(a) + \ldots + \lambda_k Dg_k(a)) = 0 \\ g(a) = 0 \end{cases}$$

2. Supposons qu'en un point $a = (a_1,\ldots,a_n) \in \Omega$, la restriction de f à la sous-variété $M = g^{-1}(\{0_{\mathbb{R}^k}\})$ admette un maximum ou un minimum relatif. On suppose aussi que la différentielle de g en a est surjective, c'est-à-dire que la matrice $J_g(a) = \left(\frac{\partial g_j}{\partial x_i}\right)$ $(1 \leq i \leq n, 1 \leq j \leq k)$ est de rang k au point a.
Alors, il existe $\lambda = (\lambda_1,\ldots,\lambda_k) \in \mathbb{R}^k$ tel que, pour chaque $1 \leq i \leq n$,

$$\frac{\partial}{\partial x_i}\left(f - \sum_{j=1}^{k} \lambda_j g_j\right)(a) = 0.$$

3. Le théorème précédent nous donne uniquement l'ensemble des points candidats à point d'extremum de f sur M (c'est-à-dire, sous la contrainte d'égalité $g(x) = 0$). Comme dans le cas d'un problème d'extremums sans contrainte, pour déterminer la solution il faut établir des conditions suffisantes (du second ordre) permettant d'identifier les maxima et minima de f sur M parmi l'ensemble des points satisfaisant la condition nécessaire (du premier ordre - théorème 5.4).

Exemple : On considère la fonction f définie sur $\mathbb{R}^2$ par $f(x,y) = (x^2+y^2)e^{x^2-y^2}$. On calcule les extrema de f sur le cercle $C = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 = 2\}$.

On pose $g(x,y) = x^2+y^2-2$. On a

$$\nabla g(x,y) = (2x,2y) = (0,0) \iff (x,y) = (0,0) \notin C.$$

Alors, $Dg(x,y) : \mathbb{R}^2 \to \mathbb{R}$ est surjective pour tout $(x,y) \in \mathbb{R}^2 \setminus \{0_{\mathbb{R}^2}\}$. On conclut que $C = g^{-1}(\{0_{\mathbb{R}^2}\}$ est une sous-variété de dimension 1 de $\mathbb{R}^2$. D'autre part, C est un fermé borné de $\mathbb{R}^2$ et donc C est un compact de $\mathbb{R}^2$. Alors, f (étant une fonction continue sur $\mathbb{R}^2$) admet un maximum et un minimum global sur C.

Soit L la fonction de Lagrange définie par :

$$L(x,y,\lambda) = f(x,y) - \lambda g(x,y) = (x^2+y^2)e^{x^2-y^2} - \lambda(x^2+y^2-2).$$

On a :
$$\begin{cases} 2x(1+x^2+y^2)e^{x^2-y^2} - \lambda 2x = 0 \\ 2y(1-x^2-y^2)e^{x^2-y^2} - \lambda 2y = 0 \\ x^2+y^2 = 2 \end{cases} \iff \begin{cases} x = 0 \text{ ou } \lambda = 3e^{x^2-y^2} \\ y = 0 \text{ ou } \lambda = -e^{x^2-y^2} \\ x^2+y^2 = 2 \end{cases}$$

Les points critiques de L sont : $(0, \sqrt{2}, -e^{-2})$, $(0, -\sqrt{2}, -e^{-2})$, $(\sqrt{2}, 0, 3e^2)$ et $(-\sqrt{2}, 0, 3e^2)$. De plus, $f(0, \sqrt{2}) = f(0, -\sqrt{2}) = 2e^{-2}$ et $f(\sqrt{2}, 0) = f(-\sqrt{2}, 0) = 2e^2 > 2e^{-2}$. Donc, $(0, \sqrt{2})$ et $(0, -\sqrt{2})$ sont points de minimum absolu de f sur C, $(\sqrt{2}, 0)$ et $(-\sqrt{2}, 0)$ sont points de maximum absolu de f sur C.

Le théorème suivant donne des conditions suffisantes pour qu'un point $a \in M$ soit un point d'extremum local de f sur la sous-variété M.

