Leçon 215 : Applications différentiables définies sur un ouvert de $\mathbb{R}^n$. Exemples et applications.

Mémoire du M2 de préparation à l'agrégation

Février 2019

Thomas CAVALLAZZI, encadré par Ismaël BAILLEUL

Table des matières

1 Introduction **3**

2 Différentiabilité et énoncés fondamentaux **4**

2.1 Différentielle d'une application . . . 4

2.2 Dérivées partielles et liens avec la différentiabilité . . . 6

2.3 Inégalité des accroissements finis et applications . . . 8

2.4 Différentielle d'ordre supérieur et formules de Taylor . . . 8

3 Théorème d'inversion locale et des fonctions implicites : applications **10**

3.1 Le théorème d'inversion locale . . . 10

3.2 Le théorème des fonctions implicites . . . 13

3.3 Applications en géométrie . . . 14

4 Optimisation et calcul différentiel **17**

4.1 Optimisation sans contraintes : conditions nécessaires et suffisante d'optimalité . . . 17

4.2 Optimisation sous contraintes . . . 18

5 Questions du jury **19**

1 Introduction

Les applications différentiables sont utilisées dans de multiples domaines des mathématiques. L'idée qui a conduit à développer cette notion est que localement, une "bonne fonction" doit être approchable par sa valeur en un point à laquelle on ajoute un terme de correction linéaire. L'idée sous-jacente étant de ramener localement n'importe quelle fonction à une fonction affine, et de se reposer sur la puissance des outils d'algèbre linéaire. Ainsi, la notion a été développée pour faire du calcul différentiel, elle généralise la notion de dérivabilité d'une fonction de la variable réelle, qui avait été introduite par Newton et Leibniz. Les applications différentiables et le calcul différentiel sont ainsi au cœur de l'étude des équations différentielles, voire aux dérivées partielles puisque ces équations font naturellement intervenir des différentielles. On peut également citer la géométrie différentielle qui offre un cadre permettant de généraliser le calcul différentiel classique qu'on peut mener sur des ouverts d'espaces vectoriels, à des ensembles plus généraux appelés sous-variétés de $\mathbb{R}^n$.

On ne peut pas parler d'applications différentiables et de calcul différentiel sans dire un mot sur la physique. En effet, c'est sous l'impulsion de la physique que le calcul différentiel a été développé. Les physiciens ayant trouvé un " langage " adapté dans les mathématiques, ils ont voulu encoder les phénomènes étudiés, pour mieux les comprendre, dans des équations mathématiques. Or ces équations sont construites grâce à des postulats, des approximations, et souvent grâce des " bilans " locaux de conservations de certaines quantités. Ces relations de conservation donnent, à la limite, des équations aux dérivées partielles. C'est donc le calcul différentiel qui rend légitime certaines approximations, certains passages à la limite utilisés en physique : de nombreux outils s'offrent ainsi aux physiciens.

Une autre problématique liée aux applications différentiables est la classification des applications, des courbes et des surfaces, à difféomorphismes près. C'est le stade au dessus de la classification topologique, qui ne s'intéresse qu'à classifier à homéomorphismes près.

Enfin, les problématiques d'optimisation sont au coeur de l'étude des applications différentiables car leur étude fournit des conditions nécessaires et/ou suffisantes d'optimalité. L'optimisation sur un ouvert d'un espace vectoriel normé (on parlera d'optimisation sans contraintes) est du ressort du calcul différentiel "classique". Cependant, on peut parfois être amené à vouloir optimiser certaines applications sur des domaines plus compliqués (on parlera d'optimisation sous contraintes) : les sous variétés de $\mathbb{R}^n$.

Ainsi la première partie est consacrée à la mise en place des premiers outils concernant les applications différentiables, ainsi qu'à quelques théorèmes fondamentaux qui permettent de comprendre l'intérêt de la notion pour l'étude locale des fonctions. La seconde partie traite des deux énoncés importants que sont le théorème d'inversion locale et le théorème des fonctions implicites ainsi que leurs multiples implications notamment en géométrie différentielle. Enfin, la dernière partie porte sur l'utilisation du calcul différentiel en ce qui concerne l'optimisation avec et sans contraintes.

2 Différentiabilité et énoncés fondamentaux

Dans cette partie, Ω désigne un ouvert de $\mathbb{R}^n$ et on considèrera une application f définie sur Ω et à valeurs dans $\mathbb{R}^m$. On notera $\|.\|$ n'importe quelle norme sur $\mathbb{R}^n$, puisqu'elles sont équivalentes. Et on notera encore $\|.\|$ une autre norme quelconque sur $\mathbb{R}^m$ par abus de notation.

2.1 Différentielle d'une application

Mettons-nous un instant à la place d'anciens mathématiciens, ayant à leur disposition uniquement la notion de dérivabilité d'une fonction de la variable réelle. On souhaite généraliser cette notion à des applications non plus définies sur $\mathbb{R}$, mais sur Ω, à valeurs dans $\mathbb{R}^m$.

Une première idée qu'on pourrait avoir serait de se ramener à la dimension 1. Cette idée s'avérera fort utile pour démontrer des résultats généraux sur les applications différentiables en se reposant sur les résultats analogues pour les fonctions de la variable réelle. Cependant la définition qui suit, qu'on pourrait vouloir prendre comme définition de la différentiabilité, n'implique même pas la continuité, propriété qu'on attend d'une fonction différentiable, comme c'est le cas en dimension 1.

Définition 2.1 (*Dérivée directionnelle*)

Soit $a \in \Omega$ et $v \in \mathbb{R}^n$. On appelle, si elle existe, dérivée directionnelle de f en a selon v, la quantité :

$$D_v f(a) := \lim_{t \to 0} \frac{f(a+tv) - f(a)}{t}.$$

Définition 2.2 (*Gateaux-différentiabilité*)

f est dite Gateaux-différentiable en a si elle admet des dérivées directionnelles selon tout vecteur.

Le contre exemple suivant illustre qu'une application Gateaux-différentiable en un point (0 en l'occurence), peut ne même pas y être continue.

Contre-exemple 2.3. *L'application f définie sur R^2 par :*

$$f(x,y) = \begin{cases} \frac{x^5}{(y-x^2)^2+x^8} & si\ (x,y) \neq (0,0) \\ 0 & si\ (x,y) = (0,0) \end{cases}$$

est Gateaux-différentiable en 0, mais n'y est pas continue.

La définition suivante, plus générale, aura les propriétés attendues d'une généralisation de la dérivabilité. Elle repose sur la notion de développement limité.

Proposition-Définition 2.4 (*Différentiabilité*)

1. *f est dite différentiable en $a \in \Omega$, s'il existe $L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ et une application $q : \Omega \to \mathbb{R}^m$ qui tend vers 0 en a telles que :*
$$\forall x \in \Omega,\ f(x) = f(a) + L(x-a) + \|x-a\| q(x).$$
2. *L est unique et on la note df_a.*
3. *On dit que f est différentiable, si elle est différentiable en tout point. Dans ce cas, sa différentielle est l'application :*
$$df : \begin{cases} \Omega & \to & \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \\ a & \mapsto & df_a \end{cases}.$$
4. *On munit $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ de sa norme subordonnée aux deux normes sur $\mathbb{R}^n$ et $\mathbb{R}^m$. Si $df : \Omega \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ est continue sur Ω, on dira que f est de classe $\mathcal{C}^1$ et on notera $f \in \mathcal{C}^1(\Omega, \mathbb{R}^m)$.*

Notation 2.5. *On notera dans la suite $o_a(\|x-a\|)$ toute application de Ω dans $\mathbb{R}^m$ qui, divisée par la norme $\|x-a\|$, tend vers 0 lorsque $x \to a$. Ainsi on peut réécrire la formule dans la définition de la différentiabilité par :*

$$\forall x \in \Omega,\ f(x) = f(a) + L(x-a) + o_a(\|x-a\|).$$

Voyons que cette notion de différentiabilité généralise bien la notion de dérivée directionnelle au sens de Gateaux.

Proposition 2.6

Si f est différentiable en $a \in \Omega$, alors f est Gateaux-différentiable en a et on a :

$$\forall v \in \mathbb{R}^n,\ D_v f(a) = df_a.v.$$

Donnons maintenant quelques exemples d'applications différentiables.

Exemple 2.7.

