



UNIVERSITÉ DE NANTES

Licence de physique
2018-2019

Calcul différentiel et intégral

Laurent Guillopé

Laboratoire de mathématiques Jean Leray
Département de mathématiques
UFR Sciences et techniques
Université de Nantes



UNIVERSITÉ DE NANTES

Licence de physique

2018-2019

Calcul différentiel et intégral

Laurent Guillopé

Laboratoire de mathématiques Jean Leray

Département de mathématiques

UFR Sciences et techniques

Université de Nantes

I : Fonction de plusieurs variables : dérivées partielles

Version : 25 mars 2020 11:23



UNIVERSITÉ DE NANTES

Licence de physique

2018-2019

Calcul différentiel et intégral

Laurent Guillopé

Laboratoire de mathématiques Jean Leray

Département de mathématiques

UFR Sciences et techniques

Université de Nantes

II : Développements limités

Version : 25 mars 2020 11:23



UNIVERSITÉ DE NANTES

Licence de physique
2018-2019

Calcul différentiel et intégral

Laurent Guillopé

Laboratoire de mathématiques Jean Leray
Département de mathématiques
UFR Sciences et techniques
Université de Nantes

III : Intégrales généralisées



UNIVERSITÉ DE NANTES

Licence de physique

2018-2019

Calcul différentiel et intégral

Laurent Guillopé

Laboratoire de mathématiques Jean Leray

Département de mathématiques

UFR Sciences et techniques

Université de Nantes

IV : Optimisation de fonctions en plusieurs variables

Version : 25 mars 2020 11:23

Table des matières

Table des matières	0
1 Fonction de plusieurs variables : dérivées partielles	1
1.1 Limites et continuité	3
1.2 Dérivées partielles, fonctions de classe \mathcal{C}^1	5
1.3 Différentielle en un point	7
1.4 Espace tangent	13
1.5 Dérivées d'ordre supérieur	14
1.6 Opérateurs différentiels : gradient, rotationnel, divergence	19
1.7 Annexe 1 : applications linéaires et matrices	20
1.8 Annexe 2 : notations, composées et différentielles	21
2 Développements limités	25
2.1 Approximations polynomiales	25
2.2 Formules de Taylor	28
2.3 Premières propriétés	31
2.4 Opérations sur les développements limités	35
2.5 Calculs de limites	42
2.6 Étude locale et asymptotique de courbes	43
3 Intégrales généralisées	45
3.1 Intégrale généralisée convergente	45
3.2 Fonctions positives et convergence absolue	49
3.3 Intégrales de Gauß, Dirichlet et Fresnel.	51
3.4 Fonctions Γ et B	57
4 Optimisation de fonctions en plusieurs variables	59
4.1 La condition d'Euler-Fermat	61
4.2 Optimisation d'une fonction d'une variable	63
4.3 Optima : condition du second ordre en plusieurs variables	64

Chapitre 1

Fonction de plusieurs variables : dérivées partielles

Ce chapitre a pour objectif de généraliser les propriétés des fonctions numériques d'une variable réelle à des fonctions de plusieurs variables réelles (x, y, z, \dots ou x_1, x_2, \dots, x_m) à valeurs dans un espace \mathbb{R}^n (*i. e.* à valeurs vectorielles)

$$f : \mathbf{x} = (x_1, \dots, x_m) \in \mathbb{R}^m \longmapsto f(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_n(\mathbf{x})) \in \mathbb{R}^n.$$

La généralisation des concepts et résultats des fonctions d'une variable est souvent aisée, mais des nouveaux phénomènes apparaissent (souvent liés à des résultats d'algèbre linéaire) et les énoncés ne seront pas toujours démontrés. On considérera plutôt les cas où m et n sont parmi 1, 2 ou 3. Les coordonnées de \mathbf{x} seront notées x_i ou $\mathbf{x}(i)$.

Pour une fonction $f : U \rightarrow \mathbb{R}$ numérique de 2 variables, on a la possibilité de représentations graphiques :

1. une courbe (ou ligne, ensemble) de niveau h pour la fonction f est la partie du plan définie par $\{(x, y) \in U, f(x, y) = h\}$.
2. le graphe G_f de la fonction f est la partie de $U \times \mathbb{R}$ définie par

$$G_f = \{(x, y, f(x, y)) \in \mathbb{R}^3, (x, y) \in U\}.$$

La ligne de niveau h est l'intersection du graphe G_f avec le plan horizontal de cote $z = h$.

En 3 variables, on peut examiner les ensembles de niveau h (en général des surfaces) définis suivant $\{(x, y, z) \in U, f(x, y, z) = h\}$.

▷ EXEMPLES 1.1:

- 1.1.1 $f : \mathbf{x} = (x_1, \dots, x_m) \in \mathbb{R}^m \mapsto a_1x_1 + \dots + a_mx_m + b \in \mathbb{R}$ fonction affine (linéaire si $b = 0$), le graphe (comme toute ligne de niveau) est un (hyper)plan de dimension m .

1.1.2 Pour la fonction¹ $f : (x, y) \mapsto x^2 + y^2$, les lignes de niveau $x^2 + y^2 = h$ ($h > 0$) sont des cercles centrés en l'origine, alors que le graphe $\{(x, y, z) | z = x^2 + y^2\}$ est un paraboloïde.

1.1.3 $\mathbf{x} = (x, y) \in (\mathbb{R}^2 \cap \{x \neq y\}) \mapsto (x + y)/(x - y) \in \mathbb{R}$; la fonction est définie en dehors de la droite $\{x = y\}$. Les lignes de niveau sont des droites.

1.1.4 $r : \mathbf{x} = (x, y, z) \in \mathbb{R}^3 \mapsto \|\mathbf{x}\|_2 = \sqrt{x^2 + y^2 + z^2}$ norme d'un vecteur de \mathbb{R}^3 . Dans \mathbb{R}^2 , on a pareillement cette fonction, qui sera notée à l'identique;

1.1.5 $r : \mathbf{x} = (x, y) \in \mathbb{R}^2 \mapsto \sqrt{x^2 + y^2} = \|(x, y)\|_2 \in \mathbb{R}$;

1.1.6

$$(x, y) \in \mathbb{R}^2 \setminus \{(x, 0), x \in \mathbb{R}^{-*}\} \mapsto \theta(x, y) = \begin{cases} \operatorname{arctg}(y/x) & \text{si } x > 0, \\ \pi/2 - \operatorname{arctg}(x/y) & \text{si } y > 0, \\ -\pi/2 - \operatorname{arctg}(x/y) & \text{si } y < 0; \end{cases}$$

La fonction θ donne l'angle (mesuré par un réel dans $] -\pi, \pi[$) du rayon de l'origine au point $m = (x, y)$ et le demi-axe horizontal positif. Cette fonction est définie sur $\mathbb{R}^2 \setminus (\mathbb{R}^-, 0)$: si $m = (x, 0)$ avec $x < 0$, les limites $\lim_{\alpha \rightarrow 0^+} \theta(m \pm (0, \alpha))$ sont distinctes, valant $\pm\pi$ resp. La fonction $(x, y) \mapsto (r(x, y), \theta(x, y))$ donne les coordonnées polaires du point m dans le plan de la géométrie euclidienne. En dimension 3, on a de manière analogue les coordonnées sphériques, bien définies en dehors de quelques coupures.

1.1.7 $(x, y) \in \mathbb{R}^2 \mapsto |x| + |y|$; on verra que cette fonction a de larges parties de non-différentiabilité, comme la suivante;

1.1.8 $(x, y, z) \in \mathbb{R}^3 \mapsto \|(x, y, z)\|_\infty = \max(|x|, |y|, |z|)$;

1.1.9 $(x, y) \in \mathbb{R}^2 \mapsto xy$: si $x, y > 0$, c'est la surface d'un rectangle avec dimensions x, y ;

1.1.10 $(x, y, z) \in \mathbb{R}^3 \mapsto 2(xy + yz + zx)$: si $x, y, z > 0$, c'est la surface des faces d'un parallélépipède x, y, z ; ◁

On considérera en général des fonctions définies sur \mathbb{R}^m tout entier. Parfois, il sera utile de considérer des fonctions avec domaine de définition plus restreint. On supposera le domaine de définition U ouvert au sens suivant :

1. Avec SageMath, les commandes

```
f=x**2+y**2
```

```
contour_plot(f, (x, -2, 2), (y, -2, 2), fill=False).show()
```

```
plot3d(f, (x, -2, 2), (y, -2, 2), viewer='threejs')
```

affichent des lignes de niveau, puis lancent une représentation du graphe de la fonction avec possibilité de déplacer ce graphe (testé sur mac).

DÉFINITION 1.1: La partie U de \mathbb{R}^2 est dite ouverte si pour tout $\mathbf{x} \in U$, il existe un $r > 0$ tel que la boule $B(\mathbf{x}, r) = \{\mathbf{y} \in \mathbb{R}^2, \|\mathbf{y} - \mathbf{x}\| < r\}$ est incluse dans \mathbb{R}^2 .

La définition en dimension 3 ou supérieure est quasi-identique.

Définissons les applications partielles relativement à une variable.

DÉFINITION 1.2: Soit $f : \mathbf{x} \in \mathbb{R}^m \mapsto f(\mathbf{x}) = (f_j(\mathbf{x}))_{j=1}^n \in \mathbb{R}^n$. L'application partielle d'indice $i \in \llbracket 1, m \rrbracket$ en \mathbf{x} est l'application

$$t \in \mathbb{R} \mapsto f(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_m) \in \mathbb{R}^n,$$

avec j -ème application coordonnée ($j \in \llbracket 1, n \rrbracket$) partielle d'indice $i \in \llbracket 1, m \rrbracket$

$$t \in \mathbb{R} \mapsto f_j(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_m) \in \mathbb{R}.$$

Si f est définie sur l'ouvert U , on considèrera une boule $B(\mathbf{x}, r)$ incluse dans U et l'application partielle d'indice i définie sur l'intervalle $]x_i - r, x_i + r[$ par

$$t \in]x_i - r, x_i + r[\mapsto f(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_m) \in \mathbb{R}^n$$

avec fonctions coordonnées partielles correspondantes.

△ REMARQUE 1.1: Comme application partielle, on considère parfois la fonction

$$s \in]-r, r[\mapsto f(x_1, \dots, x_{i-1}, x_i + s, x_{i+1}, \dots, x_m) \in \mathbb{R}^n$$

qui est liée à la précédente par le changement de variable $s = t - x_i$, modification sans bouleversement. ▽

1.1 Limites et continuité

DÉFINITION 1.3: La suite de points² $(\mathbf{x}_k)_{k \geq 0}$ dans \mathbb{R}^m est dite converger dans \mathbb{R}^m vers un point $\mathbf{x}_\infty \in \mathbb{R}^m$ si pour tout $i = 1, \dots, m$, la suite $(\mathbf{x}_k(i))_{k \geq 0}$ des i -èmes coordonnées converge dans \mathbb{R} vers la i -ème coordonnée $\mathbf{x}_\infty(i)$ de \mathbf{x}_∞ .

▷ EXEMPLE 1.2: La suite $(\mathbf{x}_k = (2^{-k}, 3^{1/k}, 2 - \sin(k)/k^2))_{k \geq 0}$ converge dans \mathbb{R}^3 vers $(0, 1, 2)$: on le voit aisément coordonnée par coordonnée. ◀

2. On ne confondra pas le statut des indices parmi $\mathbf{x}, x_i, \mathbf{x}_k(i)$. La lettre grasse \mathbf{x} désigne un vecteur de \mathbb{R}^m de coordonnées x_1, \dots, x_m , $\mathbf{x}_k(i)$ est la i -ème coordonnée du vecteur \mathbf{x}_k , de telle sorte que $\mathbf{x}(i) = x_i$ et $\mathbf{x}_\infty(i) = (x_\infty)_i$. La suite $(\mathbf{x}_k)_{k \geq 0}$ est une suite de vecteurs dans \mathbb{R}^m , alors que la suite $(\mathbf{x}_k(i))_{k \geq 0}$ est une suite numérique. Par économie ou pénurie de symboles, d'autres formulations du calcul différentiel prêtent à confusion, on essaiera d'y pallier..

△ REMARQUE 1.2: Vu que

$$\begin{aligned}\|\mathbf{x}_k - \mathbf{x}_\infty\|_\infty &= \max_{j=1}^m [|\mathbf{x}_k(j) - \mathbf{x}_\infty(j)|] \\ &\leq \|\mathbf{x}_k - \mathbf{x}_\infty\|_2 = \sqrt{(\mathbf{x}_k(1) - \mathbf{x}_\infty(1))^2 + \dots + (\mathbf{x}_k(m) - \mathbf{x}_\infty(m))^2} \\ &\leq m \max_{j=1}^m [|\mathbf{x}_k(j) - \mathbf{x}_\infty(j)|] = m \|\mathbf{x}_k - \mathbf{x}_\infty\|_\infty,\end{aligned}$$

la suite $(\mathbf{x}_k)_{k \geq 0}$ (de points dans \mathbb{R}^m) converge vers \mathbf{x}_∞ si et seulement si la suite (numérique) $(\|\mathbf{x}_k - \mathbf{x}_\infty\|_2)_{k \geq 0}$ converge vers 0, si et seulement si la suite (numérique) $(\|\mathbf{x}_k - \mathbf{x}_\infty\|_\infty)_{k \geq 0}$ converge vers 0. On notera dans la suite $\|\cdot\|_\infty$ ou $\|\cdot\|_2$. ▽

DÉFINITION 1.4: L'application $f : U(\subset \mathbb{R}^m) \rightarrow \mathbb{R}^n$ est continue en $\mathbf{x} \in U$ si pour toute suite $(\mathbf{x}_k)_{k \geq 0}$ convergente vers \mathbf{x} la suite $(f(\mathbf{x}_k))_{k \geq 0}$ converge vers $f(\mathbf{x})$.

L'application $f : U(\subset \mathbb{R}^m) \rightarrow \mathbb{R}^n$ est continue sur U si elle est continue en chacun des points de son domaine de définition U .

▷ EXEMPLES 1.3:

1.3.1 L'application i ème-coordonnée $\pi_i : \mathbf{x} \in \mathbb{R}^m \mapsto \mathbf{x}(i) \in \mathbb{R}$ est continue.

1.3.2 Si $f : \mathbb{R} \rightarrow \mathbb{R}$ est continue discontinue (comme la fonction « Partie entière ») et $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ continue, il n'est pas rare que la composée $f \circ g$ ne soit pas continue.

1.3.3 Soit $f_i : \mathbb{R} \rightarrow \mathbb{R}, i = 1, \dots, n$ des fonctions continues. Alors la fonction $x \in \mathbb{R}^m \mapsto (f_i(x_i)) \in \mathbb{R}^n$ est continue ◀

LEMME 1.1: L'application $f = (f_1, \dots, f_n) : U \rightarrow \mathbb{R}^n$ est continue en \mathbf{x} si et seulement si les fonctions coordonnées $\mathbf{x} \in U \mapsto (\pi_i \circ f)(\mathbf{x}) = f_i(\mathbf{x})$ pour $i = 1, \dots, n$ sont continues.

