
Compléments sur le calcul différentiel

Miguel Rodrigues

Le matériel exposé ici relève classiquement du calcul différentiel [BG14, Rou15].

Dans tout ce qui suit, on munit les espaces de leurs normes euclidiennes canoniques. On notera $\mathcal{L}(\mathbf{R}^n; \mathbf{R}^p)$ l'ensemble des applications linéaires (continues) de \mathbf{R}^n dans \mathbf{R}^p . Enfin on identifiera les vecteurs de \mathbf{R}^n , $n \in \mathbf{N}^*$, avec des vecteurs colonnes. Par ailleurs, on rappelle qu'une application qui s'écrit comme somme d'une application constante et d'une application linéaire est dite affine.

Même lorsque le corps de base est *a priori* \mathbf{C} , on identifiera \mathbf{C} à \mathbf{R}^2 et l'on ne considèrera la différentiabilité que sous l'angle de la \mathbf{R} -linéarité. Pour ce qui concerne la \mathbf{C} -linéarité et la notion associée d'holomorphie — beaucoup moins souple que celle de différentiabilité —, on renvoie le lecteur vers un cours d'analyse complexe.

1 Différentiel

1.1 Préliminaires

De manière informelle : pour une fonction raisonnable, on s'attend à ce qu'autour d'un point donné, une petite variation au départ corresponde essentiellement à une petite variation proportionnelle à l'arrivée. La bonne définition pour « proportionnelle » est la linéarité.

Évidemment, même localement, une application, même raisonnable, n'est affine qu'approximativement. Pour rendre cela précis, on va rappeler ce que veulent dire localement et approximativement.

Définition 1 *On dit qu'une propriété est vraie au voisinage d'un point $A \in \mathbf{R}^n$ s'il existe une boule (non triviale) centrée en A telle que la propriété soit vraie en tout point de la boule.*

Remarquons que pour $n = 1$ une boule est en fait un intervalle.

Pour définir ce qui est négligeable et ce qui ne l'est pas, il faut savoir comparer des fonctions.

Définition 2 Notations de Landau.

Soit $f : \Omega \rightarrow \mathbf{R}^p$ et $g : \Omega \rightarrow \mathbf{R}$ (où $\Omega \subset \mathbf{R}^n$).

1. On dit que f est négligeable devant g au voisinage de $A \in \Omega$ s'il existe une fonction $\varepsilon : \Omega \rightarrow \mathbf{R}^p$ tendant vers 0 en A et telle qu'au voisinage de A l'on ait $\forall X \in \Omega$, $f(X) = g(X)\varepsilon(X)$. On note alors

$$f(X) \stackrel{A}{\equiv} o(g(X)).$$

2. On dit que f est dominée par g au voisinage de $A \in \Omega$ s'il existe une fonction $M : \Omega \rightarrow \mathbf{R}^p$ bornée telle qu'au voisinage de A l'on ait $\forall X \in \Omega, f(X) = g(X)M(X)$. On note alors

$$f(X) \stackrel{A}{=} \mathcal{O}(g(X)).$$

Remarque 3 Une fois que l'on a négligé ou borné un terme, on ne peut plus revenir en arrière dans une série d'égalités.

Exemple 4 $f(X) \stackrel{A}{=} o(1)$ signifie que f tend vers 0 en A .
En particulier f est continue en A si $f(X) \stackrel{A}{=} f(A) + o(1)$.

Exemples 5 Avec $n = 1 : x^2 \stackrel{0}{=} o(x), \sin(x) \stackrel{0}{=} \mathcal{O}(1)$.

1.2 Différentielle

Définition 6 Soit $f : \Omega \rightarrow \mathbf{R}^p$ (où $\Omega \subset \mathbf{R}^n$) et $A \in \Omega$ tel que Ω soit un voisinage de A .

1. On dit que f est différentiable en A s'il existe $\ell : \mathbf{R}^n \rightarrow \mathbf{R}^p$ linéaire (continue) telle que

$$f(X) \stackrel{A}{=} f(A) + \ell(X - A) + o(\|X - A\|).$$

On note alors $\ell = df(A)$. La matrice associée $\text{Jac}(f)(A)$ est appelée matrice jacobienne ou simplement jacobienne.

2. Quand f est différentiable en tout point, on dit simplement que f est différentiable et l'on note $df : \Omega \rightarrow \mathcal{L}(\mathbf{R}^n; \mathbf{R}^p)$ sa différentielle.

Remarque 7 Il est souvent plus commode de se ramener en $(0, \dots, 0)$:

$$f(A + H) \stackrel{(0, \dots, 0)}{=} f(A) + df(A)(H) + o(\|H\|).$$

Remarque 8 Quand $n = 1$, on retrouve la notion de dérivabilité avec $df(a) : \mathbf{R} \rightarrow \mathbf{R}^p, h \mapsto h f'(a)$.