1. *Si f est constante, alors f est différentiable et sa différentielle est l'application nulle.*
2. *Si $f \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, alors f est différentiables et sa différentielle est l'application constante égale à f,*
3. *Si $f \in \mathcal{L}(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R}^m)$ est une application bilinéaire. Alors f est différentiable et on a :*

$$\forall (x,y) \in \mathbb{R}^n \times \mathbb{R}^n,\ \forall (h_x, h_y) \in \mathbb{R}^n \times \mathbb{R}^n,\ df_{(x,y)}.(h_x, h_y) = f(x, h_y) + f(h_x, y).$$

4. *L'application P_j d'élévation à la puissance j sur $\mathcal{M}_n(\mathbb{R})$ est $\mathcal{C}^1$ et sa différentielle est donnée par :*

$$\forall H \in \mathcal{M}_n(\mathbb{R}), \quad dP_{j\,M}.H = \sum_{k=0}^{j-1} M^k H M^{j-1-k}.$$

Proposition 2.8

Soient $f, g : \Omega \to \mathbb{R}^m$ deux applications différentiables en $a \in \Omega$. Alors :

1. *f est continue en a.*
2. *Pour tout $\lambda \in \mathbb{R}$, $d(f+\lambda g)_a = df_a + \lambda dg_a$.*
3. *si f et g sont à valeurs réelles, alors le produit fg est différentiables en a et on a pour tout $h \in \mathbb{R}^n$: $d(fg)_a.h = (df_a.h)g(a) + f(a)(dg_a.h)$.*

En fait l'étude de la différentiabilité peut se ramener aux applications de Ω à valeurs dans $\mathbb{R}$ comme le montre la proposition suivante.

Proposition 2.9

Notons les coordonnées de f de la manière suivante : $f = (f_1, \ldots, f_m)$. Alors f est différentiable en a si et seulement si toutes les applications coordonnées f_i sont différentiables en a. Dans de cas, la différentielle de f est donnée par :

$$df_a = (df_{1\,a}, \ldots, df_{m\,a}) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m).$$

Enfin, le théorème suivant permet d'exprimer la différentielle d'une composée d'applications différentiables, ce qui est utile en pratique.

Théorème 2.10 (*Théorème des fonctions composées*)

Soient Ω_1, Ω_2 deux ouverts de $\mathbb{R}^n$ et $\mathbb{R}^m$ respectivement, $f : \Omega_1 \to \mathbb{R}^m$ et $g : \Omega_2 \to \mathbb{R}^k$ deux applications différentiables, avec $f(\Omega_1) \subset \Omega_2$. Alors si f est différentiable en $a \in \Omega_1$ et g en $f(a)$, la composée $g \circ f$ est différentiable en a et on a :

$$d(g \circ f)_a = dg_{f(a)} \circ df_a.$$

Exemple 2.11.

1. *Si la norme $\|.\|$ est la norme euclidienne canonique sur $\mathbb{R}^n$, alors l'application $\|.\|^2$ est $\mathcal{C}^1$ sur $\mathbb{R}^n$ et on a :*

$$\forall a, h \in \mathbb{R}^n,\ d\|.\|^2_a.h = 2\langle a, h\rangle.$$

2. *L'application* det *est* $\mathcal{C}^1$ *sur* $\mathcal{M}_n(\mathbb{R})$ *et vérifie :*

$$\forall M, H \in \mathcal{M}_n(\mathbb{R}),\ d(\det)_M.H = Tr({}^t Com(M)H).$$

Application 2.12. *On considère une application* $A \in \mathcal{C}(\mathbb{R}, \mathcal{M}_n(\mathbb{R}))$. *On note* $y_1, \ldots, y_n$ *des solutions du système différentiel linéaire :*

$$y'(t) = A(t)y(t).$$

Le wronskien de ces solutions est défini par :

$$w(t) = \det(y_1(t), \ldots, y_n(t)).$$

Il s'agit d'une application $\mathcal{C}^1$*, de plus on a l'égalité suivante :*

$$\forall t \in \mathbb{R},\ w(t) = w(0)e^{\int_0^t Tr(A(s))\,ds}.$$

2.2 Dérivées partielles et liens avec la différentiabilité

Comme on l'a vu au début de la partie précédente, une application qui admet des dérivées directionnelles selon toute direction n'est pas forcément différentiable. Cependant cette notion est utile en pratique car il est souvent plus facile de se ramener à la dimension 1 via la notion de dérivée directionnelle.

Définition 2.13 (*Dérivées partielles*)

Notons $(e_1, \ldots, e_n)$ *la base canonique de* $\mathbb{R}^n$. *On définit la dérivée partielle de* f *en* a *par rapport à la variable* x_j*, comme la dérivée directionnelle de* f *en* a *selon le vecteur* e_j. *On la note des deux manières suivantes :*

$$\frac{\partial f}{\partial x_j}(a) = \partial_j f(a) := D_{e_j} f(a).$$

On a vu dans la partie précédente qu'une application différentiable est Gateaux-différentiable. En fait dans ce cas, on peut reconstruire la différentielle de f en fonction de ses dérivées partielles.

Proposition 2.14

Si f *est différentiable en* $a \in \Omega$*, on a pour tout* $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$: $df_a.h = \displaystyle\sum_{k=1}^{n} h_k \partial_k f(a)$.

Le théorème suivant est important puisqu'il permet de montrer la différentiabilité d'une application uniquement grâce à l'étude des dérivées partielles, avec une hypothèse de continuité supplémentaire.

Théorème 2.15

Il y a équivalence entre :

1. $f \in \mathcal{C}^1(\Omega, \mathbb{R}^m)$
2. f *admet des dérivées partielles* $\partial_k f(a)$ *pour tout* $k \in \{1, \ldots, n\}$*, pour tout* $a \in \Omega$. *De plus les applications dérivées partielles* $\partial_k f(.)$ *sont continues sur* Ω.

Exemple 2.16. *Soit* $f : \mathbb{R}^2 \to \mathbb{R}$ *définie par :*

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

f *n'est pas différentiable en* $(0,0)$. *En effet :*

- $f(x,0) = 0$ *donc* f *admet une dérivée partielle selon* x *et* $\frac{\partial f}{\partial x}(0,0) = 0$.
- *De même,* $f(0,y) = -y$ *donc* f *admet une dérivée partielle selon* y *en* $(0,0)$ *et* $\frac{\partial f}{\partial y}(0,0) = -1$.
- *Supposons que* f *est différentiable en* $(0,0)$*, alors pour tout* $(h_1, h_2) \in \mathbb{R}^2$,

$$df_{(0,0)}.(h_1, h_2) = \frac{\partial f}{\partial x}(0,0)h_1 + \frac{\partial f}{\partial y}(0,0)h_2 = -h_2$$

et

$$\epsilon(x,y) = \frac{f(x,y) - f(0,0) - df_{(0,0)}.(x,y)}{\|(x,y)\|} \xrightarrow[\|(x,y)\| \to 0]{} 0.$$

Pour $(x,y) \neq (0,0)$*, on a*

$$\epsilon(x,y) = \frac{1}{(x^2+y^2)^{\frac{3}{2}}} \left(y(x^2-y^2) + y(x^2+y^2)\right) = \frac{2x^2y}{(x^2+y^2)^{\frac{3}{2}}}.$$

En particulier, pour $a, b \in \mathbb{R}^*$ *et* $t > 0$*,*

$$\varepsilon(at, bt) = \frac{2a^2b}{(a^2+b^2)^{\frac{3}{2}}} \not\to 0$$

Donc f *n'est pas différentiable en* $(0,0)$*.*

Si on travaille avec des fonctions à valeurs réelles, il peut être utile de travailler avec le gradient, représentant de la différentielle en tant que forme linéaire.

Définition 2.17 (*Gradient*)

Soit $f : \Omega \to \mathbb{R}$ *différentiable. On définit le gradient de* f *en* $a \in \Omega$ *comme l'unique vecteur* $\nabla f_a \in \mathbb{R}^n$ *vérifiant :*

$$df_a = \langle \nabla f_a, . \rangle \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}).$$

De plus on a :

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

Remarque 2.18.

1. *Dans le cadre Hilbertien, l'existence d'un tel vecteur est assuré par le théorème de Riesz (on suppose l'application linéaire continue dans la définition de la différentiabilité dans le cadre du calcul différentiel sur des espaces de Banach).*
2. *Le vecteur gradient* ∇f_a *est la direction dans laquelle* f *varie le plus autour du point* a*. Cela est une conséquence directe de l'inégalité de Cauchy-Schwarz, ainsi que de la caractérisation du cas d'égalité par la colinéarité des vecteurs :*

$$\forall h \in \mathbb{R}^n, \quad |\langle \nabla f_a, h \rangle| \leq \|\nabla f_a\|.\|h\|.$$

C'est cette propriété qui sous-tend les algorithmes d'optimisation dits de "descente de gradient ".

Exemple 2.19.