Si l'application $f : U \rightarrow \mathbb{R}^n$ est continue, ses applications partielles sont continues. La réciproque n'est pas vraie en toute généralité, même si les fonctions considérées le seront, sauf exception rarissime.

▷ EXEMPLE 1.4: Comme exemple de telle fonction, prenons l'application f nulle à l'origine et vérifiant sinon

$$f(x, y) = \frac{2xy}{x^2 + y^2}, \quad (x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}.$$

Les applications partielles à l'origine $(0, 0)$, soit $t \in \mathbb{R} \mapsto f(t, 0) \in \mathbb{R}$ et $t \in \mathbb{R} \mapsto f(0, t) \in \mathbb{R}$ sont identiquement nulles, et par suite sont continues. Par ailleurs l'application f est constante égale à $\sin(2\theta)$ sur la droite $D^*(\theta)$ passant par l'origine privée de cette origine et de pente $\tan\theta$: cette droite est paramétrée par $r > 0 \mapsto (r \cos\theta, r \sin\theta)$ avec $\varphi = \theta$ ou $\theta + \pi$. La fonction f y prend la valeur constante $\sin(2\theta)$ vu

$$f(r \cos\theta, r \sin\theta) = \frac{2r \cos\theta r \sin\theta}{(r \cos\theta)^2 + (r \sin\theta)^2} = 2 \cos\theta \sin\theta = \sin(2\theta), \quad r > 0,$$

différente de $f(0,0)$: si $(r_n)_{n \geq 0}$ est une suite de réels non nuls positifs tendant vers 0, alors la suite $((r_n \cos \theta, r_n \sin \theta))_{n \geq 0}$ tend vers $(0,0)$, avec $f(r_n \cos \theta, r_n \sin \theta) = \sin(2\theta)$ qui ne tend pas vers $f(0,0)$ (si la droite $D^*(\theta)$ n'est pas l'axe vertical ou horizontal) : la fonction f n'est pas continue en $\mathbf{x} = (0,0)$. La fonction f est cependant continue sur le plan \mathbb{R}^2 privé de l'origine $(0,0)$. \triangleleft

Le théorème suivant suffit à montrer souvent la continuité d'applications d'un ouvert de \mathbb{R}^m dans \mathbb{R} .

THÉORÈME 1.1: *La somme de deux applications continues de \mathbb{R}^m dans \mathbb{R}^n est continue.*

Le produit de deux applications continues de \mathbb{R}^m dans \mathbb{R} est continue.

La composée $g \circ f : \mathbb{R}^m \rightarrow \mathbb{R}^p$ d'une application continue $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ par une application continue $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ est continue.

\triangle **REMARQUE 1.3:** Le cas de l'inverse $f^{-1} : \mathbf{x} \in (U \setminus \{\mathbf{y} : f(\mathbf{y}) \neq 0\}) \mapsto 1/f(\mathbf{x})$ d'une fonction numérique peut être vu comme la composée de la fonction f suivie de l'inverse $t \in \mathbb{R}^* \mapsto t^{-1} \in \mathbb{R}^*$. ∇

\triangleright **EXEMPLES 1.5:**

1.5.1 Soit $f_i : \mathbb{R} \rightarrow \mathbb{R}, i = 1, \dots, m$ des fonctions continues. Alors les fonctions

$$\mathbf{x} \in \mathbb{R}^m \mapsto \sum_{i=1}^m f_i(x_i) \in \mathbb{R}, \quad \mathbf{x} \in \mathbb{R}^m \mapsto \prod_{i=1}^m f_i(x_i) \in \mathbb{R}$$

sont continues.

1.5.2 L'application

$$(x, y, z) \in \mathbb{R}^3 \mapsto \left(\frac{e^{xy} \sin(z + \ln(1 + y^2))}{1 + x^2 + y^2}, \sqrt{\frac{2 \cos^2(x + y)}{1 + z^4}} \right) \in \mathbb{R}^2$$

est continue sur \mathbb{R}^3 . \triangleleft

1.2 Dérivées partielles, fonctions de classe \mathcal{C}^1

La définition suivante porte sur des fonctions numériques : on la généralise naturellement à des fonctions à valeurs dans \mathbb{R}^n en considérant les fonctions coordonnées (qui sont à valeurs numériques)

DÉFINITION 1.5: Soit U ouvert de \mathbb{R}^m , $f : U \rightarrow \mathbb{R}$, $\mathbf{x} \in U$ et $\varepsilon > 0$ tel que la boule $B(\mathbf{x}, \varepsilon)$ soit incluse dans U . Si l'application partielle d'indice i en \mathbf{x} , soit

$$t \in]x_i - \varepsilon, x_i + \varepsilon[\mapsto f(\dots, x_{i-1}, t, x_{i+1}, \dots),$$

est dérivable en $t = x_i$, alors on dit que f admet une dérivée partielle d'indice i au point \mathbf{x} , ou une dérivée partielle par rapport à la variable x_i en \mathbf{x} . On note cette dérivée partielle

$$\partial_{x_i} f(\mathbf{x}) = \frac{\partial f}{\partial x_i}(\mathbf{x}).$$

La fonction f admet une dérivée partielle par rapport à la variable x_i définie sur U si elle admet une dérivée partielle par rapport à x_i en chacun des points de U . On a alors

$$\partial_{x_i} f(\mathbf{x}) = \frac{\partial f}{\partial x_i}(\mathbf{x}) = \left[\frac{d}{dt} f(\dots, x_{i-1}, t, x_{i+1}, \dots) \right]_{|t=x_i}.$$

△ REMARQUES 1.4:

1. Pour une fonction $f : t \in U \mapsto f(t)$ numérique d'une variable, on retrouve la définition standard de la dérivée. Les notations $\partial_t f(t) = df(t)$ valent pour la dérivée $f'(t)$.
2. Si on est en 3 variables x, y, z , on écrira $\frac{\partial f}{\partial x}(\mathbf{x}), \frac{\partial f}{\partial y}(\mathbf{x}), \frac{\partial f}{\partial z}(\mathbf{x})$ ou $\partial_x f(\mathbf{x}), \partial_y f(\mathbf{x}), \partial_z f(\mathbf{x})$.
3. On a d'autres notations, par exemple $f'_x(\mathbf{x}), D_x f, D_1 f, \dots$ ▽

▷ EXEMPLES 1.6:

- 1.6.1 Si $r(x, y) = \sqrt{x^2 + y^2}$ pour $(x, y) \in \mathbb{R}^2$, la fonction r admet des dérivées partielles sur $\mathbb{R}^2 \setminus \{(0, 0)\}$, avec

$$\partial_x r(x, y) = \frac{d}{dt} \left[\sqrt{t^2 + y^2} \right]_{|t=x} = \frac{1}{2} \frac{2x}{\sqrt{x^2 + y^2}} = \frac{x}{r(x, y)}, \quad (x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}.$$

et pareillement $\partial_y r(x, y) = y/r(x, y)$ si $(x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}$. La fonction r n'admet pas de dérivées partielles en $\mathbf{m} = (0, 0)$, puisque l'application partielle en ce point et, pour la première coordonnée, est

$$t \in \mathbb{R} \mapsto r(t, 0) = \sqrt{t^2} = |t|,$$

application non dérivable en $t = 0$.

- 1.6.2 La loi des gaz parfaits s'énonce suivant la relation $pv = n\rho t$ reliant les variables de pression p , de volume v et de température t , avec comme paramètres n le nombre de moles et ρ la constante des gaz parfaits. On a les applications

$$p : (v, t) \mapsto p = \frac{n\rho t}{v}, \quad v : (p, t) \mapsto v = \frac{n\rho t}{p}, \quad t : (p, v) \mapsto t = \frac{pv}{n\rho},$$

d'où

$$\partial_v p \partial_t v \partial_p t = \left[-\frac{n\rho t}{v^2} \right] \left[\frac{n\rho}{p} \right] \left[\frac{v}{n\rho} \right] = -1, \quad t \partial_t p \partial_t v = t \left[\frac{n\rho}{v} \right] \left[\frac{n\rho}{p} \right] = n\rho.$$

◁

DÉFINITION 1.6: Soit U ouvert de \mathbb{R}^m . La fonction $f : U \rightarrow \mathbb{R}^n$ est dite de classe \mathcal{C}^1 sur U si ses fonctions coordonnées $f_j : U \rightarrow \mathbb{R}$, $j = 1, \dots, n$ admettent des dérivées partielles $\partial_{x_i} f_j$ de tout indice $i = 1, \dots, m$ sur U , qui sont de plus continues sur U .

La dérivée partielle d'indice i (ou pour la variable x_i) peut s'interpréter comme dérivée dans la direction $\mathbf{e}_i = (\dots, 0, 1, 0, \dots)$, vecteur dont seule la i -ème coordonnée est non nulle :

$$\partial_{x_i} f(\mathbf{m}) = \frac{d}{dt} [f(\mathbf{m} + t\mathbf{e}_i)]_{|t=0}.$$

Est-on capable d'établir l'existence d'une telle dérivée directionnelle

$$\frac{d}{dt} [f(\mathbf{m} + t\mathbf{v})]_{|t=0}. \quad (1.1)$$

dans toute direction $\mathbf{v} \in \mathbb{R}^2$? Autrement dit, l'existence de m dérivées partielles (relativement aux m axes de coordonnées) suffit-elle à établir l'existence de toute dérivée directionnelle ? A priori non, mais la notion de différentielle l'assure comme on le verra dans la section suivante.

1.3 Différentielle en un point

Le théorème suivant a été formulé pour des fonctions de 2 variables, mais il est facilement exprimable pour des fonctions à $m > 2$ variables. Il établit qu'une fonction de classe \mathcal{C}^1 est approchée au voisinage de \mathbf{m} par une fonction affine, à un reste r_f petit près tel que $r_f(\mathbf{v}) = \|\mathbf{v}\|\varepsilon(\mathbf{v})$.

THÉORÈME 1.2: Soit $f : U(\subset \mathbb{R}^2) \rightarrow \mathbb{R}$ de classe \mathcal{C}^1 et $\mathbf{m} \in U$. Alors

$$f(\mathbf{m} + (h, k)) = f(\mathbf{m}) + \partial_x f(\mathbf{m})h + \partial_y f(\mathbf{m})k + \|(h, k)\|\varepsilon(h, k) \quad (1.2)$$

avec $\varepsilon(h, k) \rightarrow 0$ lorsque $(h, k) \rightarrow (0, 0)$.

DÉMONSTRATION. Cette approximation de f provient d'un examen des divers accroissements et de la continuité des applications de dérivée partielle. On a

$$\begin{aligned} f(\mathbf{m} + (h, k)) &= f(\mathbf{m}) + \left[f(\mathbf{m} + (h, k)) - f(\mathbf{m} + (0, k)) \right] + \left[f(\mathbf{m} + (0, k)) - f(\mathbf{m}) \right] \\ &= f(\mathbf{m}) + \int_0^h \frac{d}{dt} [f(\mathbf{m} + (t, k))] dt + \int_0^k \frac{d}{dt} [f(\mathbf{m} + (0, t))] dt \\ &= f(\mathbf{m}) + \int_0^h \partial_{x_1} f(\mathbf{m} + (t, k)) dt + \int_0^k \partial_{x_2} f(\mathbf{m} + (0, t)) dt \\ &= f(\mathbf{m}) + h \int_0^1 \partial_{x_1} f(\mathbf{m} + (hs, k)) ds + k \int_0^1 \partial_{x_2} f(\mathbf{m} + (0, ks)) ds \\ &= f(\mathbf{m}) + h \partial_{x_1} f(\mathbf{m}) + k \partial_{x_2} f(\mathbf{m}) + r(h, k) \end{aligned}$$

avec

$$r(h, k) = h \int_0^1 \left[\partial_{x_1} f(\mathbf{m} + (hs, k)) - \partial_{x_1} f(\mathbf{m}) \right] ds + k \int_0^1 \left[\partial_{x_2} f(\mathbf{m} + (0, ks)) - \partial_{x_2} f(\mathbf{m}) \right] ds$$

On a utilisé les dérivées directionnelles par rapport aux 2 axes de coordonnées dans la deuxième ligne. Puisque les dérivées partielles sont continues (et donc uniformément continues sur tout compact), on peut majorer les intégrands par ε pour (h, k) petit, obtenant

$$|r(h, k)| \leq \|(h, k)\| \varepsilon(\mathbf{m}, h, k) \text{ avec } \varepsilon(\mathbf{m}, h, k) \rightarrow 0 \text{ quand } (h, k) \rightarrow 0$$

pour le reste dans l'expression

$$f(\mathbf{m} + (h, k)) = f(\mathbf{m}) + h \partial_{x_1} f(\mathbf{m}) + k \partial_{x_2} f(\mathbf{m}) + r(h, k). \quad \square$$

Cette approximation de la fonction $f : \mathbf{x} = (x, y) \mapsto f(x, y)$ au voisinage de \mathbf{m} par l'application linéaire affine

$$(h, k) \in \mathbb{R}^2 \mapsto f(\mathbf{m}) + \partial_x f(\mathbf{m})h + \partial_y f(\mathbf{m})k$$

justifie la définition (manipulée précédemment en deux variables)

DÉFINITION 1.7: Soit $f : U(\subset \mathbb{R}^m) \rightarrow \mathbb{R}$ et $\mathbf{m} \in U$. La fonction f est dite différentiable en \mathbf{m} si pour $\mathbf{h} \in \mathbb{R}^n$ au voisinage de 0,

$$f(\mathbf{m} + \mathbf{h}) = f(\mathbf{m}) + df(\mathbf{m})(\mathbf{h}) + \|\mathbf{h}\| \varepsilon(\mathbf{h}) \text{ avec } \varepsilon(\mathbf{h}) \rightarrow 0 \text{ lorsque } \mathbf{h} \rightarrow 0. \quad (1.3)$$

où la différentielle (ou dérivée ou encore l'application linéaire tangente) $df(\mathbf{m}) = f'(\mathbf{m})$ au point \mathbf{m} est définie comme l'application linéaire³

$$\begin{aligned} df(\mathbf{m})(\mathbf{h}) &= df(\mathbf{m})(h, k, \ell) \\ &= \partial_x f(\mathbf{m})h + \partial_y f(\mathbf{m})k + \partial_z f(\mathbf{m})\ell, \\ &= \langle \nabla f(\mathbf{m}), (h, k, \ell) \rangle = \langle \nabla f(\mathbf{m}), \mathbf{h} \rangle, \quad \mathbf{h} = (h, k, \ell) \in \mathbb{R}^3. \end{aligned}$$

3. On se reportera à la section 1.6 et sa formule (1.11) qui définit le gradient $\nabla f(\mathbf{m})$ pour une fonction f à valeurs numériques et dépendant de m variables.