Remarque 9 Un autre cas particulier intéressant est celui où $p = 1$, car toute application linéaire de \mathbf{R}^n vers \mathbf{R} (appelée forme linéaire) s'écrit comme un produit scalaire avec un certain vecteur. Le vecteur correspondant à la différentielle est appelé gradient et noté $\nabla f(A)$, il est relié à la différentielle par, pour tout H ,

$$df(A)(H) = \langle \nabla f(A), H \rangle.$$

Notons qu'alors que la notion de différentielle est intrinsèque, celle de matrice jacobienne ou de gradient dépend d'un choix, de bases ou de produit scalaire, que l'on ne précise pas quand, comme ici, on procède à des choix canoniques.

Observation 10 L'interprétation en terme de produit scalaire donne des informations géométriques intuitives : si l'on veut augmenter la valeur de f sans trop se déplacer il faut se déplacer dans la direction du gradient, si l'on veut diminuer la valeur de f sans trop se déplacer il faut se déplacer dans la direction opposée, les considérations précédentes s'appuyant sur l'optimisation de $\langle \nabla f(A), H \rangle$ en H à longueur de H fixé. De même, si l'on veut ne pas changer f il faut se déplacer perpendiculairement au gradient. Ainsi les espaces tangents des ensembles de niveaux (ensemble de points prenant une même valeur) sont en tout point orthogonaux au gradient. De plus, si f atteint sur un voisinage de A son minimum (ou son maximum) en A (on dit que f atteint un minimum (ou maximum) local), alors $\nabla f(A) = (0, \dots, 0)$, on dit que A est un point critique de f .

Exemple 11 $f : \mathbf{R}^2 \rightarrow \mathbf{R}, (x, y) \mapsto xy$.

$$\begin{aligned} f(A+H) &= (x_A + x_H)(y_A + y_H) \\ &= x_A y_A + x_H y_A + x_A y_H + x_H y_H \\ &= x_A y_A + x_H y_A + x_A y_H + o(\|(x_H, y_H)\|). \end{aligned}$$

Donc l'application f est différentiable en $A = (x_A, y_A)$ et sa différentielle en A , $df(A) : \mathbf{R}^2 \rightarrow \mathbf{R}$, est donnée par, pour $H = (x_H, y_H)$, $df(A)(H) = x_H y_A + x_A y_H$. Le gradient en A est donc $\nabla f(A) = (y_A, x_A)$.

Exemple 12 L'exemple précédent se généralise immédiatement à toutes les applications bilinéaires et plus généralement aux applications multilinéaires.

Si $B : (\mathbf{R}^n)^k \rightarrow \mathbf{R}^p$ est une application k -linéaire (continue) et que l'on définit $f : \mathbf{R}^n \rightarrow \mathbf{R}^p, X \mapsto B(X, \dots, X)$, alors f est différentiable et sa différentielle en un point A est donnée par, pour tout H ,

$$df(A)(H) = \sum_{j=1}^k B(\underbrace{A, \dots, A}_{j-1}, H, \underbrace{A, \dots, A}_{k-j}).$$

Dans le cas particulier où $p = 1, k = 2$ et B est le produit scalaire, alors f est le carré de la norme et l'on obtient $df(A)(H) = \langle A, H \rangle + \langle H, A \rangle = 2\langle A, H \rangle$ donc $\nabla f(A) = 2A$.

Exemple 13 Soit $d \in \mathbf{N}^*$. Posons $\Omega = Gl_d(\mathbf{R})$ l'ensemble des matrices inversibles ($n = p = d^2$) et définissons

$$f : Gl_d(\mathbf{R}) \rightarrow Gl_d(\mathbf{R}), \quad A \mapsto A^{-1}.$$

Pour tout $A \in Gl_d(\mathbf{C})$ et tout $H \in \mathcal{M}_d(\mathbf{C})$ tel que $\|H\| < 1$, on a

$$\begin{aligned} f(A+H) &= (A+H)^{-1} = A^{-1}(I_d + H A^{-1})^{-1} \\ &= A^{-1} \sum_{k=0}^{+\infty} (-H A^{-1})^k \\ &\stackrel{O_d}{=} A^{-1} - A^{-1} H A^{-1} + o(\|H\|). \end{aligned}$$

D'où $df(A)(H) = -A^{-1} H A^{-1}$. Quand $d = 1$, on retrouve $f'(a) = -a^{-2}$.