1. *Le gradient du carré de la norme euclidienne est donné par :*

$$\forall a \in \mathbb{R}^n, \quad \nabla \|.\|^2(a) = 2a.$$

2. *Soient* $A \in \mathcal{S}_n(\mathbb{R})$ *une matrice symétrique et* $b \in \mathbb{R}^n$*. L'application* $q : x \in \mathbb{R}^n \mapsto \frac{1}{2}\langle Ax, x \rangle - \langle b, x \rangle$ *est* $\mathcal{C}^1$ *sur* $\mathbb{R}^n$ *et son gradient est :*

$$\nabla q(x) = Ax - b$$

Dans le cas où la fonction f est à valeurs dans $\mathbb{R}^m$, on peut définir une matrice analogue au gradient.

Définition 2.20 (*Jacobienne*)

Soit $f : \Omega \to \mathbb{R}^m$ *différentiable en* $a \in \Omega$*. On définit sa jacobienne en* a *comme la matrice de l'application* df_a *dans les bases canoniques de* $\mathbb{R}^n$ *et* $\mathbb{R}^m$ *:*

$$Jf(a) := \left(\frac{\partial f_i}{\partial x_j}\right)_{(i,j) \in \{1,\ldots,n\} \times \{1,\ldots,m\}} \in \mathcal{M}_{n,m}(\mathbb{R}),$$

où l'on note $f = (f_1, \dots, f_m)$.

2.3 Inégalité des accroissements finis et applications

Dans la suite, on notera $f : \Omega \to \mathbb{R}^m$ une application différentiable.

L'inégalité des accroissements finis permet de contrôler les variations de la fonction par la norme de sa différentielle, ce qui utile pour obtenir des inégalités. La preuve consiste à appliquer l'énoncé analogue en dimension 1 en paramétrant le segment $[x, y]$ par un paramètre réel t.

Théorème 2.21 (*Inégalités des accroissements finis*)

Soient $x, y \in \Omega$ tels que $[x, y] \subset \Omega$. Soit $\phi : [0, 1[\to \mathbb{R}$ dérivable telle que :

$$\forall t \in [0,1], \quad \|df_{x+t(y-x)}.(y-x)\| \leq \phi'(t).$$

Alors

$$\|f(y) - f(x)\| \leq \phi(1) - \phi(0).$$

Corollaire 2.22

Si $[x, y] \subset \Omega$, alors on a :

$$\|f(y) - f(x)\| \leq \sup_{z \in [x,y]} \left(\|df_z\|_{\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)}\right) \|x - y\|.$$

Remarquons que le théorème s'applique dans le cas où on suppose Ω convexe.

Proposition 2.23

Supposons l'ouvert Ω connexe. Alors il y a équivalence entre :

1. *f est constante*
2. *Pour tout $x \in \Omega$, $df_x = 0_{\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)}$.*

Application 2.24. *L'inégalité des accroissements finis assure qu'une fonction $f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$ est localement lipschitzienne. Ainsi, le théorème de Cauchy-Lipschitz assure que si $f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^n)$, le problème de Cauchy :*

$$\begin{cases} y' = f(y) \\ y(0) = x_0 \in \mathbb{R}^n \end{cases}$$

a une unique solution maximale.

Application 2.25. *Soit $f : [a, \infty[\to \mathbb{R}$ une application $\mathcal{C}^1$. Alors pour tout entier $n \geq a$, on a :*

$$\left| f(n) - \int_n^{n+1} f(x)\, dx \right| \leq \frac{1}{2} \max_{n \leq x \leq n+1} |f(x)|.$$

Cela permet, par une comparaison série/intégrale, de montrer que la série de terme général $\frac{\sin(\ln(n))}{n}$ diverge.

2.4 Différentielle d'ordre supérieur et formules de Taylor

Il est parfois intéressant de pousser le développement limité d'une application différentiable à un ordre supérieur. Pour cela on itère la notion de différentiabilité sur l'application différentielle. On définit les classes de régularité de fonctions différentiables de même que dans le cas mono-dimensionnel.

Définition 2.26 (*Classe $\mathcal{C}^k$ et $\mathcal{C}^\infty$*)

Par récurrence sur n, on dit que f est k-fois différentiable en a s'il existe un voisinage ouvert U de a dans Ω tel que f soit $(k-1)$-fois différentiable sur U et $\mathrm{d}^{k-1}f : U \to \mathcal{L}((\mathbb{R}^n)^{k-1}, \mathbb{R}^m)$ est différentiable en a.

$d^k f(a)$ s'identifie à une application k-linéaire de $(\mathbb{R}^n)^k \to \mathbb{R}^m$.

On définit également :

- $\mathcal{C}^k(\Omega, \mathbb{R}^m) = \{f : \Omega \to \mathbb{R}^m,\ f$ *différentiable sur* $\Omega,\ df : \Omega \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \in \mathcal{C}^{n-1}(\Omega, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))\}$
- $\mathcal{C}^\infty(\Omega, F) = \bigcap_{n\in\mathbb{N}} \mathcal{C}^n(\Omega, \mathbb{R}^m)$.

Remarque 2.27. *A priori la différentielle seconde $d^2 f$ est une application $\Omega \to \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$. Cependant l'application suivante permet de faire l'identification entre $\mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m))$ et l'espace vectoriel $\mathcal{L}(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R}^m)$ des applications bilinéaires. L'application*

$$J : \left\{ \begin{array}{ccc} \mathcal{L}(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R}^m) & \to & \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)) \\ \alpha & \mapsto & \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \\ x_1 & \mapsto & \alpha(x_1, \cdot) \end{array} \right. \end{array} \right.$$

est une isométrie bijective.

Théorème 2.28 (*Lemme de Schwarz*)

Soient $j, k \in \{1, \ldots, n\}$. Si $\dfrac{\partial^2 f}{\partial x_j \partial x_k}$ et $\dfrac{\partial^2 f}{\partial x_k \partial x_j}$ existent sur Ω et sont continues sur Ω alors elles sont égales.

Proposition 2.29

Si f est deux fois différentiable en a alors $d^2 f(a)$ est une application bilinéaire symétrique $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$.

Dans le cas des fonctions à valeurs réelles, tout comme le gradient condense l'information de la différentielle, on synthétise l'information de la différentielle seconde dans une matrice appelée Hessienne.

Définition 2.30 (*Hessienne*)

Si f est deux fois différentiable en a à valeurs réelles ($m = 1$), on définit sa Hessienne en a comme la matrice :

$$\mathrm{Hess} f(a) := \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a) \right)_{1 \leq i,k \leq n}.$$

Il s'agit d'une matrice symétrique.

Proposition 2.31

Si f est k-fois différentiable en $a \in \Omega$ alors $d^k f(a)$ est une application k-linéaire et symétrique $(\mathbb{R}^n)^k \to \mathbb{R}^m$:

$$\forall h_1, \ldots, h_k \in \mathbb{R}^n,\ \forall \sigma \in \mathfrak{S}_k, \qquad d^n f_a.(h_1, \ldots, h_k) = d^n f_a.(h_{\sigma(1)}, \ldots, h_{\sigma(k)}).$$

Proposition 2.32

1. *Si f est k-fois différentiable en $a \in \Omega$ alors elle admet des dérivées partielles jusqu'à l'ordre k en a. La réciproque étant fausse.*
2. *$f \in \mathcal{C}^k(\Omega, \mathbb{R}^m)$ si et seulement si f admet des dérivées partielles jusqu'à l'ordre k continues sur Ω.*

La formule suivante précise à l'ordre k le développement limité d'une fonction k-fois différentiable en un point.

Théorème 2.33 (*Formule de Taylor-Young*)

Si f est k-fois différentiable en $a \in \Omega$. Alors on a le développement limité suivant :

$$f(a+h) \underset{h\to 0}{=} f(a) + df_a.h + \frac{1}{2} d^2 f_a.(h,h) + \cdots + \frac{1}{n!} d^n f_a.(h, \ldots, h) + o(\|h\|^n).$$

Enfin la formule de Taylor avec reste intégral, est une formule exacte qui permet de reconstruire une fonction de classe $\mathcal{C}^k$ en fonction de ses différentielles jusqu'à l'ordre k.

Théorème 2.34 (*Formule de Taylor avec reste intégral*)

Soit $f \in \mathcal{C}^{k+1}(\Omega, \mathbb{R}^m)$ et $[a, a+h] \subset \Omega$ alors :

$$f(a+h) = f(a) + df_a.h + \cdots + \frac{1}{n!} d^n f_a.(h, \ldots, h) + \int_0^1 \frac{(1-t)^n}{n!} d^{n+1} f_{a+th}.(h, \ldots, h) dt.$$

La preuve consiste à appliquer le résultat connu en dimension 1 à la fonction de la variable réelle $t \mapsto f(a+th)$. On en verra une application dans la partie suivante concernant le lemme de Morse.