Le théorème suivant a donc été établi :

THÉORÈME 1.3: *Toute fonction $f : U(\subset \mathbb{R}^m) \rightarrow \mathbb{R}$ de classe \mathcal{C}^1 est différentiable.*

et le lemme suivant précisant la formule (1.1) est aisé à démontrer :

LEMME 1.2: *Soit $f : U(\subset \mathbb{R}^2) \rightarrow \mathbb{R}$ de classe \mathcal{C}^1 et $\mathbf{m} \in U$. Alors la fonction f admet une dérivée directionnelle*

$$D_{\mathbf{v}}f(\mathbf{m}) := \frac{d}{dt} [f(\mathbf{m} + t\mathbf{v})]_{t=0} = \partial_{x_1}f(\mathbf{m})v_1 + \partial_{x_2}f(\mathbf{m})v_2, \quad \mathbf{v} = (v_1, v_2).$$

en toute direction \mathbf{v} au point \mathbf{m} .

DÉMONSTRATION. D'après la formule (1.2), on a pour $\mathbf{v} = (v_1, v_2)$,

$$\frac{f(\mathbf{m} + t\mathbf{v}) - f(\mathbf{m})}{t} = \partial_x f(\mathbf{m})v_1 + \partial_y f(\mathbf{m})v_2 + \|\mathbf{v}\|\varepsilon(t\mathbf{v}),$$

ce qui assure la convergence du membre de gauche lorsque $t \rightarrow 0$ et qui donne une expression de la dérivée directionnelle dans la direction \mathbf{v}

$$\frac{d}{dt} [f(\mathbf{m} + t\mathbf{v})]_{t=0} = \partial_x f(\mathbf{m})v_1 + \partial_y f(\mathbf{m})v_2, \quad \mathbf{v} = (v_1, v_2). \quad \square$$

△ REMARQUE 1.5: Les opérations usuelles (addition, produit, ...) préservent la différentiabilité et les formules du calcul en une variable se prolongent souvent aisément. Ainsi, pour des fonctions numériques $f, g : \mathbb{R}^m \rightarrow \mathbb{R}$, on a par exemple

$$d(f + g) = df + dg, \quad d(fg) = f dg + g df, \quad d(f^{-1}) = -f^{-2} df.$$

Une fonction constante a une différentielle nulle, tandis qu'une fonction linéaire L a comme différentielle L elle même. Ces propriétés découlent directement de la définition précédente. ∇

Pour des fonctions à valeurs vectorielles, la définition précédente 1.7 reste complètement valable, à quelques modifications près.

DÉFINITION 1.8: *Si $f = (f_1, \dots, f_n) : U(\subset \mathbb{R}^m) \rightarrow \mathbb{R}^n$, la différentielle au point \mathbf{m} est définie comme l'application linéaire $df(\mathbf{m}) : \mathbb{R}^m \rightarrow \mathbb{R}^n$ avec matrice dite jacobienne⁴ dans les bases canoniques*

$$J_f(\mathbf{m}) = \begin{pmatrix} \partial_{x_1} f_1(\mathbf{m}) & \dots & \partial_{x_m} f_1(\mathbf{m}) \\ \vdots & & \vdots \\ \partial_{x_1} f_n(\mathbf{m}) & \dots & \partial_{x_m} f_n(\mathbf{m}) \end{pmatrix}. \quad (1.4)$$

de telle sorte que pour $\mathbf{v} = (v_1, \dots, v_m) \in \mathbb{R}^m$

$$df(\mathbf{m})(\mathbf{v}) = \left(\sum_{i=1}^m \partial_{x_i} f_1(\mathbf{m})v_i, \dots, \sum_{i=1}^m \partial_{x_i} f_n(\mathbf{m})v_i \right) \in \mathbb{R}^n.$$

4. Charles J. K. Jacobi, 10 décembre 1804, Potsdam, Prusse – 18 février 1851, Berlin, Prusse.

△ REMARQUES 1.6:

1. Si on considère des fonctions numériques $f, g : \mathbb{R} \rightarrow \mathbb{R}$ d'une variable, on retrouve la formule classique pour la dérivée d'une fonction composée : $(g \circ f)'(t) = g'(f(t))f'(t)$.
2. La matrice $J_f(\mathbf{m})$ représente la différentielle $f'(\mathbf{m}) = df(\mathbf{m})$, application linéaire de \mathbb{R}^m dans \mathbb{R}^n , dans les bases canoniques de \mathbb{R}^m et \mathbb{R}^n . Cette matrice a comme nombre de colonnes celui des variables (la dimension de l'espace contenant l'ouvert source de f) et comme nombre de lignes celui des fonctions coordonnées (f_1, \dots, f_n) de f (la dimension de son espace but). Si on convient d'écrire le gradient ∇f comme un vecteur ligne, la matrice jacobienne reprend sur chaque ligne les gradients ∇f_j des fonctions coordonnées f_j pour $j = 1, \dots, n$.
3. Pour effectuer le produit des tableaux $J_f(\mathbf{m})$ et (v_1, \dots, v_m) il convient de placer ces tableaux comme ceci

$$\begin{pmatrix} \partial_{x_1} f_1(\mathbf{m}) & \dots & \partial_{x_m} f_1(\mathbf{m}) \\ \vdots & & \vdots \\ \partial_{x_1} f_n(\mathbf{m}) & \dots & \partial_{x_m} f_n(\mathbf{m}) \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix}$$

On effectuera le produit de chaque ligne avec le vecteur colonne des coordonnées (v_1, \dots, v_m) , de telle manière que le produit final est un vecteur colonne de m lignes. ▽

La prochaine étape est le calcul des dérivées partielles de fonctions composées : le théorème 1.4 ci-dessous l'exprime de manière lapidaire. La correspondance entre matrice et application linéaire fait l'objet d'une annexe à la fin de ce chapitre : l'équivalence des formules (1.5) et (1.6) en résultent, c'est son application en basses dimensions qui importe le plus : quelques exemples simples de calculs de dérivées partielles de fonctions composées achèvent cette section.

THÉORÈME 1.4: Soit $f : U(\subset \mathbb{R}^m) \rightarrow \mathbb{R}^n$ et $g : V(\subset \mathbb{R}^n) \rightarrow \mathbb{R}^p$, fonctions de classe \mathcal{C}^1 avec $f(U) \subset V$. Alors la composée $h = g \circ f$ est de classe \mathcal{C}^1 et

$$d(g \circ f)(\mathbf{m}) = dg(f(\mathbf{m})) \circ df(\mathbf{m}), \quad \mathbf{m} \in U, \quad (1.5)$$

ou pour les matrices jacobiennes

$$J_{g \circ f}(\mathbf{m}) = J_g(f(\mathbf{m})) \cdot J_f(\mathbf{m}), \quad \mathbf{m} \in U, \quad (1.6)$$

soit en terme de dérivées partielles (coefficients des matrices jacobiniennes), on a

$$\frac{\partial(g \circ f)_\ell}{\partial x_i}(\mathbf{m}) = \sum_{j=1}^n \frac{\partial g_\ell}{\partial y_j}(f(\mathbf{m})) \frac{\partial f_j}{\partial x_i}(\mathbf{m}), \quad \mathbf{m} \in U, i = 1, \dots, m, \ell = 1, \dots, p. \quad (1.7)$$

où on a supposé $f = (f_1, \dots, f_n)$ avec m variables x_1, \dots, x_m et n fonctions coordonnées, $g = (g_1, \dots, g_p)$ avec n variables y_1, \dots, y_n et p fonctions coordonnées et donc $h = (h_1, \dots, h_p)$ avec m variables x_1, \dots, x_m et p fonctions coordonnées.

DÉMONSTRATION. La différentiabilité de f en \mathbf{m} et de g en $f(\mathbf{m})$ signifient les approximations par des fonctions affines avec des restes « petits »

$$\begin{aligned} f(\mathbf{m} + \mathbf{v}) &= f(\mathbf{m}) + f'(\mathbf{m})(\mathbf{v}) + r_f(\mathbf{v}) \\ g(f(\mathbf{m}) + \mathbf{w}) &= g(f(\mathbf{m})) + g'(f(\mathbf{m}))(\mathbf{w}) + r_g(\mathbf{w}) \end{aligned}$$

avec r_f, r_g petits devant les autres termes en le sens indiqué par la définition de la différentiabilité. Notant $\mathbf{w} = f'(\mathbf{m})(\mathbf{v}) + r_f(\mathbf{v})$, on obtient

$$\begin{aligned} (g \circ f)(\mathbf{m} + \mathbf{v}) &= g(f(\mathbf{m} + \mathbf{v})) = g(f(\mathbf{m}) + f'(\mathbf{m})(\mathbf{v}) + r_f(\mathbf{v})) \\ &= g(f(\mathbf{m}) + f'(\mathbf{m})(\mathbf{v}) + r_f(\mathbf{v})) \\ &= g(f(\mathbf{m})) + g'(f(\mathbf{m}))(\mathbf{w}) + r_g(\mathbf{w}) \\ &= g(f(\mathbf{m})) + g'(f(\mathbf{m}))(f'(\mathbf{m})(\mathbf{v}) + r_f(\mathbf{v})) + r_g(f'(\mathbf{m})(\mathbf{v}) + r_f(\mathbf{v})) \\ &= g(f(\mathbf{m})) + g'(f(\mathbf{m}))(f'(\mathbf{m})(\mathbf{v})) + r(\mathbf{v}) \end{aligned}$$

avec le reste

$$r(\mathbf{v}) = g'(f(\mathbf{m}))(r_f(\mathbf{v})) + r_g(f'(\mathbf{m})\mathbf{v} + r_f(\mathbf{v}))$$

Pour une application linéaire $A: \mathbb{R}^p \rightarrow \mathbb{R}^q$, on a l'existence d'une constante C_A telle que $\|A\mathbf{v}\| \leq C_A \|\mathbf{v}\|$ pour tout $\mathbf{v} \in \mathbb{R}^p$. On a donc pour $r(t)$ et des constantes C_1, C_2, \dots convenables

$$\begin{aligned} \|r(\mathbf{v})\| &\leq C_1 \|\mathbf{v}\| \varepsilon_f(\mathbf{v}) + \|f'(\mathbf{m})(\mathbf{v}) + r_f(\mathbf{v})\| \varepsilon_g(f'(\mathbf{m})(\mathbf{v}) + r_f(\mathbf{v})) \\ &\leq C_1 \|\mathbf{v}\| \varepsilon_f(\mathbf{v}) + (C_2 \|\mathbf{v}\| + \|\mathbf{v}\| \varepsilon_f(\mathbf{v})) \varepsilon_g(f'(\mathbf{m})(\mathbf{v}) + r_f(\mathbf{v})) \\ &\leq \|\mathbf{v}\| \left[C_1 \varepsilon_f(\mathbf{v}) + (C_2 + \varepsilon_f(\mathbf{v})) \varepsilon_g(f'(\mathbf{m})(\mathbf{v}) + r_f(\mathbf{v})) \right] \end{aligned}$$

ce qui implique la petitesse attendue du reste $r(\mathbf{v})$ et permet de conclure : le terme d'ordre 1 dans le développement de Taylor est celui annoncé dans l'énoncé du théorème. \square

La composée $g \circ f$ de fonctions de classe \mathcal{C}^1 est une fonction de classe \mathcal{C}^1 . Les formules générales des dérivées partielles de la fonction composée en fonction des dérivées partielles des constituants de la composée sont un peu complexes, nous nous limiterons à l'exemple suivant (cf. les formules (1.5) et (1.6) de la composée de la différentielle ci-dessus)

▷ EXEMPLES 1.7:

1.7.1 Soient $f : t \in \mathbb{R} \mapsto (u(t), v(t)) \in \mathbb{R}^2$ et $g : (x, y) \in \mathbb{R}^2 \mapsto g(x, y) \in \mathbb{R}$. Les matrices jacobiniennes respectives sont

$$J_f(t) = \begin{pmatrix} u'(t) \\ v'(t) \end{pmatrix}, \quad t \in \mathbb{R}, \quad J_g(\mathbf{x}) = (\partial_x g(\mathbf{x}) \quad \partial_y g(\mathbf{x})), \quad \mathbf{x} = (x, y) \in \mathbb{R}^2,$$

d'où la matrice de la composée $g \circ f$ qui est une fonction d'une variable t

$$\begin{aligned} \frac{d(g \circ f)}{dt}(t) &= J_{g \circ f}(t) = J_g(f(t))J_f(t) \\ &= (\partial_x g(f(t)) \quad \partial_y g(f(t))) \begin{pmatrix} u'(t) \\ v'(t) \end{pmatrix} \\ &= \partial_x g(f(t))u'(t) + \partial_y g(f(t))v'(t) \\ &= \partial_x g(u(t), v(t))u'(t) + \partial_y g(u(t), v(t))v'(t). \end{aligned}$$

La composée $f \circ g$ est une fonction de deux variables à valeurs vectorielles dans \mathbb{R}^2 . Sa différentielle est représentée par une matrice carrée d'ordre 2

$$\begin{aligned} d(f \circ g)(\mathbf{x}) &= df(g(\mathbf{x})) \circ dg(\mathbf{x}) = J_f(g(\mathbf{x})) \cdot J_g(\mathbf{x}) \\ &= \begin{pmatrix} \partial_t u(g(\mathbf{x})) \\ \partial_t v(g(\mathbf{x})) \end{pmatrix} (\partial_x g(\mathbf{x}) \quad \partial_y g(\mathbf{x})) \\ &= \begin{pmatrix} \partial_t u(g(\mathbf{x}))\partial_x g(\mathbf{x}) & \partial_t u(g(\mathbf{x}))\partial_y g(\mathbf{x}) \\ \partial_t v(g(\mathbf{x}))\partial_x g(\mathbf{x}) & \partial_t v(g(\mathbf{x}))\partial_y g(\mathbf{x}) \end{pmatrix} \end{aligned}$$

1.7.2 Soient $f : \mathbf{x} = (x, y, z) \in \mathbb{R}^3 \mapsto (u(\mathbf{x}), v(\mathbf{x})) \in \mathbb{R}^2$ et $g : (u, v) \in \mathbb{R}^2 \mapsto g(u, v) \in \mathbb{R}$, toutes deux de classe \mathcal{C}^1 . Alors,

$$(g \circ f)(\mathbf{x}) = g(u(x, y, z), v(x, y, z)), \quad \mathbf{x} = (x, y, z) \in \mathbb{R}^3$$

et

$$\begin{aligned} \partial_y(g \circ f)(\mathbf{x}) &= \partial_y [g(u(x, y), v(x, y))] \\ &= \partial_u g(u(\mathbf{x}), v(\mathbf{x}))\partial_y u(\mathbf{x}) + \partial_v g(u(\mathbf{x}), v(\mathbf{x}))\partial_y v(\mathbf{x}), \quad \mathbf{x} = (x, y, z) \in \mathbb{R}^3. \end{aligned}$$