Théorème 5.5 *Soit Ω ouvert de R^n, $f : \Omega \to \mathbb{R}$ de classe C^2 et $g : \Omega \to \mathbb{R}^k$, $1 \leq k < n$, de classe C^2 telle que $M = g^{-1}(\{0_{\mathbb{R}^k}\})$ est une sous-variété de dimension $n-k$ de $\mathbb{R}^n$.*

- *Un point $a \in M$ est un point de maximum relatif de $f_{|M}$ s'il existe $\bar{\lambda} \in \mathbb{R}^k$ tel que $(a, \bar{\lambda})$ est un point critique de la fonction de Lagrange $\mathcal{L}(x, \lambda)$ et*
$$v^t D_x^2 \mathcal{L}(a, \bar{\lambda}) v < 0, \tag{7}$$
pour tout $v \in \mathbb{R}^n$ non nul tel que $Dg(a)(v) = 0$.
- *Un point $a \in M$ est un point de minimum relatif de $f_{|M}$ s'il existe $\bar{\lambda} \in \mathbb{R}^k$ tel que $(a, \bar{\lambda})$ est un point critique de la fonction de Lagrange $\mathcal{L}(x, \lambda)$ et*
$$v^t D_x^2 \mathcal{L}(a, \bar{\lambda}) v > 0, \tag{8}$$
pour tout $v \in \mathbb{R}^n$ non nul tel que $Dg(a)(v) = 0$.

Les conditions (7) et (8) stipulent que la forme quadratique associée à la matrice hessienne par rapport aux variables $(x_1, \ldots, x_n)$ du Lagrangien en $(a, \bar{\lambda})$ est respectivement définie négative et positive sur l'hyperplan tangent à M au point a. Le théorème suivant, donne deux conditions suffisantes du secon ordre plus faciles à vérifier que (7) et (8) et qui impliquent celles-ci.

Théorème 5.6 (Conditions suffisantes) *Soit Ω ouvert de R^n, $f : \Omega \to \mathbb{R}$ de classe C^2 et $g : \Omega \to \mathbb{R}^k$, $1 \leq k < n$, de classe C^2 telle que $M = g^{-1}(\{0_{\mathbb{R}^k}\})$ est une sous-variété de dimension $n-k$ de $\mathbb{R}^n$.*

- *Un point $a \in M$ est un point de maximum relatif de f sur M s'il existe $\bar{\lambda} \in \mathbb{R}^k$ tel que $(a, \bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^k$ est un point critique de la fonction de Lagrange $\mathcal{L}(x, \lambda)$ et si la matrice hessienne bordée*
$$B = \begin{pmatrix} 0 & J(g)(a) \\ J(g)(a) & H_x(\mathcal{L})(a, \bar{\lambda}) \end{pmatrix}$$
est telle que les $n-k$ derniers mineurs principaux alternent de signe, le déterminant de B étant du même signe que $(-1)^n$.
- *Un point $a \in M$ est un point de minimum relatif de f sur M s'il existe $\bar{\lambda} \in \mathbb{R}^k$ tel que $(a, \bar{\lambda})$ est un point critique de la fonction de Lagrange $\mathcal{L}(x, \lambda)$ et si la matrice hessienne bordée*
$$B = \begin{pmatrix} 0 & J(g)(a) \\ J(g)(a) & H_x(\mathcal{L})(a, \bar{\lambda}) \end{pmatrix}$$
est telle que les $n-k$ derniers mineurs principaux sont tous du signe de $(-1)^k$.