3 Théorème d'inversion locale et des fonctions implicites : applications

3.1 Le théorème d'inversion locale

Définition 3.1 (*Difféomorphisme*)

Soient Ω_1, Ω_2 deux ouverts de $\mathbb{R}^n$. $f : \Omega_1 \to \Omega_2$ est un $\mathcal{C}^1$-difféomorphisme de Ω_1 sur Ω_2 si :

1. $f \in \mathcal{C}^1(\Omega_1, \Omega_2)$
2. *f est bijective*
3. $f^{-1} \in \mathcal{C}^1(\Omega_2, \Omega_1)$.

Remarque 3.2.
1. *Si f et f^{-1} sont de classe $\mathcal{C}^k$, on dit que f est un $\mathcal{C}^k$-difféomorphisme.*
2. *On entendra par difféomorphisme un $\mathcal{C}^1$-difféomorphisme.*

L'interprétation qu'on a de deux applications qui diffèrent à composition par un difféomorphisme près sont qu'elle s'obtiennent l'une l'autre par une déformation suffisamment régulière de l'espace. Ainsi un $\mathcal{C}^1$-difféomorphisme est simplement une déformation de l'identité.

Proposition 3.3 (*Formule du changement de variables*)

Soit U un ouvert de $\mathbb{R}^n$ et ϕ un difféomorphisme entre U et $W = \phi(U) \subset \mathbb{R}^n$. Soit f mesurable sur $\mathbb{R}^n$. Alors f est intégrable sur W si et seulement si $f \circ \phi(.) \, |\det(J\phi(.))|$ est intégrable sur U. Dans ce cas, on a la formule suivante :

$$\int_{\phi(U)} f(x)\, dx = \int_U f \circ \phi(y) \, |\det(J\phi(y))| \, dy.$$

L'énoncé suivant est fondamental : il donne une condition suffisante (et nécessaire) pour qu'une application $\mathcal{C}^1$ soit un difféomorphisme local.

Théorème 3.4 (*Théorème d'inversion locale*)

Soit Ω un ouvert de $\mathbb{R}^n$ et $f \in \mathcal{C}^1(\Omega, \mathbb{R}^n)$. Si f est un difféomorphisme sur un voisinage de a alors $df_a \in GL(\mathbb{R}^n)$. Réciproquement, si $df_a \in GL(\mathbb{R}^n)$, alors il existe V un voisinage ouvert de a dans Ω et W un voisinage ouvert de $f(a)$ dans $\mathbb{R}^n$ tel que f soit un difféomorphisme de V sur W. De plus si f est de classe $\mathcal{C}^k$, alors f^{-1} est également de classe $\mathcal{C}^k$.

La preuve repose sur le théorème du point fixe de Banach.

Contre-exemple 3.5. *L'hypothèse $\mathcal{C}^1$ est nécessaire. En effet,*

$$f : \left\{ \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ x & \mapsto & x + x^2 \sin\left(\frac{\pi}{x}\right). \end{array} \right.$$

est dérivable en 0 et $f'(0) = 1 \neq 0$ *car* $f(x) = x + \underset{x \to 0}{o}(x)$. *Mais* f *n'est pas injective car non monotone au voisinage de* 0.

$$f\left(\frac{1}{2k}\right) = \frac{1}{2k} > f\left(\frac{1}{2k+1}\right) = \frac{1}{2k+1}$$

mais un calcul simple donne

$$f\left(\frac{1}{2k+\frac{1}{2}}\right) > f\left(\frac{1}{2k}\right).$$

On veut maintenant trouver des conditions pour être globalement un difféomorphisme entre deux ouverts donnés. C'est l'objet des deux prochains théorèmes.

Théorème 3.6 (*Théorème d'inversion globale*)

Si $f \in \mathcal{C}^1(\Omega, \mathbb{R}^n)$, *est injective et dont la différentielle est inversible en tout point, alors* f *est un difféomorphisme entre* Ω *et son image* $f(\Omega)$.

Définition 3.7

Une application $f : \mathbb{R}^n \to \mathbb{R}^n$ *est dite propre si l'une des deux propriétés suivantes, qui sont équivalentes, est vérifiée :*

1. *Pour tout compact* K *de* $\mathbb{R}^n$, $f^{-1}(K)$ *est compact.*
2. $\|f(x)\| \underset{\|x\| \to +\infty}{\to} +\infty$.

Le théorème suivant fait l'objet du premier développement.

Théorème 3.8 (*Théorème d'inversion globale de Hadamard-Lévy*)

Soit $f \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R}^n)$. *Il y a équivalence entre :*

1. f *est un* $\mathcal{C}^1$*-difféomorphisme global de* $\mathbb{R}^n$.
2. f *est propre et pour tout* $x \in \mathbb{R}^n$, $df_x \in GL(\mathbb{R}^n)$.

Démonstration : ($\Rightarrow$) Supposons que f soit un $\mathcal{C}^1$-difféomorphisme global de $\mathbb{R}^n$. En différentiant la relation $f^{-1} \circ f = \mathrm{Id}$, on obtient par le théorème des fonctions composées que pour tout $x \in \mathbb{R}^n$, $d(f^{-1})_{f(x)} \circ df_x = \mathrm{Id}$, ce qui assure l'inversibilité de la différentielle en tout point. Pour le caractère propre, il suffit de remarquer que si K est un compact de $\mathbb{R}^n$, alors $f^{-1}(K)$ est un compact comme image du compact K par l'application continue f^{-1}.

($\Leftarrow$) **Étape 1 :** Supposons maintenant que f est une application propre et dont la différentielle est inversible en tout point. Cette dernière hypothèse permet d'appliquer le théorème d'inversion locale à f ce qui assure que f est localement un $\mathcal{C}^1$-difféomorphisme. Il s'agit donc de montrer que f est bijective, on pourra ainsi utiliser le théorème d'inversion globale. On se donne $y \in \mathbb{R}^n$ et on cherche à montrer que y admet un unique antécédent par f, c'est-à-dire que Card $(f^{-1}(\{y\})) = 1$. Quitte à poser $g = f - y$, qui vérifie les mêmes hypothèses que f, on se ramène à étudier le nombre d'antécédents de 0 par f.

Étape 2 : L'idée est alors simple : il s'agit de trouver un flot le long duquel f décroît vers 0. Pour cela, on pose :

$$F : \left\{ \begin{array}{ll} \mathbb{R}^n & \to \mathbb{R}^n \\ x & \mapsto -(df_x)^{-1}.f(x) \end{array} \right. .$$

Il s'agit d'une application de classe $\mathcal{C}^1$ puisqu'on a supposé f de classe $\mathcal{C}^2$. Ainsi on considère le problème de Cauchy suivant :

$$\left\{ \begin{array}{ll} y' & = F(y) \\ y(0) & = q \in \mathbb{R}^n \end{array} \right. .$$

Le théorème de Cauchy-Lipschitz assure que ce problème admet une unique solution maximale qu'on note $\varphi(.,q)$ définie sur un intervalle maximal $[0, T^*[$. Voyons maintenant en quoi le choix de cette fonction F

permet d'atteindre l'objectif qu'on s'est fixé. On considère l'application $t \in [0, T^*[\mapsto g(t) := f \circ \varphi(t,q)$. Cette application est $\mathcal{C}^1$ et vérifie :

$$\begin{aligned} \forall t \in [0, T^*[, \quad g'(t) &= df_{\varphi(t,q)}.\partial_t \varphi(t,q) \\ &= df_{\varphi(t,q)}.(-(df_{\varphi(t,q)})^{-1}.f(\varphi(t,q))) \\ &= -g(t). \end{aligned}$$

Ainsi par unicité de la solution, on en déduit que :

$$\forall t \in [0, T^*[,\ f(\varphi(t,q)) = e^{-t} f(q).$$

En particulier, on en déduit que le flot φ est à valeurs dans $f^{-1}(\overline{B}(0, \|f(q)\|))$ qui est compact puisque f est propre. Le théorème de sortie de tout compact assure que la solution φ est globale, c'est-à-dire que $T^* = +\infty$. De plus comme le flot est à valeurs dans un compact, on en déduit qu'il existe $y \in \mathbb{R}^n$ et une sous-suite $(t_k)_k$ strictement croissante vers $+\infty$ telle que :

$$\varphi(t_k, q) \underset{k \to +\infty}{\to} y.$$

On déduit alors, par continuité de f, que $f(y) = 0$ puisque $g(t_k) \to 0$. On a donc l'existence d'un antécédent de 0 pour f (ce qui montre la surjectivité de f). Il s'agit maintenant de montrer que cet antécédent est unique.