1.7.3 Soient

$$\begin{aligned} f : (x, y) \in \mathbb{R}^2 &\mapsto (x + y^4, y - 3x^2, 2x^3 - 3y) = (f_1(x, y), f_2(x, y), f_3(x, y)) \in \mathbb{R}^3 \\ g : (x, y, z) \in \mathbb{R}^3 &\mapsto 2xy - 3(x + z) \in \mathbb{R}. \end{aligned}$$

Calculons les dérivées partielles de $h = g \circ f$, fonction de 2 variables à valeurs numériques. La formulation en chaîne donne

$$\begin{aligned} \partial_x h(\mathbf{m}) &= \partial_{y_1} g(f(\mathbf{m}))\partial_x(x + y^4) + \partial_{y_2} g(f(\mathbf{m}))\partial_x(y - 3x^2) + \partial_{y_3} g(f(\mathbf{m}))\partial_x(2x^2 - 3y) \\ \partial_y h(\mathbf{m}) &= \partial_{y_1} g\partial_y(x + y^4) + \partial_{y_2} g\partial_y(y - 3x^2) + \partial_{y_3} g\partial_y(2x^2 - 3y) \end{aligned}$$

On aurait pu aussi écrire

$$\partial_x h = \partial_x g \partial_x (x + y^4) + \partial_y g \partial_x (y - 3x^2) + \partial_z g \partial_x (2x^3 - 3y)$$

au risque de confusion entre les variables. Ainsi

$$\partial_x h(x, y, z) = [(2y - 3) \cdot 1] + [2x \cdot (-6x)] + [-3 \cdot 6x^2] = 2y - 3 - 30x^2.$$

◁

1.4 Espace tangent

Soit $\varphi : x \in \mathbb{R} \mapsto \varphi(x) \in \mathbb{R}$ une fonction différentiable. Son graphe

$$G_\varphi = \{(x, \varphi(x)), x \in \mathbb{R}\} \quad (\subset \mathbb{R}_{x,y}^2),$$

est une courbe dans le plan possédant au point $(x_0, \varphi(x_0)) \in \mathbb{R}^2$ une tangente d'équation

$$y = \varphi(x_0) + \varphi'(x_0)(x - x_0), \quad (x, y) \in \mathbb{R}^2.$$

Cette approximation de la courbe par sa tangente est la traduction géométrique de l'approximation analytique

$$\varphi(x) = \varphi(x_0) + \varphi'(x_0)(x - x_0) + (x - x_0)\varepsilon(x), \quad x \in \mathbb{R},$$

où $\varepsilon(x) \rightarrow 0$ lorsque $x \rightarrow x_0$.

Pour une fonction f de m (> 1) variables, on a un phénomène analogue d'approximation du graphe par un hyperplan. Si $m = 2$, cet hyperplan n'est qu'un plan habituel dans \mathbb{R}^3 . Explicitons cette approximation dans le cas d'une fonction de $m = 2$ variables.

Soit $a \in \mathbb{R}$, soit b l'application linéaire $b : \mathbf{u} = (u_1, u_2) \mapsto b_1 u_1 + b_2 u_2$ (avec b non nul, *i. e.* $(b_1, b_2) \neq (0, 0)$) et soit la fonction affine (polynomiale de degré 1) $L_{a,b} : \mathbf{u} \mapsto a + b(\mathbf{u})$. Cette fonction $L_{a,b}$ a pour graphe le plan

$$P_{a,b} = \{(u_1, u_2, L_{a,b}(\mathbf{u})), \mathbf{u} = (u_1, u_2) \in \mathbb{R}^2\} \quad (1.8)$$

d'équation $u_3 = L_{a,b}(u_1, u_2)$ dans $\mathbb{R}_{u_1, u_2, u_3}^3$.

Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ différentiable en \mathbf{x}_0 . Alors, pour \mathbf{x} voisin de \mathbf{x}_0 ,

$$f(\mathbf{x}) = f(\mathbf{x}_0) + df(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \|\mathbf{x} - \mathbf{x}_0\|\varepsilon(\mathbf{x})$$

avec $\varepsilon(\mathbf{x}) \rightarrow 0$ lorsque $\mathbf{x} \rightarrow \mathbf{x}_0$. Ainsi le graphe

$$G_f = \{(x_1, x_2, f(x_1, x_2)), (x_1, x_2) \in \mathbb{R}^2\}$$

au voisinage de x_0 est approché par le plan

$$P_{f, \mathbf{x}_0} = \{(x_1, x_2, f(\mathbf{x}_0) + df(\mathbf{x}_0)(BFx - \mathbf{x}_0)) \in \mathbb{R}^3, \quad \mathbf{x} = (x_1, x_2)\}.$$

Avec les notations précédentes de (1.8), le scalaire a a la valeur $f(\mathbf{x}_0)$ et le vecteur $\mathbf{b} = (b_1, b_2)$ est le gradient $\nabla f(\mathbf{x}_0) = (\partial_x f(\mathbf{x}_0), \partial_y f(\mathbf{x}_0))$.

On a donc démontré

PROPOSITION 1.1: Soit $f : U(\subset \mathbb{R}^2) \rightarrow \mathbb{R}$ définie au voisinage de $\mathbf{x}_0 \in \mathbb{R}^2$ et différentiable en \mathbf{x}_0 . Alors, dans l'espace \mathbb{R}^3 , le graphe de f au point $(\mathbf{x}_0, f(\mathbf{x}_0))$ est approché par le plan tangent en $(\mathbf{x}_0, f(\mathbf{x}_0))$ d'équation dans \mathbb{R}^3

$$x_3 = f(\mathbf{x}_0) + df(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0), \quad (\mathbf{x}_0, x_3) = (x_1, x_2, x_3) \in \mathbb{R}^3.$$

1.5 Dérivées d'ordre supérieur

Les dérivées partielles peuvent être dérivées, d'où la définition suivante, prolongeant celle (déf. 1.6) de fonction de classe \mathcal{C}^1 ,

DÉFINITION 1.9: Soit k un entier au moins égal à 2. La fonction $f : U(\subset \mathbb{R}^m) \rightarrow \mathbb{R}$ est dite de classe \mathcal{C}^k si elle est de classe \mathcal{C}^1 et ses m dérivées partielles $\partial_{x_i} f$ où $i = 1, \dots, m$ sont de classe \mathcal{C}^{k-1} . La fonction f est dite de classe \mathcal{C}^∞ ou indéfiniment différentiable, si elle est de classe \mathcal{C}^k pour tout entier k .

▷ EXEMPLE 1.8:

Les fonctions polynômes sont de classe \mathcal{C}^∞ , ainsi que les fonctions trigonométriques, les fonctions exp et ln (sur son domaine de définition), etc ◁

On a le théorème de Schwarz⁵ qui prend la forme suivante pour les fonctions de deux variables.

THÉORÈME 1.5: Soit $f : U \rightarrow \mathbb{R}$ de classe \mathcal{C}^2 . Alors,

$$\frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right) (\mathbf{x}) = \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right) (\mathbf{x}), \quad \mathbf{x} \in U.$$

On notera cette valeur commune $\partial_{xy}^2 f(\mathbf{x}) = \frac{\partial^2 f}{\partial x \partial y}(\mathbf{x})$.

5. H. Schwarz 25 janvier 1843, Hermsdorf, Silésie – 30 novembre 1921, Berlin, Allemagne.

DÉMONSTRATION. Il suffit de le montrer pour une fonction f de deux variables x, y différentiable à l'origine $(0, 0)$, ainsi que ses dérivées partielles d'ordre 1. Soient F et H définies au voisinage de l'origine par

$$F(x, y) = f(x, y) - f(x, 0) - f(0, y) + f(0, 0), \quad H(x, y) = F(x, y) - xy\partial_y\partial_x F(0, 0).$$

On a

$$\partial_x H(x, y) = \partial_x F(x, y) - y\partial_y\partial_x F(0, 0).$$

Par ailleurs, la différentiabilité de $\partial_x F$ à l'origine permet d'écrire

$$\partial_x F(s, t) = \partial_x F(0, 0) + \partial_x\partial_x F(0, 0)s + \partial_y\partial_x F(0, 0)t + (|s| + |t|)\varepsilon(s, t). \quad (1.9)$$

Vu les égalités

$$\partial_x F(x, y) = \partial_x f(x, y) - \partial_x f(x, 0), \quad \partial_x F(x, 0) = 0, \quad \partial_x\partial_x F(0, 0) = 0,$$

la formule précédente (1.9) se réduit à

$$\partial_x F(s, t) = t\partial_y\partial_x F(0, 0) + (|s| + |t|)\varepsilon(s, t).$$

On en déduit par l'inégalité des accroissements finis avec y fixé, et $\varepsilon_1(x, y) = \sup_{s \in [0, x]} |\varepsilon(s, y)|$,

$$|H(x, y)| \leq \sup_{s \in [0, x]} |\partial_x F(s, y) - y\partial_y\partial_x F(0, 0)| |x| \leq (|x| + |y|)\varepsilon_1(x, y)|x|$$

soit

$$|F(x, y) - xy\partial_y\partial_x F(0, 0)| \leq (|x| + |y|)\varepsilon_1(x, y)|x|.$$

On a une inégalité analogue en échangeant x et y

$$|F(x, y) - yx\partial_x\partial_y F(0, 0)| \leq (|x| + |y|)\varepsilon_2(x, y)|y|,$$

d'où en additionnant, et en prenant $t = x = y$,

$$|t^2\partial_x\partial_y F(0, 0) - t^2\partial_y\partial_x F(0, 0)| \leq 2t^2(\varepsilon_1(x, y) + \varepsilon_2(x, y))$$

soit

$$|\partial_x\partial_y F(0, 0) - \partial_y\partial_x F(0, 0)| \leq 2(\varepsilon_1(x, y) + \varepsilon_2(x, y)),$$

et donc l'annulation du membre de gauche en faisant tendre (x, y) vers $(0, 0)$, ce qu'il fallait démontrer. \square

▷ EXEMPLES 1.9:

1.9.1 Soit la fonction $f : (x, y) \in (\mathbb{R}^{+*})^2 \mapsto \sqrt{x^3 y}$. Les dérivées partielles prennent la forme

$$\begin{aligned} \partial_x f(x, y) &= \frac{3}{2}x^{1/2}y^{1/2}, & \partial_y(x, y)f(x, y) &= \frac{1}{2}x^{3/2}y^{-1/2}, \\ \partial_{x^2}^2 f(x, y) &= \frac{3}{4}x^{-1/2}y^{1/2}, & \partial_{yx}^2 f(x, y) &= \frac{3}{2}x^{1/2} \left[\frac{1}{2}y^{-1/2} \right] = \frac{3}{4}x^{1/2}y^{-1/2}, \\ \partial_{xy}^2 f(x, y) &= \frac{1}{2} \left[\frac{3}{2}x^{1/2} \right] y^{-1/2} = \frac{3}{4}x^{1/2}y^{-1/2}, & \partial_{y^2}^2 f(x, y) &= -\frac{1}{4}x^{3/2}y^{-3/2}. \end{aligned}$$

1.9.2 L'application f définie suivant

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

a des dérivées partielles d'ordre 2 à l'origine avec $\partial_x \partial_y f(0,0) = -1$ et $\partial_y \partial_x f(0,0) = 1$. En effet, la dérivée partielle $\partial_x f$ se calcule aisément en dehors de l'origine

$$\partial_x f(x, y) = \frac{y(x^2 - y^2)}{x^2 + y^2} + \frac{xy(2x)}{x^2 + y^2} + \frac{-2xy(x^2 - y^2)}{(x^2 + y^2)^2}$$

avec $\partial_x f(0,0)$ obtenu par prolongement par continuité en $(0,0)$. Ainsi, pour y non nul, $\partial_x f(0, y) = -y$ et donc $\partial_y \partial_x f(0,0) = -1$ avec, en échangeant le rôle de x et y , $\partial_x \partial_y f(0, y) = 1$, \triangleleft

DÉFINITION 1.10: Soit $f : \mathbf{x} \in U(\subset \mathbb{R}^m) \mapsto f(\mathbf{x}) \in \mathbb{R}$. Si f est de classe \mathcal{C}^2 , alors la matrice hessienne⁶ de f est la matrice d'ordre m dépendant de $\mathbf{x} \in U$ définie comme suivant

$$\text{Hess } f(\mathbf{x}) = \left(\partial_{x_i x_j}^2 f(\mathbf{x}) \right), \quad \mathbf{x} \in U.$$

Ainsi, pour une fonction de $m = 3$ variables, on a

$$\text{Hess } f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(\mathbf{x}) & \frac{\partial^2 f}{\partial x \partial y}(\mathbf{x}) & \frac{\partial^2 f}{\partial x \partial z}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial x \partial y}(\mathbf{x}) & \frac{\partial^2 f}{\partial y^2}(\mathbf{x}) & \frac{\partial^2 f}{\partial y \partial z}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial x \partial z}(\mathbf{x}) & \frac{\partial^2 f}{\partial y \partial z}(\mathbf{x}) & \frac{\partial^2 f}{\partial z^2}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \partial_{x^2}^2 f(\mathbf{x}) & \partial_{xy}^2 f(\mathbf{x}) & \partial_{xz}^2 f(\mathbf{x}) \\ \partial_{xy}^2 f(\mathbf{x}) & \partial_{y^2}^2 f(\mathbf{x}) & \partial_{yz}^2 f(\mathbf{x}) \\ \partial_{xz}^2 f(\mathbf{x}) & \partial_{yz}^2 f(\mathbf{x}) & \partial_{z^2}^2 f(\mathbf{x}) \end{pmatrix}$$

La matrice hessienne est symétrique. Elle est déterminée par 3 (6 resp.) dérivées partielles d'ordre 2 pour les fonctions de 2 (3 resp.) variables.

▷ EXEMPLE 1.10: Soit $f : \mathbf{m} = (x, y, z) \in \mathbb{R}^3 \mapsto f(\mathbf{m}) = xy + z^2 \in \mathbb{R}$. Son gradient⁷ est

$$\nabla f(\mathbf{m}) = (y, x, 2z), \quad \mathbf{m} = (x, y, z)$$

et sa matrice hessienne

$$\text{Hess } f(\mathbf{m}) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad \mathbf{m} = (x, y, z).$$

La matrice hessienne de f est constante : c'est le cas pour toute fonction polynomiale de degré au plus 2, mais pas pour une application générale. \triangleleft

Prolongeant l'approximation d'une fonction par sa différentielle, on a des développements polynomiaux de degré k qui approchent de plus en plus une fonction. Limitons nous au cas de fonctions numériques et à l'ordre $k=2$.

6. Ludwig Otto Hesse, 22 avril 1811, Königsberg, Prusse – 4 août 1874, Munich, Allemagne.

7. Le gradient est un vecteur. Il est représenté en principe par un vecteur colonne de m lignes, on l'écrira de manière abusive comme un vecteur ligne.