Remarque. La matrice bordée B au point $(a,\bar\lambda)$ est une matrice carrée d'ordre $(k+n)\times(k+n)$ et elle est donnée par :

$$B=\begin{pmatrix} 0 & \cdots & 0 & \frac{\partial g_1}{\partial x_1}(a) & \cdots & \frac{\partial g_1}{\partial x_n}(a)\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & 0 & \frac{\partial g_k}{\partial x_1}(a) & \cdots & \frac{\partial g_k}{\partial x_n}(a)\\ \frac{\partial g_1}{\partial x_1}(a) & \cdots & \frac{\partial g_k}{\partial x_1}(a) & \frac{\partial^2\mathcal{L}}{\partial x_1^2}(a,\bar\lambda) & \cdots & \frac{\partial^2\mathcal{L}}{\partial x_n\partial x_1}(a,\bar\lambda)\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ \frac{\partial g_1}{\partial x_n}(a) & \cdots & \frac{\partial g_k}{\partial x_n}(a) & \frac{\partial^2\mathcal{L}}{\partial x_1\partial x_n}(a,\bar\lambda) & \cdots & \frac{\partial^2\mathcal{L}}{\partial x_n^2}(a,\bar\lambda)\end{pmatrix}.$$

Exemple. Soit $f(x,y)=xy$ et $D=\{(x,y)\in\mathbb{R}^2 : x+4y=16\}=g^{-1}(\{0\})$ où $g(x,y)=x+4y-16$. Puisque $\nabla g(x,y)=(1,4)\neq(0,0)$ pour tout $(x,y)\in\mathbb{R}^2$, D est une courbe de $\mathbb{R}^2$. On définit la fonction de Lagrange

$$\mathcal{L}(x,y,\mu)=f(x,y)-\mu(h(x,y)-16)=xy-\mu(x+4y-16).$$

On vérifie facilement que $\nabla\mathcal{L}(x,y,\mu)=(0,0,0)$ si et seulement si $(x,y,\mu)=(8,2,2)$. Finalement,

$$B(8,2,2)=\begin{vmatrix} 0 & \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y}\\ \frac{\partial h}{\partial x} & \frac{\partial^2\mathcal{L}}{\partial x^2} & \frac{\partial^2\mathcal{L}}{\partial x\partial y}\\ \frac{\partial h}{\partial y} & \frac{\partial^2\mathcal{L}}{\partial y\partial x} & \frac{\partial^2\mathcal{L}}{\partial y^2}\end{vmatrix}(8,2,2)=\begin{vmatrix} 0 & 1 & 4\\ 1 & 0 & 1\\ 4 & 1 & 0\end{vmatrix}=8>0.$$

Donc, $(8,2)$ est un point de maximum relatif de f sur D.

Contraintes d'inégalité.

Soient Ω ouvert de $\mathbb{R}^n$, $f:\Omega\to\mathbb{R}$ et $g_i:\mathbb{R}^n\to\mathbb{R}$ avec $i=1,\ldots,k$, des fonctions d'au moins classe C^1 sur $\mathbb{R}^n$. Soit $b=(b_1,\ldots,b_k)\in\mathbb{R}^k$ et D l'ensemble de points respectant les k contraintes d'inégalité :

$$D=\{x\in\mathbb{R}^n : g_1(x)\le b_1,\ldots,g_k(x)\le b_k\}.$$

L'ensemble D est un fermé de $\mathbb{R}^n$.

On définit la *fonction de Lagrange ou Lagrangien* par

$$\mathcal{L}(x_1,\ldots,x_n,\lambda_1,\lambda_k)=f(x)-[\lambda_1(g_1(x)-b_1)+\ldots+\lambda_k(g_k(x)-b_k)], \tag{9}$$

$x=(x_1,\ldots,x_n)$ et où $\lambda_1,\ldots,\lambda_k$ sont k nouvelles variables appelées *multiplicateurs de Lagrange*. On remarque que le Lagrangien (9) ne coincide plus avec la fonction f sur l'ensemble D comme dans le cas des contraintes d'égalité mais il nous permet toujours de transformer le problème sous contraintes à n variables en un problème sans contraintes à $n+k$ variables.

Cas de la dimension 2 : Soit $f:\mathbb{R}^2\to\mathbb{R}$ et $D=\{(x,y)\in\mathbb{R}^2 : g(x,y)\le b\}$ (une unique contrainte d'inégalité). La fonction de Lagrange associée est :

$$\mathcal{L}(x,y,\lambda)=f(x,y)-\mu(g(x,y)-b). \tag{10}$$

Soit $(\bar x,\bar y)$ un maximum ou minimum local de f sur l'ensemble D. Deux cas sont possibles : $g(\bar x,\bar y)=b$ ou $g(\bar x,\bar y)<b$.