Étape 3 : On remarque que les équilibres du système différentiel introduit sont exactement les zéros de f. Nous allons voir qu'ils sont asymptotiquement stables. Pour cela appliquons le théorème d'inversion locale en y. On dispose de U^y un voisinage de y dans $\mathbb{R}^n$ et de $B^y := B(0, \delta_y)$ tel que f soit un $\mathcal{C}^1$-difféomorphisme de U^y sur B^y. Supposons qu'il existe t_0 tel que $\varphi(t_0, q) \in U^y$. Montrons que l'on a :

$$\{t \in [t_0, +\infty[,\ \varphi(t,q) \in U^y\} = \{t \in [t_0, +\infty[,\ \varphi(t,q) = f^{-1}_{|U^y}(e^{-t} f(q))\}.$$

L'inclusion directe vient du fait que $f(\varphi(t_0, q)) = e^{-t_0} f(q) \in B^y$ et comme la norme décroît quand t augmente, $f(\varphi(t,q)) \in B^y$, pour tout $t \geq t_0$. En appliquant $f^{-1}_{|U^y}$, on obtient l'inclusion directe. L'inclusion réciproque est claire.

Le premier ensemble est ouvert comme pré-image d'un ouvert par l'application continue $\varphi(., q)$. Le second est fermé comme pré-image de 0 par l'application continue $\varphi(t,q) - f^{-1}_{|U^y}(e^{-t} f(q))$. Comme les ensembles sont égaux, il s'agit d'un ouvert fermé non vide, donc égal à $[t_0, +\infty[$, par connexité de ce dernier ensemble. En laissant tendre t vers $+\infty$ dans l'égalité : $\varphi(t,q) = f^{-1}_{|U^y}(e^{-t} f(q))$, on obtient que :

$$\varphi(t,q) \underset{t \to +\infty}{\to} y,$$

puisque y est le seul 0 de f dans U^y. Or dans l'étape 2, on a fait converger le flot vers y à extraction près, on en déduit qu'en fait il converge globalement vers y lorsque t tend vers $+\infty$.

Étape 4 : On utilise un argument de connexité pour conclure. Posons, pour $y \in f^{-1}(\{0\})$:

$$W^y := \left\{ q \in \mathbb{R}^n,\ \varphi(t,q) \underset{t \to +\infty}{\to} y \right\}.$$

Dans les étapes précédentes, on a vu que :

$$(*) \quad \mathbb{R}^n = \bigcup_{y \in f^{-1}(\{0\})} W^y.$$

En effet, le flot issu de toute donnée initiale converge vers un 0 de f.

Montrons que les W^y sont des ouverts non vides. Le caractère non vide provient du fait que $y \in W^y$, puisque y est un équilibre du système différentiel. Pour le caractère ouvert, on reprend le voisinage U^y de y

obtenu à l'étape 2 par le théorème d'inversion locale. On dispose de $\eta_y > 0$, tel que $B(y, 2\eta_y) \subset U^y$. Soit $q \in W^y$, on dispose de $T > 0$ tel que $\varphi(T, q) \in B(y, \eta_y)$. De plus, la continuité du flot par rapport à la donnée initiale assure qu'il existe $\delta > 0$ tel que :

$$\|q - q'\| < \delta \Rightarrow \|\varphi(T, q) - \varphi(T, q')\| < \eta_y.$$

L'inégalité triangulaire assure alors que pour $\|q - q'\| < \delta$, on a : $\varphi(T, q') \in B(y, 2\eta_y) \subset U^y$. L'étape 3 assure alors que :

$$\varphi(t, q') \underset{t \to +\infty}{\to} y.$$

Ainsi $B(q, \delta)$ est inclus dans W^y, qui est donc ouvert. L'égalité $(*)$ ainsi que la connexité de $\mathbb{R}^n$ assurent que :

$$\text{Card } f^{-1}(\{0\}) = 1,$$

ce qui achève la preuve.

$\square$

Remarque 3.9. *Le théorème reste vrai dans le cas où f est seulement supposée de classe $\mathcal{C}^1$ mais la preuve est plus compliquée (on ne peut pas directement utiliser le théorème de Cauchy-Lipschitz).*

Application 3.10. *La fonction exp : $\mathcal{M}_n(\mathbb{R}) \to GL_n(\mathbb{R})$ est un difféomorphisme local entre un voisinage de 0 et un voisinage de I_n. En effet la différentielle de exp est 0 l'identité de $\mathcal{L}(\mathcal{M}_n(\mathbb{R}))$ qui est bien inversible.*

Une application de ce résultat est qu'il existe un voisinage V de I_n dans $GL_n(\mathbb{R})$ tel que si G est un sous-groupe de $GL_n(\mathbb{R})$ inclus dans V, alors $G = \{I_n\}$.

L'énoncé du théorème suivant ne parle pas d'applications différentiables, cependant dans la preuve on se ramène à montrer le théorème avec une hypothèse de régularité en plus sur f, ce qui permet d'utiliser toute la puissance du calcul différentiel.

Théorème 3.11 (*Théorème du point fixe de Brouwer*)

On note $\overline{B}(0, 1)$ la boule unité fermée de $\mathbb{R}^n$ pour n'importe quelle norme. Soit $f : \overline{B}(0, 1) \to \overline{B}(0, 1)$ une application continue. Alors f admet un point fixe.

3.2 Le théorème des fonctions implicites

Théorème 3.12 (*Théorème des fonctions implicites*)

Soit U un ouvert de $\mathbb{R}^n \times \mathbb{R}^m$ et $f \in \mathcal{C}^1(U, \mathbb{R}^m)$. Soit $(a, b) \in U$ tel que $f(a, b) = 0$. On note f_y l'application partielle $f(a, .)$ qui est différentiable en b. Supposons que $df_y(b) \in GL(\mathbb{R}^m)$. Alors il existe V un voisinage de a dans $\mathbb{R}^n$, W un voisinage de b dans $\mathbb{R}^m$, avec $V \times W \subset U$, et une application $\phi \in \mathcal{C}^1(V, W)$ tels que :

$$(x \in V,\ y \in W,\ f(x, y) = 0) \Leftrightarrow (x \in V,\ y = \phi(x))$$

De plus si f est de classe $\mathcal{C}^p$, alors ϕ est également de classe $\mathcal{C}^p$.

Application 3.13. *On considère un système de deux équations à deux inconnues x et y paramétré par un paramètre $t \in \mathbb{R}$:*

$$\begin{cases} x &= \frac{1}{2}\sin(x + y) + t - 1 \\ y &= \frac{1}{2}\cos(x - y) - t + \frac{1}{2} \end{cases}.$$

Ce système admet une unique solution $t \mapsto (x(t), y(t))$ qui est de classe $\mathcal{C}^\infty$.

L'existence résulte du théorème de point fixe de Banach et la régularité du théorème des fonctions implicites.

Application 3.14. *On considère P un polynôme de $\mathbb{R}_n[X]$ qui admet une racine simple en x_0. Alors il existe un voisinage V de P dans $\mathbb{R}_n[X]$ et une application ϕ une application $\mathcal{C}^\infty$ entre V et U un voisinage de x_0 dans $\mathbb{R}$, tel que pour tout $Q \in V$, $\phi(Q)$ est une racine simple de Q.*

On peut ainsi montrer que l'ensemble des polynômes scindés à racines simples de $\mathbb{R}_n[X]$ est un ouvert de $\mathbb{R}_n[X]$.

3.3 Applications en géométrie

La notion de sous-variété de $\mathbb{R}^n$ a été introduite pour généraliser le calcul différentiel sur des espaces qui ne sont pas des ouverts d'espaces vectoriels. Ce sont justement les espaces qui sont localement un ouvert d'un espace vectoriel à déformation près par un difféomorphisme. On voit donc tout l'intérêt de la notion de difféomorphisme. La proposition suivante donne des définitions équivalentes d'une sous-variété de $\mathbb{R}^n$.