THÉORÈME 1.6: Soit $f : U(\subset \mathbb{R}^m) \rightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^2 et $\mathbf{x} \in U$. Alors on a le développement de Taylor à l'ordre 2 au voisinage de \mathbf{x} :

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \sum_{j=1}^m \partial_{x_j} f(\mathbf{x}) h_j + \frac{\sum_{i,j=1}^m \partial_{x_i x_j}^2 f(\mathbf{x}) h_i h_j}{2} + r(\mathbf{h})$$

soit avec les opérateurs Hess $f, \nabla f$

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{h} \rangle + \frac{\langle \mathbf{h}, \text{Hess } f(\mathbf{x}) \mathbf{h} \rangle}{2} + r(\mathbf{h}). \quad (1.10)$$

où $\mathbf{h} \in \mathbb{R}^m$ est suffisamment petit (de telle manière que $\mathbf{x} + \mathbf{h} \in U$) et où le reste r est de la forme $r(\mathbf{h}) = \|\mathbf{h}\|^2 \varepsilon(\mathbf{h})$ avec $\varepsilon(\mathbf{h}) \rightarrow 0$ lorsque $\mathbf{h} \rightarrow 0$.

△ REMARQUES 1.7:

1. En dimension $m = 2$, on a en général

$$f(\mathbf{x} + (h, k)) = f(\mathbf{x}) + \partial_x f(\mathbf{x}) h + \partial_y f(\mathbf{x}) k + \frac{1}{2} \partial_{xx}^2 f(\mathbf{x}) h^2 + \partial_{xy}^2 f(\mathbf{x}) h k + \frac{1}{2} \partial_{yy}^2 f(\mathbf{x}) k^2 + (h^2 + k^2) \varepsilon(h, k).$$

2. En 3 variables, l'exemple précédent 1.10 donne le développement

$$f(\mathbf{m} + \mathbf{h}) = xy + z^2 + yh + xk + 2z\ell + \frac{hk + \ell^2}{2}, \quad \mathbf{m} = (x, y, z), \mathbf{h} = (h, k, \ell) \in \mathbb{R}^3,$$

où le reste r (de la formule générale (1.10)) est identiquement nul : une fonction quadratique a un développement de Taylor se terminant par le terme d'ordre 2 (cf. la remarque suivante).

3. Soit ℓ un vecteur de \mathbb{R}^m et A une matrice (non nécessairement) symétrique d'ordre m . La fonction quadratique

$$F_{\ell, A} : \mathbf{x} \in \mathbb{R}^m \mapsto \langle \mathbf{x}, \ell \rangle + \langle \mathbf{x}, A\mathbf{x} \rangle.$$

a pour gradient $\nabla F_{\ell, A}$ et hessienne Hess $F_{\ell, A}$ les applications

$$\nabla F_{\ell, A}(\mathbf{x}) = \ell + (A + {}^T A)\mathbf{x}, \quad \text{Hess } F_{\ell, A}(\mathbf{x}) = (A + {}^T A)/2, \quad \mathbf{x} \in \mathbb{R}^m,$$

la hessienne étant constante. En effet, on a le développement de Taylor directement calculé

$$\begin{aligned} F_{\ell, A}(\mathbf{x} + \mathbf{v}) &= \langle \mathbf{x} + \mathbf{v}, \ell \rangle + \langle \mathbf{x} + \mathbf{v}, A(\mathbf{x} + \mathbf{v}) \rangle \\ &= \langle \mathbf{x}, A\mathbf{x} \rangle + \langle \mathbf{x}, \ell \rangle + \langle \mathbf{v}, A(\mathbf{x} + \mathbf{v}) \rangle + \langle \mathbf{x}, A\mathbf{v} \rangle + \langle \mathbf{v}, \ell \rangle \\ &= F_{\ell, A}(\mathbf{x}) + \langle \mathbf{v}, A\mathbf{x} \rangle + \langle \mathbf{x}, A\mathbf{v} \rangle + \langle \mathbf{v}, \ell \rangle + \langle \mathbf{v}, A\mathbf{v} \rangle \\ &= F_{\ell, A}(\mathbf{x}) + \langle \mathbf{v}, A\mathbf{x} + {}^T A\mathbf{x} + \ell \rangle + \frac{\langle \mathbf{v}, (A + {}^T A)\mathbf{v} \rangle}{2}, \end{aligned}$$

qui sera identifié au développement général (1.10) (ici le reste $r(\mathbf{v})$ est identiquement nul)

$$F_{\ell,A}(\mathbf{x} + \mathbf{v}) = F_{\ell,A}(\mathbf{x}) + \langle \nabla F_{\ell,A}(\mathbf{x}), \mathbf{v} \rangle + \frac{\langle \mathbf{v}, \text{Hess } F_{\ell,A}(\mathbf{x}) \mathbf{v} \rangle}{2}$$

Si la matrice A est symétrique, alors le gradient est égal à $\nabla F_{A,\ell}(\mathbf{x}) = 2A\mathbf{x} + \ell$ et la hessienne égale à la matrice $2A$. En dimension $m = 1$, on retrouve la fonction $F_{A,\ell} : t \in \mathbb{R} \mapsto Ax^2 + \ell t$ (où A, ℓ sont des réels), de dérivée première $2Ax + \ell$ et dérivée seconde $2A$. ∇

DÉMONSTRATION. La preuve consiste à faire quelques intégrations du type $\varphi(b) - \varphi(a) = \int_0^1 \frac{d}{dt} [\varphi(a + t(b-a))] dt$. Ainsi

$$\begin{aligned} f(\mathbf{x} + \mathbf{v}) &= f(\mathbf{x}) + \int_0^1 \frac{d}{dt} [f(\mathbf{x} + t\mathbf{v})] dt \\ &= f(\mathbf{x}) + \int_0^1 \sum_{j=1}^m \partial_{x_j} f(\mathbf{x} + t\mathbf{v}) v_j dt \\ &= f(\mathbf{x}) + \int_0^1 \sum_{j=1}^m \partial_{x_j} f(\mathbf{x}) v_j dt + \int_0^1 \sum_{j=1}^m [\partial_{x_j} f(\mathbf{x} + t\mathbf{v}) - \partial_{x_j} f(\mathbf{x})] v_j dt \\ &= f(\mathbf{x}) + \sum_{j=1}^m \partial_{x_j} f(\mathbf{x}) v_j + \sum_{j=1}^m \int_0^1 [\partial_{x_j} f(\mathbf{x} + t\mathbf{v}) - \partial_{x_j} f(\mathbf{x})] v_j dt \\ &= f(\mathbf{x}) + \sum_{j=1}^m \partial_{x_j} f(\mathbf{x}) v_j + \sum_{j=1}^m \int_0^1 \left[\int_0^1 \frac{d}{ds} [\partial_{x_j} f(\mathbf{x} + t\mathbf{sv})] ds \right] v_j dt \\ &= f(\mathbf{x}) + \sum_{j=1}^m \partial_{x_j} f(\mathbf{x}) v_j + \sum_{i,j=1}^m \int_0^1 \left[\int_0^1 t [\partial_{x_i x_j}^2 f(\mathbf{x} + t\mathbf{sv})] ds \right] dt v_i v_j \\ &= f(\mathbf{x}) + \sum_{j=1}^m \partial_{x_j} f(\mathbf{x}) v_j + \sum_{i,j=1}^m \int_0^1 t \left[\int_0^1 [\partial_{x_i x_j}^2 f(\mathbf{x})] ds \right] dt v_i v_j + r(\mathbf{v}) \\ &= f(\mathbf{x}) + \sum_{j=1}^m \partial_{x_j} f(\mathbf{x}) v_j + \frac{1}{2} \sum_{i,j=1}^m [\partial_{x_i x_j}^2 f(\mathbf{x}) v_i v_j] + r(\mathbf{v}) \end{aligned}$$

Le $1/2$ dans la dernière équation provient de l'intégrale $\int_0^1 \left[t \int_0^1 ds \right] dt = 1/2$. Le reste $r(\mathbf{v})$ est

$$r(\mathbf{v}) = \sum_{i,j=1}^m \int_0^1 t \left[\int_0^1 [\partial_{x_i x_j}^2 f(\mathbf{x} + t\mathbf{sv}) - \partial_{x_i x_j}^2 f(\mathbf{x})] v_i v_j ds \right] dt$$

avec l'estimation

$$|r(\mathbf{v})| \leq \sum_{i,j=1}^m \int_0^1 t \left[\int_0^t |\partial_{x_i x_j}^2 f(\mathbf{x} + \mathbf{sv}) - \partial_{x_i x_j}^2 f(\mathbf{x})| |v_i| \cdot |v_j| ds \right] dt \leq \frac{1}{2} \sum_{i,j=1}^m M_{i,j}(\|\mathbf{v}\|) \|\mathbf{v}\|^2.$$

où les majorants $M_{i,j}$ sont définis suivant

$$M_{i,j}(r) = \sup_{\|v\| \leq r} \left| \partial_{x_i x_j} f(\mathbf{x} + v) - \partial_{x_i x_j} f(\mathbf{x}) \right|$$

et tendent vers 0 quand $r \rightarrow 0$ par continuité uniforme des dérivées partielles $\partial_{x_i x_j}^2 f$ sur la boule fermée $\overline{B(0, r)}$ (incluse dans U) centrée en 0 et de rayon r . \square

1.6 Opérateurs différentiels : gradient, rotationnel, divergence

Dans la suite, on supposera les fonctions toutes de classe \mathcal{C}^∞ . L'analyse vectorielle considère des champs de scalaires (*i. e.* des fonctions à valeurs numériques, réelles ou complexes), des champs de vecteurs (*i. e.* des fonctions à valeurs dans \mathbb{R}^m où m est le nombre de variable, dont les valeurs sont interprétées souvent comme des vecteurs vitesse) et des opérateurs qui transforment un de ces champs en un champ du même type ou non.

DÉFINITION 1.11: 1. Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ et $g : \mathbb{R}^3 \rightarrow \mathbb{R}$.

Les gradients $\nabla f, \nabla g$ et les laplaciens $\Delta f, \Delta g$ sont définis par

$$\begin{aligned} \nabla f &= {}^T(\partial_x f, \partial_y f), & \nabla g &= {}^T(\partial_x g, \partial_y g, \partial_z g), \\ \Delta f &= \partial_{xx}^2 f + \partial_{yy}^2 f, & \Delta g &= \partial_{xx}^2 g + \partial_{yy}^2 g + \partial_{zz}^2 g. \end{aligned} \quad (1.11)$$

soit en général pour $f : \mathbb{R}^m \rightarrow \mathbb{R}$

$$\nabla f = {}^T(\partial_{x_1} f, \dots, \partial_{x_m} f), \quad \Delta f = \partial_{x_1 x_1}^2 f + \dots + \partial_{x_m x_m}^2 f.$$

Soit $V = (u, v) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ et $W = (u, v, w) : \mathbb{R}^3 \rightarrow \mathbb{R}^3$.

Les rotationnels $\text{rot } V, \text{rot } W$ et divergence $\text{div } V, \text{div } W$ sont définis suivant

$$\begin{aligned} \text{rot } V &= \partial_x v - \partial_y u, & \text{rot } W &= (\partial_y w - \partial_z v, \partial_z u - \partial_x w, \partial_x v - \partial_y u), \\ \text{div } V &= \partial_x u + \partial_y v, & \text{div } W &= \partial_x u + \partial_y v + \partial_z w. \end{aligned}$$

On a les relations classiques

$$\text{div}(\nabla f) = \Delta f, \quad \text{div}(\nabla g) = \Delta g, \quad \text{div rot } W = 0, \quad \text{rot } \nabla f = 0, \quad \text{rot } \nabla g = 0,$$

En effet, par exemple

$$\begin{aligned} \text{div}(\nabla f) &= \text{div}(\partial_x f, \partial_y f) = \partial_x(\partial_x f) + \partial_y(\partial_y f) = \Delta f, \\ \text{div rot } W &= \partial_x(\partial_y w - \partial_z v) + \partial_y(\partial_z u - \partial_x w) + \partial_z(\partial_x v - \partial_y u) = 0, \\ \text{rot } \nabla f &= \partial_x(\partial_y f) - \partial_y(\partial_x f) = 0. \end{aligned}$$

Une *équation aux dérivées partielles* est une équation dont l'inconnue est une fonction f de plusieurs variables à valeur scalaires ou vectorielles, fonction qui annule une expression contenant certaines dérivées partielles de f .

Ces opérateurs interviennent dans les équations aux dérivées partielles décrivant des phénomènes physiques basiques

$\partial_t u + \partial_x(a(x)u)$	t, x	transport (1d)
$\Delta u + \lambda u = 0$	x, y, z	Schrödinger stationnaire (3d)
$i\partial_t u = \Delta u + V$	t, x, y, z	Schrödinger avec potentiel (3d)
$\partial_t u = \Delta u$	t, x, y	chaleur (2d)
$\partial_{tt}^2 u = \Delta u$	t, x, y	ondes (2d)
$\rho(\partial_t u + (u \cdot \nabla)u) + \nabla p = F$	t, x, y	Euler (2d)

et les ÉDP en dimension 1 d'espace correspondantes

$\partial_{x^2}^2 u + \lambda u = 0$	x	Schrödinger stationnaire (1d)
$i\partial_t u = \partial_{x^2}^2 u + V$	t, x	Schrödinger avec potentiel (1d)
$\partial_t u = \partial_{x^2}^2 u$	t, x	chaleur (1d)
$\partial_{x^2}^2 u = \partial_{t^2}^2 u$	t, x	ondes (1d)

1.7 Annexe 1 : applications linéaires et matrices

Généralisant les applications linéaires $L_1 : v \in \mathbb{R} \mapsto \ell v \in \mathbb{R}$, on considère les applications linéaires ℓ de \mathbb{R}^m dans \mathbb{R}^n de la forme

$$\ell(\mathbf{v}) = \left(\sum_{i=1}^m \ell_{ji} v_i \right) = \left(\sum_{i=1}^m \ell_{1i} v_i, \dots, \sum_{i=1}^m \ell_{ni} v_i \right), \quad \mathbf{v} = (v_1, \dots, v_m), j = 1, \dots, n. \quad (1.12)$$

Le tableau de nombres associé

$$L = \begin{pmatrix} \ell_{11} & \dots & \ell_{1m} \\ \vdots & & \vdots \\ \ell_{n1} & \dots & \ell_{nm} \end{pmatrix}$$

sera appelé *matrice* : il a n lignes et m colonnes. La connaissance de la matrice L associé à ℓ permet de retrouver l'image $\ell(\mathbf{v})$ par une multiplication de la matrice L et du vecteur v

$$L \cdot \mathbf{v} = \begin{pmatrix} \ell_{11} & \dots & \ell_{1m} \\ \vdots & & \vdots \\ \ell_{n1} & \dots & \ell_{nm} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix} = \left(\sum_{i=1}^m \ell_{1i} v_i \quad \dots \quad \sum_{i=1}^m \ell_{ni} v_i \right) \in \mathbb{R}^n. \quad (1.13)$$

On a représenté le vecteur \mathbf{v} indifféremment par une liste de coordonnées (avec des virgules les séparant) ou bien comme une matrice à 1 colonne et m lignes (avec le blanc comme séparateur des coefficients de la matrice). Pour deux applications linéaires $\ell : \mathbb{R}^m \rightarrow \mathbb{R}^n$ et $k : \mathbb{R}^n \rightarrow \mathbb{R}^p$, leur composée $f = k \circ \ell : \mathbb{R}^m \rightarrow \mathbb{R}^p$ est une application linéaire de \mathbb{R}^m dans \mathbb{R}^p , avec comme tableau associé F le produit des tableaux K et L

$$F = K \cdot L = \begin{pmatrix} k_{11} & \dots & k_{1n} \\ \vdots & & \vdots \\ k_{p1} & \dots & k_{pn} \end{pmatrix} \begin{pmatrix} \ell_{11} & \dots & \ell_{1m} \\ \vdots & & \vdots \\ \ell_{n1} & \dots & \ell_{nm} \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n k_{1j} \ell_{j1} & \dots & \sum_{j=1}^n k_{1j} \ell_{jm} \\ \vdots & & \vdots \\ \sum_{j=1}^n k_{pj} \ell_{j1} & \dots & \sum_{j=1}^n k_{pj} \ell_{jm} \end{pmatrix}$$

et on vérifie bien

$$(K \cdot L) \cdot \mathbf{v} = k(\ell(\mathbf{v})).$$

suivant les identifications de l'application linéaire ℓ et du tableau L et de leur action sur les vecteurs $\mathbf{v} \in \mathbb{R}^m$ suivant (1.13). On identifie donc application linéaire $\ell : \mathbb{R}^m \rightarrow \mathbb{R}^n$ et matrice $L = (\ell_{ji})$ à $i = 1, \dots, m$ colonnes et $j = 1, \dots, n$ lignes.