Exemple 14 $A = (x_A, y_A)$ et $f : \mathbf{R}^2 \rightarrow \mathbf{R}^3, (x, y) \mapsto (y^2, \cos(x+y), x e^y)$. f est différentiable en A et $df(A) : \mathbf{R}^2 \rightarrow \mathbf{R}^3$ est donné par, pour $H = (x_H, y_H)$,

$$\begin{aligned} df(A)(H) &= \begin{bmatrix} 0 & 2y_A \\ -\sin(x_A + y_A) & -\sin(x_A + y_A) \\ e^{y_A} & x_A e^{y_A} \end{bmatrix} \begin{pmatrix} x_H \\ y_H \end{pmatrix} \\ &= (2y_A y_H, -\sin(x_A + y_A)(x_H + y_H), e^{y_A} x_H + x_A e^{y_A} y_H). \end{aligned}$$

1.3 Dérivées partielles

Comme le laisse penser ce dernier exemple, il peut parfois — mais pas toujours! — être plus commode de calculer en faisant varier coordonnée par coordonnée, en dérivant direction par direction.

Définition 15 Soit $f : \Omega \rightarrow \mathbf{R}^p$ (où $\Omega \subset \mathbf{R}^n$) et $A \in \Omega$ tel que Ω soit un voisinage de A .

1. On dit que f est dérivable en $A \in \Omega$ dans la direction H si la fonction $I \rightarrow \mathbf{R}^p$, $t \mapsto f(A + tH)$ (où I est un intervalle ouvert centré en 0) est dérivable en 0. Cette dérivée directionnelle est alors notée $\partial_H f(A) = \frac{\partial f}{\partial H}(A)$.
2. Quand H est l'un des vecteurs de la base (canonique) e_i correspondant à une coordonnée X_i , on parle plutôt de dérivée partielle et l'on note $\partial_i f(A) = \partial_{X_i} f(A) = \frac{\partial f}{\partial X_i}(A)$.

Théorème 16 1. Si f est différentiable en A , alors f y possède des dérivées dans toutes les directions, $\partial_H f(A) = df(A)(H)$ et

$$df(A)(H) = \frac{\partial f}{\partial X_1}(A) H_1 + \cdots + \frac{\partial f}{\partial X_n}(A) H_n.$$

En particulier, alors $\frac{\partial f}{\partial X_i}(A) = df(A)(e_i)$.

2. Réciproquement, si f possède dans un voisinage dans A des dérivées partielles (par rapport à toutes les coordonnées) **continues**, alors f possède une différentielle en A .

Démonstration. Donnons une idée de la démonstration de la réciproque quand $n = 2$ (pour simplifier les notations) :

$$\begin{aligned} f(A + H) - f(A) &= f(A + H_1 e_1 + H_2 e_2) - f(A + H_1 e_1) + f(A + H_1 e_1) - f(A) \\ &\stackrel{(0,0)}{=} \frac{\partial f}{\partial X_2}(A + H_1 e_1) H_2 + o(H_2) + \frac{\partial f}{\partial X_1}(A) H_1 + o(H_1) \\ &\stackrel{(0,0)}{=} \frac{\partial f}{\partial X_1}(A) H_1 + \frac{\partial f}{\partial X_2}(A) H_2 + o(\|H\|). \end{aligned}$$

■

Observation 17 En remarquant que les différentielles des coordonnées sont données par $dX_i(A)(H) = H_i$, on a donc

$$df(A) = \sum_i \frac{\partial f}{\partial X_i}(A) dX_i(A)$$

ou encore

$$df(A)(H) = \sum_i \frac{\partial f}{\partial X_i}(A) dX_i(A)(H) = \sum_i \frac{\partial f}{\partial X_i}(A) H_i.$$

Si f est différentiable on a donc $df = \sum_i \frac{\partial f}{\partial X_i} dX_i$. Avec $n = 1$, on a en particulier $df = f' dx$.

Observation 18 Cela implique qu'en écrivant f en coordonnées $f = (f_1, \dots, f_p)$,

$$[\text{Jac}(f)(A)]_{i,j} = \frac{\partial f_i}{\partial X_j}(A).$$

Quand $p = 1$, on trouve aussi

$$\nabla f(A) = \left(\frac{\partial f}{\partial X_1}(A), \dots, \frac{\partial f}{\partial X_n}(A) \right).$$

Définition 19 Soit $\Omega \subset \mathbf{R}^n$ un ouvert. On note $\mathcal{C}^1(\Omega, \mathbf{R}^p)$, ou simplement $\mathcal{C}^1(\Omega)$ si $p = 1$, l'ensemble des fonctions de Ω dans \mathbf{R}^p différentiables à différentielle continue. De telles fonctions sont dites de classe \mathcal{C}^1 ou continûment différentiable.