Proposition-Définition 3.15

Soit $p \in \mathbb{N}^* \cup \{\infty\}$ et $k \in \{1, \ldots, n\}$. Un sous-ensemble $N \subset \mathbb{R}^n$ est une sous-variété de $\mathbb{R}^n$ de dimension k et de classe $\mathcal{C}^p$, si pour tout $x_0 \in N$, il existe un voisinage de W de x_0 dans $\mathbb{R}^n$ tel que l'une des condition suivante soit vérifiée :

1. ***(Carte locale)*** *Il existe $\mathcal{C}^p$-difféomorphisme $\phi : W \to \phi(W) \subset \mathbb{R}^n$ tel que :*
$$\phi(N \cap W) = \mathbb{R}^k \times \{0\}^{n-k} \cap \phi(W).$$
2. ***(Graphe)*** *Il existe $u : \mathbb{R}^k \to \mathbb{R}^{n-k}$ une application de classe $\mathcal{C}^p$ et $A \in GL_n(\mathbb{R})$ tels que :*
$$\left\{A(z, u(z))\, z \in \mathbb{R}^k\right\} \cap W = W \cap N.$$
3. ***(Équations)*** *Il existe $F : W \to \mathbb{R}^{n-k}$ de classe $\mathcal{C}^p$, telle que dF_{x_0} est surjective et :*
$$W \cap N = F^{-1}(\{0\}).$$
4. ***(Nappe paramétrée)*** *Il existe un voisinage U de 0 dans $\mathbb{R}^k$ et $j : U \to \mathbb{R}^n$ de classe $\mathcal{C}^p$ telle que :*
   (a) $j(0) = x_0$
   (b) *dj_0 est injective*
   (c) *$j : U \to N \cap W$ est un homéomorphisme.*

Exemple 3.16. 1. *La parabole $P := \{(x, x^2),\, x \in \mathbb{R}\}$ est une sous-variété de $\mathbb{R}^2$, de classe $\mathcal{C}^\infty$ et de dimension 1.*

2. *La sphère $S^{n-1} := \{x \in \mathbb{R}^n,\, \|x\| = 1\}$, où $\|.\|$ désigne la norme euclidienne usuelle, est une sous-variété de $\mathbb{R}^n$ de classe $\mathcal{C}^\infty$ et de dimension $n-1$.*

Définissons maintenant la notion d'espace tangent, qui généralise la notion de tangente à une courbe en dimension 1.

Définition 3.17

L'espace tangent à N en x_0 est
$$T_{x_0}N = \{\gamma'(0), \gamma \in \mathcal{C}^1(I, \mathbb{R}^n),\ I \text{ un intervalle ouvert contenant 0, } \gamma(I) \subset N \text{ et } \gamma(0) = x_0\}.$$

Proposition 3.18

On a les égalités suivantes :

1. ***Carte locale*** $T_{x_0}N = d\varphi(x_0)^{-1}.(\mathbb{R}^k \times \{0^{n-k}\}) = d\varphi^{-1}(\varphi(x_0)).(\mathbb{R}^k \times \{0^{n-k}\})$.
2. ***Graphe*** $T_{x_0}N = \{A(h, du(z_0).h), h \in \mathbb{R}^k\}$.
3. ***Équations*** $T_{x_0}N = \ker(dF(x_0))$.
4. ***Nappe paramétrée*** $T_{x_0}N = dj(0)(\mathbb{R}^k)$.

Remarquons que l'expression 3. montre que l'espace tangent est un espace vectoriel de dimension k (la dimension de la sous-variété).

Exemple 3.19. *Le folium de Descartes est défini par :*
$$C := \{(x, y) \in \mathbb{R}^2,\, x^3 + y^3 - 3xy = 0\}.$$

C n'est pas une sous-variétés de $\mathbb{R}^2$, en revanche, $C\backslash\{0\}$ est une sous-variété de $\mathbb{R}^2$ de dimension 1 et de classe $\mathcal{C}^\infty$. L'espace tangent en un point de $C\backslash\{0\}$ est donc une droite.

1. *Si $(a,b) = (2^{\frac{2}{3}}, 2^{\frac{1}{3}}) \in C\backslash\{0\}$, la tangente est une droite verticale (pente infinie).*
2. *Sinon, la tangente en $(a,b) \in C\backslash\{(0,0) \cup (2^{\frac{2}{3}}, 2^{\frac{1}{3}})\}$ est la droite d'équation $y = b + \frac{b-a^2}{b^2-a}(x-a)$.*

Les deux énoncés suivants s'inscrivent dans l'idée de la classification des fonctions et courbes à difféomorphismes près.

Théorème 3.20 (*Théorème du rang constant*)

Soit U un ouvert de $\mathbb{R}^n$, $a \in U$, et $f : U \to \mathbb{R}^p$ une application de classe $\mathcal{C}^1$. On suppose que le rang de la différentielle de f en tout point de U est constant égal à r. Alors il existe :

1. *ϕ un $\mathcal{C}^1$-difféomorphisme d'un voisinage V de 0 dans $\mathbb{R}^n$ sur un voisinage de a dans U, tel que $\phi(0) = a$.*
2. *ψ un $\mathcal{C}^1$-difféomorphisme d'un ouvert de $\mathbb{R}^p$ contenant $f(\phi(V))$ sur un ouvert de $\mathbb{R}^p$*

tels que :

$$\forall x \in V,\ \psi \circ f \circ \phi(x_1, \ldots, x_n) = (x_1, \ldots, x_r, 0, \ldots).$$

Remarque 3.21. *Ce théorème s'applique en particulier dans le cas où le rang de la différentielle est maximal en un point, en effet par semi-continuité inférieure du rang, le rang restera localement maximal. On peut ainsi retrouver la forme "usuelle" des immersions et des submersions.*

Théorème 3.22 (*Lemme de Morse*)

Soit U un ouvert de $\mathbb{R}^n$ contenant 0 et $f : U \to \mathbb{R}$ de classe $\mathcal{C}^3$. On suppose également que $df_0 = 0$ et que la Hessienne $d^2 f_0$ est non dégénérée (ie. inversible) et de signature $(p, n-p)$. Alors il existe ϕ un $\mathcal{C}^1$-difféomorphisme entre deux voisinages V_0 et $W_0 = \phi(V_0)$ de 0 dans $\mathbb{R}^n$ tel que $\phi(0) = 0$ et :

$$\forall x \in V_0,\ f(x) - f(0) = \phi_1(x)^2 + \cdots + \phi_p(x)^2 - \phi_{p+1}(x)^2 - \cdots - \phi_n(x)^2.$$

Pour démontrer ce résultat, nous allons avoir besoin d'un lemme.

Lemme 3.23

On se donne $A_0 \in \mathcal{S}_n(\mathbb{R})$ une matrice symétrique et inversible. Alors il existe V un voisinage de A_0 dans $\mathcal{S}_n(\mathbb{R})$ et une application $\rho \in \mathcal{C}^1(V, GL_n(\mathbb{R}))$ telle que :

$$\forall A \in V, \quad A = {}^t(\rho(A)) A_0 \rho(A).$$

Démonstration : On considère l'application :

$$f : \left\{ \begin{array}{ll} \mathcal{M}_n(\mathbb{R}) & \to \mathcal{S}_n(\mathbb{R}) \\ M & \mapsto {}^t M A_0 M \end{array} \right. .$$

Il s'agit d'une application de classe $\mathcal{C}^1$ car les coordonnées à l'arrivée sont polynômiales en celles de M. Calculons la différentielle de f en I_n. Soit $H \in \mathcal{M}_n(\mathbb{R})$. On a :

$$f(I_n + H) \underset{H \to 0}{=} A_0 + {}^t H A_0 + A_0 H + o(\|H\|),$$

puisque A_0 est symétrique. Ainsi :

$$\forall H \in \mathcal{M}_n(\mathbb{R}), \quad df_{I_n}.H = {}^t H A_0 + A_0 H.$$

On en déduit que $\ker df_{I_n} = A_0^{-1} \mathcal{A}_n(\mathbb{R})$, qui est de dimension $\frac{n(n-1)}{2}$. On considère un supplémentaire F de $\ker df_{I_n}$ dans $\mathcal{M}_n(\mathbb{R})$, et ψ la restriction de f à F. Puisque $I_n \in F$, la différentielle de ψ en I_n est la restriction à F de la différentielle de f en I_n.

Par choix de F, $d\psi_{I_n}$ est injective et donc inversible par le théorème du rang (puisque la dimension de $\mathcal{S}_n(\mathbb{R})$ est égale à la dimension de F). Le théorème d'inversion locale assure qu'il existe U un voisinage de I_n dans F, qu'on peut supposer inclus dans $\mathrm{GL}_n(\mathbb{R})$ puisque I_n est inversible et que $\mathrm{GL}_n(\mathbb{R})$ est un ouvert, et V un voisinage de A_0 dans $\mathcal{S}_n(\mathbb{R})$ tels que ψ soit un $\mathcal{C}^1$-difféomorphisme de U sur V. L'application $\rho := \psi^{-1}$ convient puisqu'on a alors :

$$\forall A \in V, \quad A = \psi(\rho(A)) = {}^t(\rho(A))A_0\rho(A).$$

□

On peut maintenant passer à la preuve du lemme de Morse.