Le vecteur \mathbf{v} dans les formules $\ell(\mathbf{v})$ de (1.12) et $L \cdot \mathbf{v}$ de (1.13) coïncident quasiment : rigoureusement le premier considère des vecteurs ligne de coordonnées pour les vecteurs \mathbf{v} et $\ell(\mathbf{v})$ alors que le second est un vecteur colonne où le produit $L \cdot \mathbf{v}$ est un produit matriciel d'une matrice de n lignes et m colonnes et d'un vecteur colonne de m lignes donnant un vecteur colonne de n lignes. Ces identifications seront systématiquement faites.

1.8 Annexe 2 : notations, composées et différentielles

Cette annexe commence par décrire les choix de notations (redondants, voire parfois incompatibles) de variables et fonctions utilisés dans ces notes. Cette liste est suivie de 4 formules de différentielles de fonctions composées entre des espaces de diverses (basses) dimensions : ils reprennent pour des (basses) dimensions la formule générale (1.7) du théorème 1.4 et complètent les exemples 1.7.

1. U : ouvert de \mathbb{R}^2 , \mathbb{R}^3 ou \mathbb{R}^m (et \mathbb{R} !)
2. \mathbf{m} , \mathbf{x} , \mathbf{v} , \mathbf{h} : points de \mathbb{R}^2 , \mathbb{R}^3 , \mathbb{R}^m . Lettres en gras dans l'édition papier, mais au tableau les lettres soulignées 1 ou 2 fois : $\underline{\underline{m}}$, $\underline{\underline{x}}$, $\underline{\underline{v}}$.
3. $\mathbf{x} = (x_1, x_2, x_3)$, $\mathbf{m} = (x, y, z)$, $\mathbf{v} = (v_1, v_2)$: points et leur coordonnées :
4. (h, k) , (h, k, ℓ) , (v_1, v_2) : points en coordonnées dans \mathbb{R}^2 , \mathbb{R}^3 ou $\mathbb{R}^m \dots$:
5. $f = (f_1, f_2, f_3)$: fonction à valeurs vectorielles, avec leurs fonctions coordonnées (ici $\text{Im } f \subset \mathbb{R}^3$)
6. $f(\mathbf{x})$, $g(\mathbf{m} + \mathbf{v})$, ... : valeurs de fonctions :
7. df : différentielle (comme en une variable f' est associée à f)

8. $df(\mathbf{m})$: valeur de la différentielle au point \mathbf{m} . C'est une application linéaire de \mathbb{R}^m dans \mathbb{R}^n si $f : U(\subset \mathbb{R}^m) \rightarrow \mathbb{R}^n$. En une variable (ou variable unidimensionnelle) $f'(\mathbf{m})$ est un nombre, associé de manière équivalente à l'application linéaire $v \in \mathbb{R} \rightarrow f'(\mathbf{m})v \in \mathbb{R}$
9. \mathbf{v} : vecteur de \mathbb{R}^m ; $df(\mathbf{m})(\mathbf{v})$: vecteur de \mathbb{R}^n
10. $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ $h = g \circ f : \mathbb{R}^m \rightarrow \mathbb{R}^p$
11. $f = (f_1, \dots, f_n) : \mathbb{R}^m \rightarrow \mathbb{R}^n$ avec $f_i : \mathbb{R}^m \rightarrow \mathbb{R}$, $i = 1, \dots, n$
12. $f(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_n(\mathbf{x}))$, $\mathbf{x} = (x_1, \dots, x_m) \in \mathbb{R}^m$
 $f(\mathbf{x}) = f_1(x_1, \dots, x_m), \dots, f_n(x_1, \dots, x_m)$
13. $g(\mathbf{y}) = (g_1(\mathbf{y}), \dots, g_p(\mathbf{y}))$, $\mathbf{y} = (y_1, \dots, y_n)$
14. $g(\mathbf{y}) = (g_1(\mathbf{y}), \dots, g_p(\mathbf{y})) = (g_1(y_1, \dots, y_n), \dots, g_p(y_1, \dots, y_n))$, $\mathbf{y} \in \mathbb{R}^n$
15. $g \circ f(\mathbf{x}) = g[f(\mathbf{x})] = g(f_1(\mathbf{x}), \dots, f_n(\mathbf{x}))$

Calcul des dérivées partielles de $g \circ f$

Exemple 0 : $m = 1, n = 1, p = 1$. Avec la variable t pour f et y pour g , on a

$$\frac{d(g \circ f)}{dt}(t) = \frac{dg}{dy}(f(t)) \frac{df}{dt}(t)$$

$$\partial_t(g \circ f)(t) = \partial_y g(f(t)) \partial_t f(t)$$

Exemple 1 : $m = 1, n, p = 1$. La composée

$$g \circ f : t \in \mathbb{R} \mapsto g(f_1(t), \dots, f_n(t)) \in \mathbb{R}$$

a pour dérivée

$$\begin{aligned} \frac{d}{dt}(g \circ f(t)) &= \frac{d}{dt}(g(f_1(t), \dots, f_n(t))) \\ &= \partial_{x_1} g(f_1(t), \dots, f_n(t)) \frac{df_1}{dt}(t) + \dots + \partial_{x_n} g(f_1(t), \dots, f_n(t)) \frac{df_n}{dt}(t) \\ &= \partial_{x_1} g(f_1(t), \dots, f_n(t)) f_1'(t) + \dots + \partial_{x_n} g(f_1(t), \dots, f_n(t)) f_n'(t) \end{aligned}$$

Si $n = 1$, $f(t) = f_1(t)$ et $g \circ f(t) = g(f(t))$

$$\frac{d}{dt}(g \circ f(t)) = \frac{d}{dt}(g(f(t))) = \partial_{x_1} g(f_1(t)) \frac{df_1}{dt}(t) = g'(f(t)) f'(t)$$

Exemple 2 : $m = 2, n = 2, p = 1$. La composée

$$f : (x, y) \in \mathbb{R}^2 \mapsto (u(x, y), v(x, y)) \mapsto g(u(x, y), v(x, y)) \in \mathbb{R}$$

a une différentielle caractérisée par deux coefficients

$$\begin{aligned} \partial_x [g(u(x, y), v(x, y))] &= \partial_u g(u(x, y), v(x, y)) \partial_x u(x, y) + \partial_v g(u(x, y), v(x, y)) \partial_x v(x, y) \\ \partial_y [g(u(x, y), v(x, y))] &= \partial_u g(u(x, y), v(x, y)) \partial_y u(x, y) + \partial_v g(u(x, y), v(x, y)) \partial_y v(x, y) \end{aligned}$$

1 Exemple 3 : $m = 3, n = 1, p = 2$. Soit les fonctions f et g

$$f : (x, y, z) \in \mathbb{R}^3 \mapsto f(x, y, z) \in \mathbb{R}, \quad g : u \in \mathbb{R} \mapsto (g_1(u), g_2(u)) \in \mathbb{R}^2$$

et donc la composée

$$(g \circ f)(x, y, z) = ((g_1 \circ f)(x, y, z), (g_2 \circ f)(x, y, z)) \in \mathbb{R}^2$$

Sa jacobienne (cf. (1.4)) a 2 lignes et 3 colonnes

$$\begin{pmatrix} g'_1(f(x, y, z))\partial_x f(x, y, z) & g'_1(f(x, y, z))\partial_y f(x, y, z) & g'_1(f(x, y, z))\partial_z f(x, y, z) \\ g'_2(f(x, y, z))\partial_x f(x, y, z) & g'_2(f(x, y, z))\partial_y f(x, y, z) & g'_2(f(x, y, z))\partial_z f(x, y, z) \end{pmatrix}.$$

Chapitre 2

Développements limités

Une fonction dérivable est approchée au voisinage d'un point a par une fonction affine :

$$f(a+h) \approx_{h \approx 0} f(a) + f'(a)h.$$

En effet, d'après la définition de la dérivée

$$\frac{f(a+h) - f(a)}{h} = f'(a) + \varepsilon(h)$$

avec $\varepsilon(h) \rightarrow 0$ lorsque $h \rightarrow 0$, ce qu'on peut écrire suivant

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

où l'approximation de $f(a+h)$ par $f(a) + f'(a)h$ a lieu avec un reste $h\varepsilon(h)$ petit par rapport à l'incrément h .

La notion de développement limité affine ce type d'approximation en introduisant des polynômes $a_0 + a_1h + \dots + a_nh^n$, dont les monômes a_0, a_1h, \dots, a_nh^n , ordonnés par degré croissant, constituent une échelle de quantités de plus en plus petites (quand le degré augmente) permettant par une combinaison linéaire d'approcher de plus en plus précisément une fonction au voisinage d'un point. La figure 2 illustre cette remarque pour le graphe de la fonction sin.

La complexité de calcul et des ressources (pour l'humain ou l'ordinateur) demeure abordable et peu onéreuse, le calcul sur des polynômes étant beaucoup plus facile que le calcul sur des fonctions trigonométriques ou des expressions exponentielles et autres fonctions transcendentes.

2.1 Approximations polynomiales

DÉFINITION 2.1: Soit I un intervalle de \mathbb{R} , f une fonction $f : I \rightarrow \mathbb{R}$ et $a \in I$. On dit que f admet un développement limité¹ d'ordre n au voisinage de a s'il existe un

1. Dans la littérature anglophone, on parle plutôt de *Taylor series expansion* ou de *series expansion*.

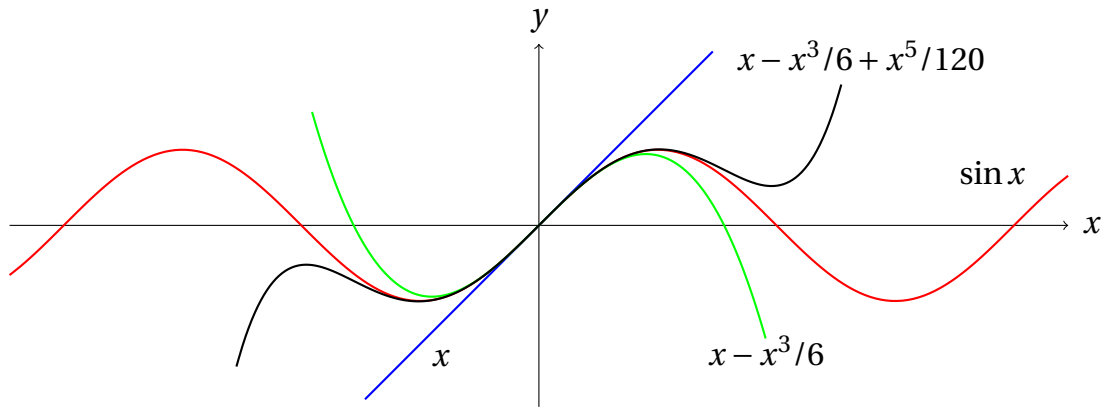


FIGURE II.1 – Le graphe de la fonction $x \in I \mapsto \sin x$ et de ses approximations de Taylor en $x = 0$ à l'ordre 1, 3 et 5 sur l'intervalle $I = [-4, 4]$.

polynôme $P_{f,n}(h) = f_0 + f_1 h + \dots + f_n h^n$ de degré au plus n tel que

$$f(a+h) = f_0 + f_1 h + \dots + f_n h^n + h^n \varepsilon_n(h) = P_{f,n}(h) + h^n \varepsilon_n(h)$$

où $\varepsilon_n(h) \rightarrow 0$ lorsque $h \rightarrow 0$.

Le polynôme $P_{f,n}(h)$ est dit polynôme de Maclaurin. Le terme final, la différence $r_n(h) = f(a+h) - P_{f,n}(h)$, est nommé reste de Maclaurin.