Dans le premier cas, $g(\bar{x},\bar{y}) = b$, on dit que la contrainte est *saturée* en $(\bar{x},\bar{y})$. Comme dans le cas d'une contrainte d'égalité, la courbe de niveau $f(x,y) = f(\bar{x},\bar{y})$ est tangente en $(\bar{x},\bar{y})$ à la courbe répresentant l'ensemble $\{(x,y) \in \mathbb{R}^2 : g(x,y) = b\}$. Donc, les vecteurs gradients $\nabla f(\bar{x},\bar{y})$ et $\nabla g(\bar{x},\bar{y})$ ont la même direction en $(\bar{x},\bar{y})$, c'est-à-dire qu'il existe $\bar{\lambda} \in \mathbb{R}$ tel que

$$\nabla f(\bar{x},\bar{y}) = \bar{\lambda}\nabla g(\bar{x},\bar{y}). \tag{11}$$

De plus, puisque le vecteur gradient d'une fonction admet comme direction celle pour laquelle la fonction s'accroît le plus rapidement, les vecteurs gradients $\nabla f(\bar{x},\bar{y})$ et $\nabla g(\bar{x},\bar{y})$ doivent s'orienter dans le même sens si $(\bar{x},\bar{y})$ est un point de maximum et dans le sens opposé si $(\bar{x},\bar{y})$ est un point de minimum. Donc, le multiplicateur de Lagrange dans (11) doit vérifier $\bar{\lambda} \geq 0$ si $(\bar{x},\bar{y})$ est un point de maximum et $\bar{\lambda} \leq 0$ si $(\bar{x},\bar{y})$ est un point de minimum.

Dans le cas $g(\bar{x},\bar{y}) < b$, on dit que la contrainte *n'est pas saturée* en $(\bar{x},\bar{y})$ et le point $(\bar{x},\bar{y})$ est un point de maximum ou minimum local sans contraintes car il appartient à l'ensemble ouvert $\{(x,y) \in \mathbb{R}^2 : g(x,y) < b\}$. Il doit donc vérifier

$$\nabla f(\bar{x},\bar{y}) = (0,0),$$

et les dérivées de g n'interviennent pas dans la caractérisation de $(\bar{x},\bar{y})$.

On obtient le théorème suivant.

Théorème 5.7 (Condition nécessaire en dimension 2 et une contrainte) *Soient f et g deux fonctions de classe C^1 dans $\mathbb{R}^2$ et soit $(\bar{x},\bar{y})$ un point de maximum ou minimum local de f sur l'ensemble $D = \{(x,y) \in \mathbb{R}^2 : g(x,y) \leq b\}$ tel que $\nabla g(\bar{x},\bar{y}) \neq (0,0)$ si $g(\bar{x},\bar{y}) = b$. Alors, il existe un unique $\bar{\lambda} \in \mathbb{R}$ tel que*

$$\frac{\partial \mathcal{L}}{\partial x}(\bar{x},\bar{y},\bar{\lambda}) = 0, \quad \frac{\partial \mathcal{L}}{\partial y}(\bar{x},\bar{y},\bar{\lambda}) = 0,$$

et

$$\bar{\lambda}(g(\bar{x},\bar{y}) - b) = 0,$$

avec $\bar{\lambda} \geq 0$ si $(\bar{x},\bar{y})$ est un point de maximum et $\bar{\lambda} \leq 0$ si $(\bar{x},\bar{y})$ est un point de minimum.

La généralisation naturelle du théorème précédent au cas des fonctions de n variables sous k contraintes d'inégalité est la suivante.