Démonstration : Quitte à restreindre U, on peut supposer que U est une boule ouverte centrée en 0. On peut ainsi utiliser la formule de Taylor avec reste intégral à l'ordre 2 (ce qui est possible puisque les boules sont convexes). Ainsi pour tout $x \in U$:

$$f(x) - f(0) = \int_0^1 (1-t)\,{}^t x \,\mathrm{Hess} f(tx)\, x\, dt,$$

puisque 0 est un point critique de f. En posant $Q(x) := \displaystyle\int_0^1 (1-t)\,\mathrm{Hess} f(tx)\, dt \in \mathcal{S}_n(\mathbb{R})$, on a :

$$\forall x \in U, \quad f(x) - f(0) = {}^t x\, Q(x)\, x.$$

De plus Q est de classe $\mathcal{C}^1$ d'après le théorème de régularité sous l'intégrale, qui s'applique puisque f est supposée $\mathcal{C}^3$. Or un calcul direct donne $Q(0) = \frac{1}{2}\mathrm{Hess} f(0)$ qui est inversible et de signature $(p, n-p)$ par hypothèse. Le lemme précédent assure alors qu'il existe V un voisinage de $Q(0)$ dans $\mathcal{S}_n(\mathbb{R})$ et une application $\rho \in \mathcal{C}^1(V, \mathrm{GL}_n(\mathbb{R}))$ tels que :

$$\forall A \in V, \quad A = {}^t(\rho(A))\, Q(0)\, \rho(A).$$

La continuité de Q et le fait qu'elle soit à valeurs $\mathcal{S}_n(\mathbb{R})$ assure qu'il existe un voisinage W de 0 dans $\mathbb{R}^n$ tel que $Q(W) \subset V$. Il s'ensuit que :

$$\forall x \in W, \quad f(x) - f(0) = {}^t(\rho(Q(x))x)\, Q(0)\, \rho(Q(x))x.$$

Le théorème d'inertie de Sylvester assure qu'il existe $A \in \mathrm{GL}_n(\mathbb{R})$ telle que $Q(0) = {}^t A\, \Delta_{p,n-p}\, A$, où :

$$\Delta_{p,n-p} := \begin{pmatrix} I_p & 0 \\ 0 & -I_{n-p} \end{pmatrix}.$$

Si on pose $\phi : x \in W \mapsto A\rho(Q(x))x \in \mathbb{R}^n$, on définit une application de classe $\mathcal{C}^1$ par composition de telles applications. De plus on a :

$$\forall x \in W, \quad f(x) - f(0) = {}^t(\phi(x))\, \Delta_{p,n-p}\, \phi(x) \quad (*).$$

Calculons la différentielle de ϕ en 0. Si $h \in W$:

$$\phi(h) \underset{h \to 0}{=} A\rho(Q(h))h = A\rho(Q(0))h + o(\|h\|).$$

Ainsi la jacobienne de ϕ en 0 est la matrice $A\rho(Q(0)) = A \in \mathrm{GL}_n(\mathbb{R})$. Le théorème d'inversion locale assure que ϕ est un $\mathcal{C}^1$-difféomorphisme entre deux ouverts V_0 et $\phi(V_0)$ voisinages de 0. De plus l'égalité $(*)$ assure que :

$$\forall x \in V_0,\ f(x) - f(0) = \phi_1(x)^2 + \cdots + \phi_p(x)^2 - \phi_{p+1}(x)^2 - \cdots - \phi_n(x)^2.$$

□

4 Optimisation et calcul différentiel

L'optimisation des applications différentiables est une problématique au coeur du calcul différentiel. Nous allons en illustrer quelques aspects, en distinguant l'optimisation dite sans contraintes, c'est-à-dire lorsque l'espace de départ est un ouvert d'un espace vectoriel, de l'optimisation sous contraintes, c'est-à-dire lorsque l'espace de départ est une sous-variétés de $\mathbb{R}^n$.

4.1 Optimisation sans contraintes : conditions nécessaires et suffisante d'optimalité

On se donne Ω un ouvert de $\mathbb{R}^n$ avec $a \in \Omega$ et $f : \Omega \to \mathbb{R}$ une application différentiable.

Définition 4.1

1. *On dit que f admet un extremum local en a si $f(a)$ est la valeur minimale ou maximale de f sur un voisinage de a.*
2. *On dit que f admet un extremum global en a si $f(a)$ est la valeur minimale ou maximale que prend f sur tout Ω.*

Définition 4.2 (*Point critique*)

Un point $a \in \Omega$ est un point critique de f si $df_a = 0$.

On a une première condition nécessaire d'optimalité du premier ordre.

Proposition 4.3

Si f admet un extremum local en a, alors a est un point critique de f.

Remarque 4.4. *Le caractère ouvert de Ω est essentiel. Pour s'en convaincre il suffit de considérer l'application $x \in [0,1] \mapsto x$ qui atteint son maximum en 1 qui n'est évidemment pas un point critique. D'autres contre-exemples génériques de ce phénomène sont les applications holomorphes non constantes grâce au principe du maximum.*

Proposition 4.5 (*Conditions nécessaires d'optimalité du second ordre*)

On suppose que f est deux fois différentiable en a. Alors :

1. *Si f admet un minimum local en a, alors $df_a = 0$ et $\mathrm{Hess}f(a)$ est positive (en tant que matrice symétrique).*
2. *Si f admet un maximum local en a, alors $df_a = 0$ et $\mathrm{Hess}f(a)$ est négative (en tant que matrice symétrique).*

Proposition 4.6 (*Conditions suffisantes d'optimalité du second ordre*)

1. *Si $\mathrm{Hess}f(a)$ est une matrice définie positive, alors f atteint un minimum local en a.*
2. *Si $\mathrm{Hess}f(a)$ est une matrice définie négative, alors f atteint un maximum local en a.*

Remarque 4.7. *Le caractères positif ou négatif de la Hessienne se lit sur son spectre. Dans le cas où la Hessienne n'est pas inversible, on ne peut a priori rien dire et il faut pousser le développement limité. Dans le cas où elle a des valeurs propres strictement positives et strictement négatives, $f(a)$ n'est pas un extremum local de f : f croît selon les directions propres associées aux valeurs propres strictement positives (de la Hessienne), et décroît selon les directions propres associées aux valeurs propres strictement négatives. Le lemme de Morse précise ce fait.*

A priori, les conditions précédentes proposent des critères concernant les extrema locaux. Dans le cas des fonctions convexes différentiables, les extrema sont en fait globaux.

Proposition 4.8

On suppose Ω convexe et f convexe différentiable. Il y a équivalence entre :

1. *f atteint son minimum global en a.*
2. *$df_a = 0$.*

Citons deux autres applications de la recherche d'extrema grâce au calcul différentiel : l'une en statistiques et l'autre pour résoudre des systèmes linéaires.

Application 4.9. *En statistiques, on peut obtenir des estimateurs par la méthode de maximisation de la vraisemblance. L'heuristique de la méthode étant de maximiser la " probabilité " d'obtenir l'observation qu'on a par rapport au paramètre qu'on désire estimer. Sous réserve d'existence et d'unicité, pour trouver l'estimateur du maximum de vraisemblance, on cherche d'abord à annuler le gradient de la vraisemblance, ou de son logarithme. Et on vérifie qu'il s'agit bien d'un maximum. On peut par exemple citer le cas du modèle gaussien où la moyenne et l'écart-type sont inconnus. Les estimateurs du maximum de vraisemblence de la moyenne m et de la variance σ^2 sont :*

$$\widehat{m_n} = \overline{X_n} := \frac{1}{n}\sum_{k=1}^{n} X_k \qquad \widehat{\sigma_n^2} = S_n^2 := \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X_n})^2.$$

L'application suivante permet de résoudre un système linéaire en minimisant une fonctionnelle quadratique.

Application 4.10. *Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice symétrique définie positive et soit $b \in \mathbb{R}^n$. La fonctionnelle quadratique*

$$f : \left\{ \begin{array}{ll} \mathbb{R}^n & \to \mathbb{R} \\ x & \mapsto \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle \end{array} \right.$$

atteint un unique minimum en $x = A^{-1}b$.

Si A est seulement positive, non définie positive, alors f admet un minimum global si et seulement si $b \in Im(A)$.

4.2 Optimisation sous contraintes

Le théorème fondamental est le suivant.

Théorème 4.11

Soit $f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R})$ et N une sous-variétés de $\mathbb{R}^n$. Si la restriction de f à N admet un extremum local en un point $x_0 \in N$, alors la différentielle de f en x_0 s'annule sur l'espace tangent $T_{x_0}N$.