Le polynôme $P_{f,a,n}(x) = P_{f,n}(x-a)$ est dit polynôme de Taylor. Il permet d'écrire

$$f(x) = P_{f,a,n}(x-a) + r_n(x-a) = f_0 + f_1(x-a) + \dots + f_n(x-a)^n + r_n(x-a)$$

où le reste $r_n(x-a)$ vérifie $\lim_{x \rightarrow a} r_n(x-a)/(x-a)^n = 0$

△ REMARQUE 2.1: Les fonctions notées ci-dessus $\varepsilon_n(h)$ et $\tilde{\varepsilon}_n(x)$ partagent la propriété de converger vers 0 lorsque la variable h ou x converge vers la valeur de la variable adéquate (0 ou a) : c'est ceci qu'on retiendra, en notant simplement $\varepsilon(h)$ (ou $\varepsilon(x)$) toute quantité tendant vers 0 lorsque $h \rightarrow 0$, ou ce qui est équivalent $x \rightarrow a$. L'addition (ou la multiplication) de deux telles fonctions, la multiplication par une fonction bornée, donnent des fonctions ayant cette même propriété de limite : on notera toutes ces fonctions de la même manière, en écrivant (fort abusivement) $\varepsilon_1(x) + \varepsilon_2(x) = \varepsilon(x)$ par exemple. ▽

▷ EXEMPLES 2.1:

2.1.1 La relation

$$\frac{1}{1-h} = 1 + h + h^2 + \dots + h^n + \frac{h^{n+1}}{1-h}$$

où l'estimation sur le dernier terme $h^n \times \varepsilon(h)$ avec $\varepsilon(h) = h/(1-h)$

$$|\varepsilon(h)| = \left| \frac{h}{1-h} \right| \leq 2|h|, \quad |h| \leq 1/2,$$

donne le développement limité en $x = 0$ et à l'ordre n du membre de gauche

$$\boxed{\frac{1}{1-h} = 1 + h + h^2 + \dots + h^n + h^n \varepsilon(h)}$$

2.1.2 On a facilement le développement limité de $1/(1-x^2)$ en $x = 0$ en reprenant les formules précédentes

$$\frac{1}{1-h^2} = 1 + h^2 + h^4 + \dots + h^{2n} + \frac{h^{2(n+1)}}{1-h^2}$$

avec l'estimation sur le dernier terme

$$\left| \frac{h^{2(n+1)}}{1-h^2} \right| \leq |h|^{2n+1} \frac{4}{3} |h|, \quad |h| \leq 1/2$$

ce qui donne un développement à l'ordre $2n+1$

$$\frac{1}{1-h^2} = 1 + h^2 + h^4 + \dots + h^{2n} + h^{2n+1} \frac{h}{1-h^2}$$

avec $|h/(1-h^2)| \leq 4/3|h|$ si $|h| \leq 1/2$. Si on souhaite un développement limité à l'ordre n , il suffit de supprimer les termes d'ordre au moins $n+1$, soit

$$\frac{1}{1-h^2} = \begin{cases} 1 + h^2 + h^4 + \dots + h^n + h^{n+1} \varepsilon(h) & \text{si } n \text{ est pair,} \\ 1 + h^2 + h^4 + \dots + h^{n-1} + h^n \varepsilon(h) & \text{si } n \text{ est impair.} \end{cases}$$

Si n est impair, le terme $a_n h^n$ est nul : il n'apparaît pas dans le développement.

2.1.3 Le développement de $1/(1-2h)$ en $h = 0$ est pareillement aisément atteignable

$$\begin{aligned} \frac{1}{1-2h} &= 1 + 2h + (2h)^2 + \dots + (2h)^n + \frac{(2h)^{n+1}}{1-2h} \\ &= 1 + 2h + 2^2 h^2 + \dots + 2^n h^n + h^n \left[\frac{2^{n+1} h}{1-2h} \right] \end{aligned}$$

ainsi que le développement de $1/(1-x)$ en $x = 1/2$ en posant $h = x - 1/2$ qui tend vers 0 si et seulement si x tend vers $1/2$: ce changement de variable permet de se ramener au développement précédent

$$\frac{1}{1-x} = \frac{1}{1/2-h} = \frac{2}{1-2h}$$

et de conclure

$$\frac{1}{1-x} = 2 + 4 \left(x - \frac{1}{2} \right) + \dots + 2^{n+1} \left(x - \frac{1}{2} \right)^n + \left(x - \frac{1}{2} \right)^n \varepsilon(x)$$

où $\varepsilon(x) \rightarrow 0$ lorsque $x \rightarrow 0$.

◁

2.2 Formules de Taylor

Le théorème des accroissements finis

$$f(b) = f(a) + f'(c)(b - a),$$

est établi pour des fonctions dérivables. La formule de Taylor-Lagrange étend cette formule en exprimant, sous couvert d'une hypothèse de n -dérivabilité, la fonction f comme la somme d'un terme polynomial en $b - a$ de degré n et d'un reste bien contrôlé.

THÉORÈME 2.1 (Formule de Taylor-Lagrange²): *Soit I intervalle ouvert, $f : I \rightarrow \mathbb{R}$ de classe \mathcal{C}^n et $[a, b] \subset I$ ($a < b$). Alors, il existe $c \in]a, b[$ tel que*

$$\begin{aligned} f(b) = & f(a) + f'(a)(b - a) + f''(a) \frac{(b - a)^2}{2!} + \dots + f^{(k)}(a) \frac{(b - a)^k}{k!} + \dots \\ & + f^{(n-1)}(a) \frac{(b - a)^{n-1}}{(n - 1)!} + f^{(n)}(c) \frac{(b - a)^n}{n!}. \end{aligned}$$

DÉMONSTRATION. Pour $n = 1$, on retrouve le théorème des accroissements finis, *i. e.* l'existence d'un c tel que $(f(b) - f(a))/(b - a) = f'(c)$. Le réel c est un point du graphe de f où la pente de la tangente au graphe de f est la même que celle du segment joignant $(a, f(a))$ et $(b, f(b))$.

Pour $n > 1$ et une constante C , on introduit la fonction

$$F_{n,C}(x) = f(b) - f(x) - f'(x)(b - x) - f''(x) \frac{(b - x)^2}{2!} - \dots - f^{(n-1)}(x) \frac{(b - x)^{n-1}}{(n - 1)!} - C \frac{(b - x)^n}{n!}$$

qui est de classe \mathcal{C}^1 , vu que f est de classe \mathcal{C}^n , et où la constante C est choisie telle que $F_{n,C}(a) = 0$, *i. e.*

$$C \frac{(b - a)^n}{n!} = f(b) - f(a) - f'(a)(b - a) - f''(a) \frac{(b - a)^2}{2!} - \dots - f^{(n-1)}(a) \frac{(b - a)^{n-1}}{(n - 1)!}.$$

Vu la nullité $F_{n,C}(a) = F_{n,C}(b) = 0$ et la dérivabilité de $F_{n,C}$, le théorème de Rolle assure de l'existence de $c \in]a, b[$ annulant la dérivée $F'_{n,C}$. Or

$$\begin{aligned} F'_{n,C}(x) = & -f'(x) - [f'(x)(-1) + f''(x)(b - x)] - \left[f''(x)(-1) \frac{(b - x)}{1!} + f'''(x) \frac{(b - x)^2}{2!} \right] - \dots \\ & - \left[f^{(n-1)}(x)(-1) \frac{(b - x)^{n-2}}{(n - 2)!} + f^{(n)}(x) \frac{(b - x)^{n-1}}{(n - 1)!} \right] - C(-1) \frac{(b - x)^{n-1}}{(n - 1)!} \\ = & \left[C - f^{(n)}(x) \right] \frac{(b - x)^{n-1}}{(n - 1)!}, \end{aligned}$$

soit $C = f^{(n)}(c)$, cette égalité résultant directement de la nullité $F'_{n,C}(c) = 0$. \square

Δ REMARQUES 2.2:

2. B. Taylor, 1685–1731; J. L. Lagrange, 1736–1813.

1. En appliquant la formule précédente à la fonction $\varphi : x \rightarrow f(a+b-x)$ pour laquelle $\varphi(a) = f(b)$, $\varphi(b) = f(a)$ et $\varphi^{(k)}(a) = (-1)^k f^{(k)}(b)$, on obtient l'existence de $c \in (a, b)$ tel que

$$f(a) = f(b) + f'(b)(a-b) + \dots + f^{(k)}(b) \frac{(a-b)^k}{k!} + \dots \\ + f^{(n-1)}(b) \frac{(a-b)^{n-1}}{(n-1)!} + f^{(n)}(c) \frac{(a-b)^n}{n!}.$$

Ainsi, pour h réel (négatif ou positif) tel que l'intervalle d'extrémités a et $a+h$ soit dans I , on a

$$f(a+h) = f(a) + f'(a)h + f''(a) \frac{h^2}{2!} + \dots + f^{(k)}(a) \frac{h^k}{k!} + \dots \\ + f^{(n-1)}(a) \frac{h^{n-1}}{(n-1)!} + f^{(n)}(a+\theta h) \frac{h^n}{n!}$$

avec $\theta \in (0, 1)$.

2. Si $M_{|f|}$ note un majorant de la fonction $|f|$ sur un voisinage de x , le reste $f^{(n)}(c)(b-a)^n/(n!)$ dans la formule de Lagrange est majoré par $\frac{(b-a)^n}{n!} M_{|f^{(n)}|}$, assurant un développement limité à l'ordre $n-1$ au point a . ∇

▷ EXEMPLE 2.2: Les dérivées successives des fonctions sin, cos sont majorées uniformément par 1. Cela permet de majorer simplement les restes dans les formules de Taylor par $h^n/n!$

Ainsi, par ex., pour $\sin(0.1) = 0.099\ 833\ 416\dots$, la formule de Taylor-Lagrange à l'ordre $n=3$

$$\sin(0.1) = \sin(0) + \cos(0)0.1 - \frac{\sin(0)}{2}(0.1)^2 + \frac{-\cos(0)}{6}(0.1)^3 + R_4 \\ = 0.1 - (0.1)^3/6 + R_4 \\ = 0.099\ 833\ 33\dots + R_4$$

et à l'ordre 4

$$\sin(0.1) = \sin(0) + \cos(0)0.1 - \frac{\sin(0)}{2}(0.1)^2 + \frac{-\cos(0)}{6}(0.1)^3 + \frac{\sin(0)}{24}(0.1)^4 + R_5 \\ = 0.1 - (0.1)^3/6 + R_5 \\ = 0.099\ 833\ 33\dots + R_5$$

avec des restes majorés suivant

$$|R_4| \leq (0.1)^4/4! \sim 0.000\ 004\ 166 \sim 4.2 * 10^{-6} \\ |R_5| \leq (0.1)^5/5! \sim 0.000\ 000\ 083\ 166 \sim 8.3 * 10^{-8}$$

les valeur approchées $P_3 = P_4 = 0.099\ 833\ 33$ diffèrent de la valeur exacte de $|\sin 0.1 - P_3| = 0.000\ 000\ 083\ 313\dots < 8.34 * 10^{-8}$. Ce développement du sin est très particulier, car cette fonction est impaire, ce qui annule tous les termes d'ordre pair. \triangleleft

THÉORÈME 2.2 (Formule de Taylor-Young³): Soit I intervalle ouvert, $f : I \rightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^n et $a \in I$. Alors, la formule de Taylor-Young est

$$f(x) = f(a) + f'(a)(x-a) + f''(a)\frac{(x-a)^2}{2!} + \dots + f^{(k)}(a)\frac{(x-a)^k}{k!} + \dots \\ + f^{(n)}(a)\frac{(x-a)^n}{n!} + (x-a)^n \varepsilon(x).$$

où $\varepsilon(x) \rightarrow 0$ lorsque $x \rightarrow a$.

Le cas particulier de formule de Taylor en $x = 0$ est appelé *Formule de Maclaurin*⁴. Pour considérer de telles formules au voisinage de $h = 0$, il suffit de faire le changement de variable $h = x - a$, c'est à dire de considérer la fonction $F : h \mapsto f(a+h)$ pour laquelle la formule de Taylor-Young prend la forme

$$F(h) = F(0) + F'(0)h + F''(0)\frac{h^2}{2!} + \dots + F^{(k)}(0)\frac{h^k}{k!} + \dots + F^{(n)}(0)\frac{h^n}{n!} + h^n \varepsilon(h)$$

soit

$$f(a+h) = f(a) + f'(a)h + f''(a)\frac{h^2}{2!} + \dots + f^{(k)}(a)\frac{h^k}{k!} + \dots + f^{(n)}(a)\frac{h^n}{n!} + h^n \varepsilon(h)$$

DÉMONSTRATION. Si $n = 1$, la formule de Taylor-Young se réduit à l'expression

$$f(x) = f(a) + f'(a)(x-a) + (x-a)\varepsilon(x),$$

qui dérive de la formule des accroissements finis où $|c_x - a| \leq |x - a|$

$$f(x) = f(a) + f'(c_x)(x-a) = f(a) + f'(a)(x-a) + [f'(c_x) - f'(a)](x-a) \\ = f(a) + f'(a)(x-a) + (x-a)\varepsilon(x)$$

vu la continuité supposée de la dérivée f' . Pour $n \geq 2$, on reprend la formule de Taylor-Lagrange : il existe $c = c_x$ tel que

$$f(x) = f(a) + f'(a)(x-a) + f''(a)\frac{(x-a)^2}{2!} + f^{(k)}(a)\frac{(x-a)^k}{k!} + \dots \\ + f^{(n-1)}(a)\frac{(x-a)^{n-1}}{(n-1)!} + f^{(n)}(c_x)\frac{(x-a)^n}{n!} \\ = f(a) + f'(a)(x-a) + f''(a)\frac{(x-a)^2}{2!} + f^{(k)}(a)\frac{(x-a)^k}{k!} + \dots \\ + f^{(n-1)}(a)\frac{(x-a)^{n-1}}{(n-1)!} + f^{(n)}(a)\frac{(x-a)^n}{n!} + [f^{(n)}(c_x) - f^{(n)}(a)]\frac{(x-a)^n}{n!}$$

où le dernier terme est de la forme $(x-a)^n \varepsilon(x)$ vu la continuité de $f^{(n)}$ en a , ce qui achève la démonstration. \square

3. W. H. Young, 1863–1942.

4. C. Maclaurin, 1698–1746.

En résumé, on retiendra la formule de Taylor avec reste

THÉORÈME 2.3: *Soit $f : I \rightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^∞ au voisinage de a . On a le développement*

$$f(a+h) = f(a) + f'(a)h + \dots + \frac{f^{(k)}(a)h^k}{k!} + \dots + \frac{f^{(n)}(a)h^n}{n!} + h^n \varepsilon(h). \quad (2.1)$$

Il est parfois utile (cf. les calculs sur la fonction sin de l'exemple 2.2 d'avoir une expression plus précise du reste $R_n(h) = h^n \varepsilon(h)$ sous la forme

$$R_n(h) = \frac{f^{(n+1)}(a+\theta h)h^{n+1}}{(n+1)!}.$$

Mais la forme du théorème suffit pour le calcul des développements limités.

2.3 Premières propriétés

La formule de Taylor-Young donne des développements limités pour les fonctions de base. On introduit les espaces $\mathcal{C}^k(I)$ des fonctions de classe \mathcal{C}^k sur un intervalle I .

DÉFINITION 2.2: *Soit I un intervalle. L'espace $\mathcal{C}^0(I)$ est l'espace des fonctions continues sur I . Pour $k > 0$, on définit par récursivité l'espace $\mathcal{C}^k(I)$ des fonctions dérivables dont la dérivée est dans l'espace $\mathcal{C}^{k-1}(I)$ des fonctions continues dérivables⁵.*

La formule de Taylor-Young donne l'existence d'un développement limité à l'ordre n pour toute fonction de classe \mathcal{C}^n , comme cela est dit dans le théorème suivant.