Théorème 5.8 (Condition nécessaire en dimension n et k contraintes) *Soient f et $g_1,\ldots,g_k$, $k+1$ fonctions de classe C^1 dans $\mathbb{R}^n$ et soit $\bar{x} = (\bar{x}_1,\ldots,\bar{x}_n)$ un point de maximum ou minimum local de f sur l'ensemble*

$$D = \{x \in \mathbb{R}^n : g_1(x) \leq b_1, \ldots, g_k(x) \leq b_k\}.$$

Sans perte de généralité, on suppose qu'il existe $k_0 \leq k$ tel que

$$g_1(\bar{x}) - b_1 = \ldots = g_{k_0}(\bar{x}) - b_{k_0} = 0 \quad \text{et} \quad g_{k_0+1}(\bar{x}) < b_{k_0+1}, \ldots, g_k(\bar{x}) < b_k,$$

c'est-à-dire les k_0 premières contraintes sont saturées en $\bar{x}$ et les dernières $k - k_0$ ne le sont pas. Supposons enfin que le rang de la matrice jacobienne des k_0 contraintes saturées calculée en $\bar{x}$

$$\begin{pmatrix} \frac{\partial g_1}{\partial x_1}(\bar{x}) & \ldots & \frac{\partial g_1}{\partial x_n}(\bar{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_{k_0}}{\partial x_1}(\bar{x}) & \ldots & \frac{\partial g_{k_0}}{\partial x_n}(\bar{x}) \end{pmatrix}$$

est maximal et donc égal à k_0. Alors, il existe un unique $\bar\lambda = (\bar\lambda_1, \dots, \bar\lambda_k) \in \mathbb{R}^k$ tel que

$$\frac{\partial \mathcal{L}}{\partial x_i}(\bar x, \bar\lambda) = 0, \quad i = 1, \dots, n,$$

et

$$\bar\lambda_j (g_j(\bar x) - b_j) = 0, \quad j = 1, \dots, k,$$

avec $\bar\lambda_j \geq 0$ pour tout $j = 1, \dots, k$, si $\bar x$ est un point de maximum et $\bar\lambda_j \leq 0$ pour tout $j = 1, \dots, k$, si $\bar x$ est un point de minimum.

Remarque. Dans les théorèmes précédents, le signe des multiplicateurs de Lagrange est lié aux inégalités définies pas les k contraintes

$$g_i(x) \leq b_i \quad i = 1, \dots, k.$$

Si l'on change les inégalités en

$$g_i(x) \geq b_i \quad i = 1, \dots, k,$$

il faut changer aussi le signe des multiplicateurs de Lagrange.

Exemple. Soit $f(x, y) = xy$ et $D = \{(x, y) \in R^2 : x^2 + y^2 \leq 1\}$. On calcule les éventuels maxima et minima de f sur D. La fonction contrainte est $g(x, y) = x^2 + y^2$ et $\nabla g(x, y) = (2x, 2y) \neq (0, 0)$ pour tout $(x, y) \neq (0, 0)$. Donc, $M = g^{-1}(\{0\})$ est une sous-variété de $\mathbb{R}^2$ de dimension 1. On définit la fonction de Lagrange

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda(g(x, y) - 1) = xy - (x^2 + y^2 - 1).$$

Les conditions nécessaires impliquent :

$$\frac{\partial \mathcal{L}}{\partial x}(x, y, \lambda) = y - 2\lambda x = 0, \quad \frac{\partial \mathcal{L}}{\partial y}(x, y, \lambda) = x - 2\lambda y = 0, \quad \lambda(x^2 + y^2 - 1) = 0.$$

Si $\lambda = 0$, alors $(x, y) = (0, 0)$. Mais la fonction f atteint des valeurs positives et négatives en toute boule centrée en $(0, 0)$. Donc, $(0, 0)$ est un point selle.

Si $\lambda \neq 0$, alors la contrainte est saturée, c'est-à-dire $x^2 + y^2 = 1$, et

$$\lambda = \frac{y}{2x} = \frac{x}{2y}.$$

Alors, les points critiques de $\mathcal{L}$ sont :

$$(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, \frac{1}{2}), \quad (-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, \frac{1}{2}), \quad (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, -\frac{1}{2}), \quad (-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, -\frac{1}{2}).$$

Comme la contrainte est saturée en ces points, les conditions suffisantes du second ordre pour des contraintes d'égalité impliquent que les points $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$ et $(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$ sont deux points de maximum de f sur D et que $(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$ et $(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$ sont deux points de minimum de f sur D.