Corollaire 4.12 (*Théorème des extrema liés*)

Soient W un ouvert de $\mathbb{R}^n$, $x_0 \in W$, $f^1, \dots, f^{n-k} \in \mathcal{C}^1(W, \mathbb{R})$ tels que $df^1_{x_0}, \dots, df^{n-k}_{x_0}$ soient des formes linéaires sur $\mathbb{R}^n$ linéairement indépendantes et $N = \{x \in W,\ f^1(x) = \dots = f^{n-k}(x) = 0\}$. Si $g \in \mathcal{C}^1(W, \mathbb{R})$ et $g_{|N}$ admet un extremum local en x_0 alors il existe des multiplicateurs de Lagrange $\lambda_1, \dots, \lambda_{n-k} \in \mathbb{R}$ tels que

$$dg_{x_0} = \lambda_1 df^1_{x_0} + \dots + \lambda_{n-k} df^{n-k}_{x_0}.$$

Application 4.13 (Théorème spectral)**.** *On se donne un endomorphisme autoadjoint $u \in \mathcal{L}(\mathbb{R}^n)$. L'application :*

$$\phi : \left\{ \begin{array}{ll} \mathbb{R}^n & \to \mathbb{R} \\ x & \mapsto \langle x, u(x)\rangle \end{array} \right.$$

atteint son maximum sur la sphère S^{n-1} en vecteur x_0. Le théorème des extrema liés assure que le gradient de ϕ en x_0 est orthogonal à l'hyperplan tangent à la sphère en x_0, c'est-à-dire au plan de direction $x_0^\perp$. Cela assure que x_0 est un vecteur propre de u. Ce dernier étant autoadjoint, son orthogonal est stable par u. Une récurrence sur la dimension permet de conclure.

5 Questions du jury

1. **Interprétation du jacobien dans la formule du changement de variables** (Voir l'exercice 29 page 89 de [1])

   On se donne f un $\mathcal{C}^1$-difféomorphisme entre deux ouverts U et V de $\mathbb{R}^n$ qui contiennent l'origine. Quitte à translater, on peut supposer $f(0) = 0$. Il s'agit de montrer, en notant λ la mesure de Lebesgue sur $\mathbb{R}^n$ que :
$$|\det df_0| = \lim_{r\to 0} \frac{\lambda(f(B(0,r)))}{\lambda(B(0,r))}.$$

   Solution :

   Étape 1 : On se donne $\epsilon \in]0,1[$. L'application $x \mapsto df_0^{-1}.f(x) - x$ est différentiable en 0 et le théorème des fonctions composées assure que sa différentielle y est nulle. Par Taylor-Young, on dispose de $R > 0$ tel que :
$$\forall x \in B(0,R), \quad \|df_0^{-1}.f(x) - x\| \leq \epsilon\|x\|.$$

   Étape 2 : Notons $A = df_0$. Montrons qu'il existe $R' > 0$ tel que pour tout $r \in [0, R']$:
$$(1-\epsilon)A(B(0,r)) \subset f(B(0,r)) \subset (1+\epsilon)A(B(0,r)).$$
   D'après l'étape 1, si $r \leq R$ et si $x \in B(0,r)$, alors par inégalité triangulaire :
$$\|A^{-1}f(x)\| \leq (1+\epsilon)r,$$
   cela donne la seconde inclusion. Pour la première inclusion, on considère l'application $g : x \mapsto f^{-1}\left(\frac{1}{1+\epsilon}Ax\right)$, qui est, tout comme f, un $\mathcal{C}^1$-difféomorphisme entre deux voisinages de 0. De plus g vérifie :
$$g(0) = 0 \quad \text{et} \quad dg_0 = \frac{1}{1+\epsilon}\mathrm{Id}.$$
   Le point précédent assure qu'il existe $R' > 0$ tel que pour tout $r \leq R'$:
$$g(B(0,r)) \subset (1+\epsilon)dg_0(B(0,r)) = B(0,r).$$
   On a donc :
$$(1-\epsilon)A(B(0,r)) \subset \frac{1}{1+\epsilon}A(B(0,r)) \subset f(B(0,r)),$$
   d'après ce qui précède.

   Étape 3 : On conclut : l'étape précédente assure que pour tout $r \in [0, R]$:
$$\lambda((1-\epsilon)A(B(0,r))) \leq \lambda(f(B(0,r))) \leq ((1+\epsilon)A(B(0,r))).$$
   Or si $c > 0$, on a :
$$\lambda(cA(B(0,r))) = c^n\lambda(A(B(0,r))) = c^n|\det A|\lambda(B(0,r)),$$
   la dernière égalité se démontrant pour les dilatations et les transvections, ce qui suffit puisque de telles matrices engendrent $\mathrm{GL}_n(\mathbb{R})$. En combinant les résultats précédents, on obtient pour tout $r \in [0, R]$:
$$(1-\epsilon)^n|\det A| \leq \frac{\lambda(f(B(0,r)))}{\lambda(B(0,r))} \leq (1+\epsilon)^n|\det A|.$$
   En prenant les limites supérieure et inférieure quand $r \to 0$ dans les inégalités précédentes et en laissant tendre ϵ vers 0, on obtient la convergence voulue.

2. **Donner une interprétation du rotationnel**
   Voir l'exercice 31 page 95 de [1]

3. **Paramétrisation des racines d'un polynômes de degré 3** (Voir l'exercice 79 page 245 de [1]).

   On considère deux réels $a < b$ et on note la fonction polynômiale du troisième degré définie pour $\epsilon > 0$:

   $$f(x, \epsilon) = (x-a)(b-x) + \epsilon x^3.$$

   Alors pour ϵ assez petit, ce polynôme est scindé à racines simples et on peut donner un développement limité des 3 racines.

   Solution : On va appliquer le théorème des fonctions implicites. On commence par le comportement local autour de a. On a $f(a, 0) = 0$. De plus, on a :

   $$\partial_x f(x, \epsilon) = a + b - 2x + 3\epsilon x^2.$$

   Ainsi $\partial_x f(a, 0) = b - a > 0$. Le théorème des fonctions implicites assure qu'il existe V_1 un voisinage de 0 et W_1 un voisinage de a et une fonction $x_1 : V_1 \to W_1$ de classe $\mathcal{C}^\infty$ telle que :

   $$(\epsilon \in V_1,\, x \in W_1 \text{ et } f(x, \epsilon) = 0) \Leftrightarrow (\epsilon \in V_1,\, x = x_1(\epsilon)).$$

   On peut faire de même autour de b. Ainsi pour ϵ assez petit, le polynôme de degré 3 admet deux racines réelles, donc à fortiori toutes ses racines sont réelles. On peut donner un développement asymptotique de la racine proche de a. Comme x_1 est $\mathcal{C}^\infty$ elle admet un développement limité à tout ordre. Or $x_1(0) = a$ et en différentiant par rapport à ϵ la relation $f(x_1(\epsilon), \epsilon) = 0$, qui est valable pour $\epsilon \in V_1$, on a :

   $$\forall \epsilon \in V_1, \quad \left(a + b - 2x_1(\epsilon) + 3x_1(\epsilon)^2\right) x_1'(\epsilon) + x_1(\epsilon)^3 = 0.$$

   En prenant $\epsilon = 0$, on obtient : $x_1'(0) = \frac{-a^3}{b-a}$ et on a donc le développement limité :

   $$x_1(\epsilon) \underset{0}{=} a - \frac{a^3}{b-a}\epsilon + O(\epsilon^2).$$

   On peut obtenir de même un développement limité de la racine autour de b et pour la troisième racine, on utilise les relations coefficients/racines. Remarquons qu'en différentiant à nouveau la relation, on peut pousser plus loin le développement limité.

4. **Le groupe $SL_n(\mathbb{R})$ est-il une sous-variété de $\mathbb{R}^{n^2}$?**

   Solution : Il s'agit bien d'une sous-variété de $\mathbb{R}^{n^2}$. Pour le voir, on considère l'application

   $$\phi : \begin{cases} \mathcal{M}_n(\mathbb{R}) & \to \mathbb{R} \\ M & \mapsto \det(M) - 1 \end{cases}.$$

   Il s'agit d'une application $\mathcal{C}^\infty$ puisque le déterminant est une application polynômiale. De plus $SL_n(\mathbb{R}) = \phi^{-1}(\{0\})$. On vérifie qu'il s'agit bien d'une sous-variété de classe $\mathcal{C}^\infty$ et de dimension $n^2 - 1$ grâce à la définition d'une sous-variétés par des équations. Il s'agit de voir que la différentielle de ϕ en $M \in SL_n(\mathbb{R})$ est surjective. Or on a :

   $$\forall H \in \mathcal{M}_n(\mathbb{R}), \quad d\phi_M.H = \mathrm{Tr}({}^t\mathrm{Com}(M)H).$$

   Ainsi $d\phi_M$ est surjective puisque ${}^t\mathrm{Com}(M)$ est non nulle comme le montre la formule de la comatrice puisque M est inversible.

-
  -
  -
  -
  -
  -

Références

[1] F. Rouviere, *Petit guide du calcul différentiel à l'usage de la licence et de l'agrégation*, Cassini, 2014.

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

[3] V. Beck, J. Malick, G. Peyré, *Objectif agrégation*, Broché, 2005.