1.4 Opérations

Proposition 20 *Si*

— $\Omega \subset \mathbf{R}^n$ est un voisinage de A , $f : \Omega \rightarrow \mathbf{R}^p$ est différentiable en A ,

— $\Omega' \subset \mathbf{R}^p$ est un voisinage de $f(A)$ et $g : \Omega' \rightarrow \mathbf{R}^q$ est différentiable en $f(A)$,

alors $g \circ f$ est différentiable en A et $d(g \circ f)(A) = dg(f(A)) \circ df(A)$, c'est-à-dire

$$d(g \circ f)(A)(H) = dg(f(A))(df(A)(H)).$$

Démonstration. Cela se démontre avec

$$\begin{aligned} (g \circ f)(A + H) &\stackrel{(0, \dots, 0)}{=} g(f(A) + df(A)(H) + o(\|H\|)) \\ &\stackrel{(0, \dots, 0)}{=} g(f(A) + dg(f(A))(df(A)(H) + o(\|H\|)) + o(\|df(A)(H) + o(H)\|)) \\ &\stackrel{(0, \dots, 0)}{=} g(f(A) + dg(f(A))(df(A)(H)) + o(\|H\|)). \end{aligned}$$

■

Observation 21 *En coordonnées cela se récrit*

$$\partial_i(g \circ f)(A) = \sum_{j=1}^n \partial_j g(f(A)) \partial_i f_j(A).$$

Matriciellement cela se récrit

$$\text{Jac}(g \circ f)(A) = \text{Jac}(g)(f(A)) \text{Jac}(f)(A).$$

Observation 22 *Cela permet de montrer beaucoup de relations de différentiation. On pourra en particulier considérer le cas où g est une application multilinéaire.*

Exemple 23 *Soit $V \in \mathbf{R}^n$ et $u_0 : \mathbf{R}^n \rightarrow \mathbf{R}$ de classe \mathcal{C}^1 . Définissons*

$$u : \mathbf{R} \times \mathbf{R}^n \rightarrow \mathbf{R}, \quad (t, X) \mapsto u_0(X - tV).$$

Alors u est de classe \mathcal{C}^1 et

$$\partial_t u + V \cdot \nabla_X u \equiv 0.$$

1.5 Ordre supérieur et développement de Taylor

Lorsque f est différentiable (en tout point de Ω), la différentielle df étant elle-même une fonction, on peut vouloir la différentier. Quand c'est possible, on dit que f est *deux fois différentiable* en A et l'on introduit la notation $d^2f(A)$ (définissant une application bi-linéaire) pour la différentielle seconde au point A

$$df(A + H') \stackrel{(0, \dots, 0)}{=} df(A) + d^2f(A)(\cdot, H') + o(\|H'\|)$$

c'est-à-dire, pour tout H ,

$$df(A + H')(H) \stackrel{(0, \dots, 0)}{=} df(A)(H) + d^2f(A)(H, H') + o(\|H\| \|H'\|).$$

Quand $n = 1$, on trouve $df^2(a)(h, h') = h h' f''(a)$.

En sachant comment varie la différentielle, on obtient une information plus précise sur f .

Proposition 24 Développement de Taylor à l'ordre 2. Si $f : \Omega \rightarrow \mathbf{R}^p$ est deux fois différentiable en A , alors

$$f(A + H) \stackrel{(0, \dots, 0)}{=} f(A) + df(A)(H) + \frac{1}{2}d^2f(A)(H, H) + o(\|H\|^2).$$

Remarque 25 Contrairement au cas des dérivées premières, il n'y a pas de réciproque.

Corollaire 26 Si $f : \Omega \rightarrow \mathbf{R}$ ($p = 1$) est deux fois différentiable en A , que A est un point critique (c'est-à-dire que $df(A)$ est nulle) et que $d^2f(A)$ est définie positive, alors f atteint un minimum local strict en A .

Observation 27 La formule est exacte quand f est quadratique. En effet si $B : \mathbf{R}^n \times \mathbf{R}^n \rightarrow \mathbf{R}^p$ est bilinéaire et $f : \mathbf{R}^n \rightarrow \mathbf{R}^p$, $X \rightarrow B(X, X)$, alors d'une part

$$f(A + H) = B(A + H, A + H) = B(A, A) + B(A, H) + B(H, A) + B(H, H)$$

et d'autre part $df(A)(H) = B(A, H) + B(H, A)$, $d^2f(A)(H, H') = B(H', H) + B(H, H')$.

On peut évidemment passer par les dérivées partielles.