Références

[1] R. Abraham, J. Marsden, *Fondations of mechanics,* 2nd edition, Benjamin, New York, 1978.

[2] G. Auliac, J. Y. Caby, *Mathématiques, topologie et analyse,* Editions EdiScience, 2007.

[3] A. Avez, *Calcul différentiel,* Masson, 1983.

[4] M. Berger, B. Gostiaux, *Géométie différentielle,* Armand Colin, Paris, 1971.

[5] H. Cartan, *Cours de calcul différentiel,* Hermann, Éditeurs des Sciences et des Arts, 1967.

[6] F. Cottet-Emard, *Calcul différentiel et intégral,* De Boeck, 2007.

[7] J. Dieudonné, *Éléments d'analyse,* Tome 1, Gauthiers-Villars, Paris, 1968.

[8] P. Donato, *Calcul différentiel pour la licence*, Dunod, Paris, 2000.

[9] J. M. Souriau, *Calcul linéaire,* Tomes I et II, Jacques Gabay Éditeur, Paris, 1998.

[10] J. Stewart, *Analyse : concepts et contextes*, Volume 2 : *Fonctions de plusieurs variables*, De Boeck University, 2001.

Fonctions convexes et fonctions concaves'

On dit qu'un ensemble A de $\mathbb{R}^n$ est convexe si et seulement si pour tout $\lambda \in [0,1]$ et tout point $x, y \in A$ on a $\lambda x + (1-\lambda)y \in A$. Cela signifie que le segment d'extrémités x et y est entièrement contenu dans A.

Une fonction définie sur un sous-ensemble convexe A de $\mathbb{R}^n$ est *convexe* si, pour tout $\lambda \in [0,1]$ et tout point $x, y \in A$ on a

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y).$$

De même, une fonction définie sur un sous-ensemble convexe A de $\mathbb{R}^n$ est *concave* si, pour tout $\lambda \in [0,1]$ et tout point $x, y \in A$ on a

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y).$$

Théorème 5.9 (Caractérisation des fonctions convexes) *Soit U un ouvert convexe de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$ une fonction de classe C^1. Alors f est convexe si et seulement si, pour tout $x, y \in U$, on a*

$$f(y) - f(x) \geq Df(x)(y-x). \tag{12}$$

De plus, si f est de classe C^2 dans U, alors f est convexe si et seulement si, pour tout $x \in U$, tout $h \in \mathbb{R}^n$, on a

$$D^2 f(x)(h,h) \geq 0,$$

c'est-à-dire, la matrice hessienne $H(f)(x)$ est définie positive.

Théorème 5.10 (Caractérisation des fonctions convexes) *Soit U un ouvert convexe de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$ une fonction de classe C^1. Alors f est concave si et seulement si, pour tout $x, y \in U$, on a*

$$f(y) - f(x) \leq Df(x)(y-x). \tag{13}$$

De plus, si f est de classe C^2 dans U, alors f est concave si et seulement si, pour tout $x \in U$, tout $h \in \mathbb{R}^n$, on a

$$D^2 f(x)(h,h) \leq 0,$$

c'est-à-dire, la matrice hessienne $H(f)(x)$ est définie négative.

La condition (12) (respectivement, (13) dit qu'une fonction de classe C^1 est convexe (respectivement, concave) si et seulement si le plan tangent au graphe de la fonction est toujours au dessous (respectivement, au-dessus) du graphe.

On peut démontrer facilement l'existence d'un minimum absolu pour une fonction convexe et l'existence d'un maximum absolu pour une fonction concave.

Théorème 5.11 *Soit U un ouvert convexe de $\mathbb{R}^n$ et $f : U \to \mathbb{R}$ une fonction de classe C^1 et $a \in U$ un point critique de f. Alors, si f est convexe, a est un point de minimum absolu de f sur U. Et, si f est concave, a est un point de maximum absolu de f sur U.*