Si f possède (dans un voisinage de A) une dérivée dans la direction H' et que cette dérivée partielle est dérivable dans la direction H en A , on note

$$\frac{\partial^2 f}{\partial H \partial H'}(A) = \frac{\partial}{\partial H} \left(\frac{\partial f}{\partial H'} \right) (A).$$

Si $H = H'$, on note aussi

$$\frac{\partial^2 f}{\partial H \partial H}(A) = \frac{\partial^2 f}{\partial H^2}(A).$$

Comme pour les dérivées d'ordre 1, si $H = e_i$ et $H' = e_j$, on parle de dérivée partielle et l'on note

$$\frac{\partial^2 f}{\partial X_i \partial X_j}(A), \quad \text{et} \quad \frac{\partial^2 f}{\partial X_i^2}(A) \quad \text{si } i = j.$$

Corollaire 28 1. Si f est deux fois différentiable en A , alors f y possède des dérivées partielles d'ordre 2 dans toutes les directions $\frac{\partial^2 f}{\partial H \partial H'}(A) = d^2f(A)(H, H')$ et

$$d^2f(A)(H, H') = \sum_{i,j} \frac{\partial^2 f}{\partial X_j \partial X_i}(A) H_i H'_j.$$

En particulier,

$$\frac{\partial^2 f}{\partial X_j \partial X_i}(A) = d^2f(A)(e_i, e_j).$$

2. Réciproquement, si f possède dans un voisinage de A des dérivées partielles d'ordre 2 (par rapport à toutes les coordonnées) continues, alors f possède une différentielle d'ordre 2 en A .

Théorème 29 Théorème de Schwarz. Si f possède dans un voisinage dans A des dérivées partielles d'ordre 2 (par rapport à toutes les coordonnées) continues, alors $d^2f(A)$ est symétrique, c'est-à-dire

$$d^2f(A)(H, H') = d^2f(A)(H', H).$$

En particulier,

$$\frac{\partial^2 f}{\partial X_i \partial X_j}(A) = \frac{\partial^2 f}{\partial X_j \partial X_i}(A).$$

La relation de Schwarz exprime que pour une fonction raisonnable un petit accroissement H suivi d'un petit accroissement H' a (approximativement) le même effet qu'un petit accroissement de H' suivi d'un petit accroissement H .

Remarque 30 Quand $p = 1$, on introduit la matrice hessienne $\text{Hess}(f)(A)$ la matrice telle que

$$d^2f(A)(H, H') = \langle H, \text{Hess}(f)(A)H' \rangle.$$

En particulier, $[\text{Hess}(f)(A)]_{i,j} = \frac{\partial^2 f}{\partial X_j \partial X_i}(A)$.

Définition 31 Soit $\Omega \subset \mathbf{R}^n$ un ouvert. On note $\mathcal{C}^2(\Omega, \mathbf{R}^p)$, ou simplement $\mathcal{C}^2(\Omega)$ si $p = 1$, l'ensemble des fonctions de Ω dans \mathbf{R}^p deux fois différentiables à différentielle seconde continue. De telles fonctions sont dites de classe \mathcal{C}^2 ou deux fois continûment différentiable.

Évidemment, on peut obtenir des équivalents avec plus de différentiations que l'on admettra par la suite.

2 Du local au global

2.1 Développements de Taylor

On a énoncé ci-dessus les formules de Taylor-Young. Pour obtenir une version plus précise, il faut se ramener en dimension 1 en choisissant un chemin $\gamma : [0, 1] \rightarrow \Omega$ tel que $\gamma(0) = A$ et $\gamma(1) = B$ puis en développant $f(B) - f(A) = f \circ \gamma(1) - f \circ \gamma(0)$.

Évidemment le chemin le plus simple est le segment $[A, B]$ paramétré par

$$\gamma : [0, 1] \rightarrow \Omega, \quad t \mapsto A + t(B - A) = (1 - t)A + tB.$$

Avec ce choix, si f est k fois différentiable, on a, pour tout $t \in [0, 1]$,

$$(f \circ \gamma)^{(k)}(t) = d^k f(\gamma(t))(B - A, \dots, B - A).$$

D'où ce qui suit.

Théorème 32 1. **Développement de Taylor avec reste intégral.**

Soit $\Omega \subset \mathbf{R}^n$ un ouvert et $k \in \mathbf{N}$.

Si $f : \Omega \rightarrow \mathbf{R}^p$ est $(k + 1)$ fois continûment différentiable et $[A, B] \subset \Omega$, alors

$$\begin{aligned} f(B) &= \sum_{j=0}^k \frac{1}{j!} d^j f(A)(B - A, \dots, B - A) \\ &+ \frac{1}{k!} \int_0^1 (1 - t)^k d^{k+1} f(A + t(B - A))(B - A, \dots, B - A) dt. \end{aligned}$$

2. Formule de Taylor-Lagrange.

Soit $\Omega \subset \mathbf{R}^n$ un ouvert et $k \in \mathbf{N}$.

Si $f : \Omega \rightarrow \mathbf{R}$ ($p = 1$) est $(k + 1)$ fois différentiable et $[A, B] \subset \Omega$, alors il existe $\xi \in [A, B]$ tel que

$$\begin{aligned} f(B) &= \sum_{j=0}^k \frac{1}{j!} d^j f(A)(B - A, \dots, B - A) \\ &+ \frac{1}{(k + 1)!} d^{k+1} f(\xi)(B - A, \dots, B - A). \end{aligned}$$

Remarque 33 Dans les deux cas, on déduit

$$\begin{aligned} \left\| f(B) - \sum_{j=0}^k \frac{1}{j!} d^j f(A)(B-A, \dots, B-A) \right\| \\ \leq \frac{1}{(k+1)!} \left(\sup_{[A,B]} \|d^{k+1} f\| \right) \|B-A\|^{k+1}. \end{aligned}$$

Mais les hypothèses pour la formule de Taylor-Lagrange n'impliquent pas que la borne supérieure ci-dessus est finie.

Remarque 34 Un ensemble Ω tel que $[A, B] \subset \Omega$ pour toute paire $(A, B) \in \Omega^2$ est dit convexe. Les convexes de \mathbf{R} ($n = 1$) sont précisément les intervalles.

2.2 Convexité

Quand $p = 1$, on peut chercher des minoration ou des majoration plutôt que des bornes.

Définition 35 Soit $\Omega \subset \mathbf{R}^n$ un convexe et $f : \Omega \mapsto \mathbf{R}$.

1. On dit que f est convexe si, pour toute paire $(A, B) \in \Omega^2$, pour tout $t \in [0, 1]$,

$$f((1-t)A + tB) \leq (1-t)f(A) + tf(B).$$

On dit que f est concave si, pour toute paire $(A, B) \in \Omega^2$, pour tout $t \in [0, 1]$,

$$f((1-t)A + tB) \geq (1-t)f(A) + tf(B).$$

2. On dit que f est strictement convexe si, pour toute paire $(A, B) \in \Omega^2$ telle que $A \neq B$, pour tout $t \in]0, 1[$,

$$f((1-t)A + tB) < (1-t)f(A) + tf(B).$$

On dit que f est strictement concave si, pour toute paire $(A, B) \in \Omega^2$ telle que $A \neq B$, pour tout $t \in]0, 1[$,

$$f((1-t)A + tB) > (1-t)f(A) + tf(B).$$

Remarque 36 f est concave (respectivement strictement concave) si et seulement si $-f$ est convexe (respectivement strictement convexe). On ne discutera donc que la convexité.

Proposition 37 Soit $\Omega \subset \mathbf{R}^n$ un ouvert convexe et $f : \Omega \mapsto \mathbf{R}$.

1. Si f est différentiable, alors f est convexe si et seulement si, pour toute paire $(A, B) \in \Omega^2$,

$$f(B) \geq f(A) + df(A)(B-A).$$

2. Si f est différentiable, alors f est strictement convexe si et seulement si, pour toute paire $(A, B) \in \Omega^2$ telle que $A \neq B$,

$$f(B) > f(A) + df(A)(B-A).$$

3. Si f est deux fois différentiable, alors f est convexe si et seulement si, en tout point $A \in \Omega$, $d^2 f(A)$ est symétrique positive.

4. Si f est deux fois différentiable, alors si, en tout point $A \in \Omega$, $d^2f(A)$ est symétrique définie positive, f est strictement convexe.

Démonstration. On peut le déduire directement du cas $n = 1$. Pour une démonstration *ab initio* : les points 1. et 2. découlent de simples manipulations algébriques, les points 3. et 4. sont des conséquences directes de la formule de Taylor-Lagrange à l'ordre 2. ■

3 Inversibilité et points fixes

Il est également important de comprendre comment convertir l'inversibilité du linéarisé en inversibilité (locale) non linéaire ou comment décrire la dynamique au voisinage d'un point fixe à partir d'un développement de Taylor. Sur ces sujets on renvoie le lecteur au cours de complément consacré aux équations non linéaires.

4 Intégration et divergence

4.1 Sur un domaine

Il est facile d'intégrer sur des rectangles (ou des parallélépipèdes droits) parallèles aux axes. En effet, si $f : [a, b] \times [c, d] \rightarrow \mathbf{R}^p$ est continue alors

$$\int_{[a,b] \times [c,d]} f = \int_{[a,b] \times [c,d]} f(x, y) \, dx \otimes dy = \int_a^b \left(\int_c^d f(x, y) \, dy \right) dx = \int_c^d \left(\int_a^b f(x, y) \, dx \right) dy.$$

Le fait que l'on puisse calculer les intégrales en les découpant en tranches repose sur le **théorème de Fubini**.

Fréquemment l'objectif de changer de variables est de se ramener à ces cas. Avant d'énoncer la formule générale, discutons le cas d'un changement de variable linéaire. Si $\ell : \mathbf{R}^n \rightarrow \mathbf{R}^n$ est linéaire, alors en appliquant ℓ à un parallélépipède, on multiplie son volume par $|\det \ell|$.

Proposition 38 Changement de variables. Si $\phi : \Omega \rightarrow \Omega'$ est bijective avec ϕ et ϕ^{-1} de classe \mathcal{C}^1 et $f : \Omega' \rightarrow \mathbf{R}^p$ est continue, alors

$$\int_{\Omega'} f(X) dX = \int_{\Omega} f(\phi(Y)) |\det d\phi(Y)| dY.$$

Le déterminant de la différentielle (qui intervient dans la formule) est appelé **déterminant jacobien** ou juste jacobien. Il est utile de connaître par cœur et de savoir retrouver rapidement les jacobiens des changements de variables les plus classiques.

Coordonnées polaires :

$$\phi : \mathbf{R}_+^* \times [0, 2\pi[\rightarrow \mathbf{R}^2 \setminus \{(0, 0)\}, \quad (r, \theta) \mapsto (r \cos(\theta), r \sin(\theta))$$

de jacobien r en (r, θ) .

Coordonnées cylindriques :

$$\phi : \mathbf{R}_+^* \times [0, 2\pi[\times \mathbf{R} \rightarrow \mathbf{R}^3 \setminus \{(0, 0)\} \times \mathbf{R}, \quad (r, \theta, z) \mapsto (r \cos(\theta), r \sin(\theta), z)$$

de jacobien r en (r, θ, z) .

Coordonnées sphériques :

$$\phi : \mathbf{R}_+^* \times [0, 2\pi[\times]-\pi/2, \pi/2[\rightarrow \mathbf{R}^3 \setminus \{(0, 0)\} \times \mathbf{R}, \quad (r, \theta, \varphi) \mapsto (r \cos(\theta) \cos(\varphi), r \sin(\theta) \cos(\varphi), r \sin(\varphi))$$

de jacobien $r^2 \cos(\varphi)$ en (r, θ, φ) .

Pour l'exemple, calculons le volume de la boule centrée en $(0, 0, 0)$ et de rayon R . Ce volume est

$$\int_{]0, R[\times [0, 2\pi[\times]-\pi/2, \pi/2[} r^2 \cos(\varphi) \, dr \otimes d\theta \otimes d\varphi = \frac{R^3}{3} \times 2\pi \times 2 = \frac{4\pi}{3} R^3.$$

4.2 Intégrale sur une sous-variété

On va se restreindre à intégrer des fonctions¹.

Définition 39 Soit $\Gamma \subset \mathbf{R}^n$ une sous-variété de \mathbf{R}^n de dimension $d \in \mathbf{N}^*$, paramétrée par $\gamma : \Omega \rightarrow \mathbf{R}^n$ de classe \mathcal{C}^1 , avec $\Omega \subset \mathbf{R}^d$ ouvert.

Pour $f : \Gamma \rightarrow \mathbf{R}^p$ continue, on définit l'intégrale de f sur Γ par

$$\int_{\Gamma} f d\sigma := \int_{\Omega} f(\gamma(p)) |\det((d\gamma(p))^* d\gamma(p))|^{\frac{1}{2}} dp.$$

La formule de changement de variable en dimension d montre que cela ne dépend pas du paramétrage choisi. Quand $d = 1$, on retrouve la notion d'intégrale curviligne

$$\int_{\Gamma} f d\sigma := \int_{\Omega} f(\gamma(p)) \|\gamma'(p)\| dp.$$

À l'aide d'une partition de l'unité, on peut généraliser la définition au cas où la sous-variété n'est pas globalement paramétrable. On peut aussi relaxer la régularité \mathcal{C}^1 du paramétrage en une régularité Lipschitz, ce qui permet de considérer des polyèdres.

Quand f est constante égale à 1, on trouve le volume de Γ . On parle traditionnellement plutôt de longueur de la courbe quand $d = 1$ et d'aire de la surface quand $d = 2$.

Remarque 40 Toute la construction des objets associés à une sous-variété peut se faire avec des paramétrages, mais ces objets ne dépendent pas du paramétrage choisi. Donnons-en un autre exemple avec le fait que l'espace tangent à Γ en $x = \gamma(p)$ peut être identifié avec $d\gamma(p)(\mathbf{R}^d)$.

4.3 Opérateur divergence

Définition 41 À tout champ de vecteur $V : \Omega \rightarrow \mathbf{R}^d$, \mathcal{C}^1 sur un ouvert Ω de \mathbf{R}^d , on peut associer sa **divergence** $\operatorname{div}(V) : \Omega \rightarrow \mathbf{R}$ définie par

$$\operatorname{div}(V) = \operatorname{Tr}(dV) = \sum_{j=1}^d \partial_j V_j.$$

La définition de la divergence à l'aide de la trace et de la différentielle montre qu'il s'agit bien d'un objet intrinsèque, qui ne dépend pas des coordonnées choisies pour le calculer.

L'utilité sans doute principale de la divergence est de permettre la généralisation multidimensionnelle suivante du théorème fondamental de l'analyse. Ce théorème est lui-même un cas particulier d'un théorème plus général appelé théorème de Stokes.

1. On peut aussi intégrer des formes différentielles, c'est-à-dire des applications qui à tout point de la sous-variété associe des applications multi-linéaires sur l'espace tangent et/ou co-tangent au point.

Théorème 42 Théorème de la divergence ou de Green-Ostrogradski. Soit Ω un ouvert borné de \mathbf{R}^d dont le bord est une sous-variété \mathcal{C}^1 . Soit $V : \bar{\Omega} \rightarrow \mathbf{R}^d$ continu, \mathcal{C}^1 sur Ω . Alors

$$\int_{\Omega} \operatorname{div}(V) = \int_{\partial\Omega} V \cdot n \, d\sigma$$

où, pour tout $x \in \partial\Omega$, $n(x)$ est la normale unitaire de $\partial\Omega$ sortant de Ω .

Le théorème s'étend au cas où le bord est seulement Lipschitz. L'une de ses démonstrations consiste d'ailleurs à le déduire du cas où Ω est un parallélépipède droit.

Explicitons le sur un rectangle en dimension deux, un cas particulièrement facile à démontrer et utile pour l'agrégation. Si u et v sont \mathcal{C}^1 , on a :

$$\begin{aligned} \int_{[t_0, t_1] \times [x_0, x_1]} (\partial_t u + \partial_x v)(t, x) \, dt \, dx &= \int_{[x_0, x_1]} u(t_1, x) \, dx - \int_{[x_0, x_1]} u(t_0, x) \, dx \\ &+ \int_{[t_0, t_1]} v(t, x_1) \, dt - \int_{[t_0, t_1]} v(t, x_0) \, dt. \end{aligned}$$

On utilise souvent la conséquence suivante : $\partial_t u + \partial_x v$ est nul partout si et seulement si pour tout $t \in \mathbf{R}$ et tous réels x_0, x_1 , on a :

$$\frac{d}{dt} \left(\int_{[x_0, x_1]} u(\cdot, x) \, dx \right) (t) = -(v(t, x_1) - v(t, x_0)).$$

Évidemment on peut aussi obtenir l'équivalence en dérivant en x plutôt qu'en intégrant en t .

4.4 Flot et divergence

Connaître la divergence d'un champ de vecteur permet de savoir comment il fait sortir ou entrer, concentrer ou diluer quand on le suit. Cette interprétation est déjà contenue dans le théorème de la divergence. Ainsi quand la divergence est nulle, il y a compensation entre tout ce qui entre par le bord et ce qui en sort.

Une autre manière de se convaincre de cette interprétation est de considérer les trajectoires suivant le champ de vecteur.

Donnons-nous un champ de vecteur $V : \mathbf{R}^n \rightarrow \mathbf{R}^n$ \mathcal{C}^1 et Lipschitzien². Notons Φ le flot différentiel associé, c'est-à-dire que $\Phi(t, t_0, x_0)$ est la valeur au temps t de la solution de $x' = V(x)$ partant de x_0 en t_0 .

Alors pour tout (t, t_0, x_0) , on a :

$$\det(d_x \Phi)(t, t_0, x_0) = \exp \left(\int_{t_0}^t \operatorname{div}(V)(\Phi(s, t_0, x_0)) \, ds \right).$$

En particulier, d'après la formule de changement de variable, si V est à divergence nulle, son flot préserve la mesure.

La formule ci-dessus est une version non linéaire de la formule

$$\det(e^{tA}) = e^{t \operatorname{Tr}(A)}$$

valable pour toute matrice carrée A et tout réel t . On peut d'ailleurs déduire la version non linéaire de la version linéaire par approximation.

2. L'hypothèse est seulement faite pour évacuer les questions liées au temps d'existence des solutions.

Références

- [BG14] S. Benzoni-Gavage. *Calcul différentiel et équations différentielles*. Dunod, 2014.
- [Rou15] F. Rouvière. *Petit guide de calcul différentiel à l'usage de la licence et de l'agrégation*. Cassini, 2015.