THÉORÈME 2.4: *Soit I intervalle de \mathbb{R} et $a \in I$. Toute fonction $f \in \mathcal{C}^n(I)$ admet un développement limité d'ordre n en $x = a$ avec*

$$f(x) = \sum_{\ell=0}^n f^{(\ell)}(a) \frac{(x-a)^\ell}{\ell!} + (x-a)^n \varepsilon(x),$$

soit sous la forme de MacLaurin, pour h tendant vers 0,

$$f(a+h) = \sum_{\ell=0}^n f^{(\ell)}(a) \frac{h^\ell}{\ell!} + h^n \varepsilon(h).$$

5. Si I contient l'extrémité a , par exemple la plus petite, la fonction est supposée k -fois dérivable à droite en a .

▷ EXEMPLES 2.3:

2.3.1 Vu que $\cos^{(2k)} = (-1)^k \cos$ et $\cos^{(2k+1)} = (-1)^{k+1} \sin$, on a les développements limités en $x = 0$ de la fonction \cos aux ordres 0, 1, 2, 3 et 4

$$\begin{aligned} \cos x &= 1 + \varepsilon(x) \\ &= 1 + x\varepsilon(x) \\ &= 1 - \frac{x^2}{2} + x^2\varepsilon(x) \\ &= 1 - \frac{x^2}{2} + x^3\varepsilon(x) \\ &= 1 - \frac{x^2}{2} + \frac{x^4}{24} + x^4\varepsilon(x) + \dots \end{aligned}$$

Le développement limité de \cos en $x = 0$ ne contient que des termes d'ordre pair : si on a un développement limité à l'ordre $2k$, on en a un *de facto* à l'ordre $2k + 1$: cela explique les stationarités apparentes des lignes 1-2 et 3-4.

2.3.2 La fonction qui vaut 0 sur les rationnels et x^4 pour x non rationnel admet un développement limité (nul) à l'ordre 3 en 0 : $f(x) = x^3\varepsilon(x)$. Néanmoins, cette fonction n'est, en dehors de $x = 0$, nulle part continue! Il y a des développements limités qui ne proviennent pas de l'application de la formule Taylor-Young. ◀

△ REMARQUES 2.3:

1. Si un développement limité à l'ordre 1 en a est équivalent à la différentiabilité de la fonction, cette propriété ne vaut plus pour les dérivées d'ordre supérieur et un développement limité d'ordre au moins 1.
2. Le facteur $\varepsilon(h)$ est important uniquement pour indiquer un terme tendant vers 0, sa valeur exacte important peu. On le notera *génériquement* ε .
3. Un développement limité donne une approximation d'une fonction par un polynôme de degré n au voisinage d'un point a : les monômes de ce polynôme de degré croissant, donnant des termes correctifs d'ordre inférieur, *i. e.* de plus en plus petits, $h^{n+1} = h^n\varepsilon(h)$: si $h = 0.1$, on a

$$h^2 = 0.01 \geq h^3 = 0.001 \geq h^4 = 0.0001 \geq \dots$$

Un polynôme de développement limité sera écrit comme une somme de monômes de degré croissant.

4. Le tableau II.1 rassemble les développements limités de fonctions de base tous justifiés par le théorème 2.4 précédent, à l'exception du premier qui a été établi directement. Si α est un entier naturel, le polynôme P_n^0 du développement limité d'ordre n du polynôme $(1+x)^\alpha$ est constant de degré α si

$n > \alpha$, avec reste nul : ce développement provient directement de la formule du binôme

$$(1+x)^\alpha = \sum_{k=0}^{\alpha} \binom{\alpha}{k} x^k = \sum_{k=0}^{\alpha} \frac{\alpha(\alpha-1)\dots(\alpha-k+1)}{k!} x^k.$$

avec

$$\binom{\alpha}{k} = \frac{\alpha!}{k!(\alpha-k)!} = \frac{[\alpha(\alpha-1)\dots(\alpha-k+1)]}{k!}.$$

5. Existe-t-il des développements limités issus de ce tableau qui donnent un développement limité en tout point du domaine de définition? La réponse est oui, grâce à des « formules d'addition »

$$\begin{aligned} e^{x_0+h} &= e^{x_0} e^h = e^{x_0} \left(1 + h + \frac{h^2}{2!} + \dots \right) \\ &= 1 + e^{x_0} h + \frac{e^{x_0}}{2!} h^2 + \dots \\ \cos(x_0+h) &= \cos x_0 \cos h - \sin x_0 \sin h \\ &= \cos x_0 (1 - h^2/2 + \dots) - \sin x_0 (h - h^3/6 + \dots) \\ (1+x_0+h)^\alpha &= (1+x_0)^\alpha \left(1 + \frac{h}{1+x_0} \right)^\alpha = \dots \end{aligned}$$

On voit ainsi à l'œuvre pour l'obtention de développement limité en un point quelconque des calculs sur les polynômes, opérations systématisées dans la suite. ∇

PROPOSITION 2.1: *Le polynôme $P_{f,a,n}$ de développement limité à l'ordre n en a est uniquement déterminé parmi les polynômes de degré n .*

Si f admet un développement limité à l'ordre p en a , alors, pour tout $n < p$, f admet un développement limité en a d'ordre n , obtenu par l'opérateur de troncature T_n défini suivant

$$T_n \left[\sum_{j=0}^p f_j h^j \right] = \sum_{j=0}^n f_j h^j.$$

DÉMONSTRATION. Il y a unicité d'un développement limité à l'ordre n . En effet, si deux polynômes $P_{f,n}, Q_{f,n}$ de degré au plus n donnent un développement limité de f , alors $[P_{f,n} - Q_{f,n}](h) = h^n \varepsilon(h)$ n'est possible que si le polynôme $R = P_{f,n} - Q_{f,n}$ est nul : si $R(h) = r_k h^k + \dots + r_n h^n$ avec $r_k \neq 0$, alors

$$\frac{R(h)}{h^n} = \frac{h^k [r_k + r_{k+1} h + \dots + r_n h^{n-k}]}{h^n} = h^{k-n} [r_k + r_{k+1} h + \dots + r_n h^{n-k}]$$

Lorsque $h \rightarrow 0$, le terme $\varepsilon(h) = R(h)/h^n$ tend vers 0 par hypothèse, alors que la valeur absolue du dernier membre des égalités ci-dessus tend vers $|r_k| > 0$ si $k = n$

TABLE II.1 – Développements limités de base en $x_0 = 0$ (I)

$f(x_0 + h)$	$P_{n,f}(h) + h^n \varepsilon(h)$
$\frac{1}{1-h}$	$1 + h + h^2 + \dots + h^n + h^n \varepsilon(h)$
$(1+h)^\alpha$	$1 + \alpha h + \frac{\alpha(\alpha-1)}{2!} h^2 + \dots + \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!} h^n + h^n \varepsilon(h)$
e^h	$1 + h + \frac{h^2}{2} + \dots + \frac{h^n}{n!} + h^n \varepsilon(h)$
$\sin h$	$h - \frac{h^3}{3!} + \frac{h^5}{5!} - \frac{h^7}{7!} + \dots + \frac{(-1)^n h^{2n+1}}{(2n+1)!} + h^{2n+2} \varepsilon(h)$
$\cos h$	$1 - \frac{h^2}{2!} + \frac{h^4}{4!} - \frac{h^6}{6!} + \dots + \frac{(-1)^n h^{2n}}{(2n)!} + h^{2n+1} \varepsilon(h)$

ou $+\infty$ si $k < n$, ce qui n'est pas possible. L'hypothèse $R \neq 0$ est donc absurde : $R(h) = h^n \varepsilon(h)$ n'est pas possible pour un polynôme non nul de degré au plus n et il y a unicité du polynôme $P_{f,n}$.

Si $p > n$, $h^p = h^n \varepsilon(h)$. Ainsi le polynôme de degré au plus p est de la forme $P(h) = T_n(P) + h^n \varepsilon(h)$ où $T_n(P)$ désigne le polynôme P où les monômes de degré au moins $n+1$ ont été omis, *i. e.* le polynôme P a été tronqué à l'ordre n : si f a un développement limité d'ordre $p > n$ en 0,

$$f(h) = P_{f,p}(h) + h^p \varepsilon(h) = T_n[P_{f,p}](h) + \varepsilon(h)h^n + h^p \varepsilon(h) = T_n[P_{f,p}](h) + h^n \varepsilon(h),$$

alors f admet un développement limité à l'ordre n , avec $T_n(P_{f,p})$ comme polynôme de Maclaurin à l'ordre n . \square

COROLLAIRE 2.1: *Le développement limité en $h = 0$ d'une fonction paire (impair resp.) n'a que des termes de degré pair (impair resp.).*

DÉMONSTRATION. Si f est paire, on a

$$P_{f,n,0}(h) + h^n \varepsilon(h) = f(h) = f(-h) = P_{f,n,0}(-h) + (-h)^n \varepsilon(-h) = P_{f,n,0}(-h) + h^n \varepsilon(h),$$

d'où $P_{f,n,0}(h) = P_{f,n,0}(-h)$: le polynôme $P_{f,n,0}$ est pair, il n'admet que des monômes de degré pair. \square

▷ EXEMPLE 2.4: Les fonctions cosh et sinh ont des développements limités à tout ordre

$$\begin{aligned}\cosh x &= 1 + x^2/2 + x^2\varepsilon(x) = 1 + \sum_{k=1}^n \frac{x^{2k}}{(2k)!} + x^{2n+1}\varepsilon(x), \\ \sinh x &= x + x^3/6 + x^4\varepsilon(x) = \sum_{k=0}^n \frac{x^{2k+1}}{(2k+1)!} + x^{2n+2}\varepsilon(x).\end{aligned}$$

Le terme d'ordre $2k+1$ ($2k$ resp.) pour la fonction cosh (sinh resp.) est nul, comme il résulte des calculs de dérivée : $\cosh^{(2k+1)} x = \sinh x$ et $\cosh^{(2k)} x = \cosh x$. ◁

2.4 Opérations sur les développements limités

En introduisant le changement de variable $x = a + h$, on peut se limiter à ne considérer que des développements en 0 et des polynômes de Maclaurin : les polynômes de développement limité de $x \mapsto f(x)$ en a et de $h \mapsto F(h) = f(a + h)$ en 0 se correspondent via $P_{f,n,a}(x) = P_{f,n}(x - a)$

$$f(x) = P_{f,n,a}(x) + (x - a)^n \varepsilon(x), \quad f(a + h) = P_{f,n}(h) + h^n \varepsilon(h).$$

On considère diverses fonctions du type de f admettant un développement limité en 0 à l'ordre n : $f(h) = P_{f,n}(h) + R_{f,n}(h)$ où $P_{f,n}$ est un polynôme de Maclaurin de degré n et le reste $R_{f,n} = h^n \varepsilon(h)$.

PROPOSITION 2.2 (addition et produit): Soient f et g admettant des développements limités d'ordre n en a .

Alors la somme $f + g$ admet un développement limité d'ordre n en a avec comme polynôme de Maclaurin $P_{f+g,n} = P_{f,n} + P_{g,n}$.

Le produit fg admet un développement limité d'ordre n en a avec comme polynôme de Maclaurin le polynôme de degré n obtenu par troncature T_n du produit des polynômes de Maclaurin d'ordre n de f et g : $P_{fg,n} = T_n[P_{f,n}P_{g,n}]$.

DÉMONSTRATION. On a pour la somme

$$(f + g)(a + h) = P_{f,n}(h) + h^n \varepsilon(h) + P_{g,n}(h) + h^n \varepsilon(h) = P_{f,n}(h) + P_{g,n}(h) + h^n \varepsilon(h)$$

alors que pour le produit

$$\begin{aligned}(fg)(a + h) &= (P_{f,n}(h) + h^n \varepsilon(h))(P_{g,n}(h) + h^n \varepsilon(h)) \\ &= P_{f,n}(h)P_{g,n}(h) + h^n \varepsilon(h) \\ &= T_n[P_{f,n}P_{g,n}](h) + h^n \varepsilon(h) + h^n \varepsilon(h) \\ &= T_n[P_{f,n}P_{g,n}](h) + h^n \varepsilon(h)\end{aligned}$$

où on a utilisé le fait que les termes $P_{f,n}(h)h^n\varepsilon(h)$, $P_{g,n}(h)h^n\varepsilon(h)$ et $h^n\varepsilon(h)h^n\varepsilon(h)$ sont tous des $h^n\varepsilon(h)$. On a donc démontré

$$P_{f+g,n} = P_{f,n} + P_{g,n}, \quad P_{fg,n} = T_n[P_{f,n}P_{g,n}].$$

□

▷ EXEMPLE 2.5: Utilisant les développements limités de l'exponentielle et de $(1+x)^\alpha_{\alpha=-2}$ dans le tableau II.1, on obtient le développement limité de $e^x(1+x)^{-2}$ en $x=0$ et à l'ordre 2

$$\begin{aligned} e^x(1+x)^{-2} &= (1+x+x^2/2+x^2\varepsilon(x))(1-2x+3x^2+x^2\varepsilon(x)) \\ &= 1+(1-2)x+(1/2-2+3)x^2+x^2\varepsilon(x) \\ &= 1-x+3/2x^2+x^2\varepsilon(x), \end{aligned}$$

et à l'ordre 3

$$\begin{aligned} e^x(1+x)^{-2} &= (1+x+x^2/2+x^3/3+x^3\varepsilon(x))(1-2x+3x^2-4x^3+x^3\varepsilon(x)) \\ &= 1+(1-2)x+(3-2+1/2)x^2+(-4+3-1+1/3)x^3+x^3\varepsilon(x) \\ &= 1-x+3/2x^2-5/3x^3+x^3\varepsilon(x). \end{aligned}$$

Pour avoir le développement en $x_0=2$ à l'ordre 2, on se ramène à des expressions avec la variable $h \sim 0$ telle que $x=2+h \sim 2$

$$e^x(1+x)^{-2} = e^{2+h}(1+2+h)^{-2} = e^2e^h(3+h)^{-2} = e^2/9e^h(1+h/3)^{-2}$$

et on calcule le développement limité du second facteur en revenant à la définition (anticipant un peu le calcul de développement limité de composées)

$$\begin{aligned} (1+h/3)^{-2} &= 1+(-2)h/3+\frac{(-2)(-3)}{2}(h/3)^2+h/3)^2\varepsilon(h/3) \\ &= 1-2h/3+h^2/3+h^2\varepsilon(h) \end{aligned}$$

et donc

$$\begin{aligned} e^x(1+x)^{-2}_{|x=2+h} &= e^2/9(1+h+h^2/2+h^2\varepsilon(h))(1-2h/3+h^2/3+h^2\varepsilon(h)) \\ &= e^2/9\left(1+\frac{h}{3}+\left(\frac{1}{2}+\frac{1}{3}-\frac{2}{3}\right)h^2+h^2\varepsilon(h)\right) \\ &= e^2/9(1+h/3+h^2/6)+h^2\varepsilon(h) \\ &= e^2/9\left(1+\frac{x-2}{3}+\frac{(x-2)^2}{6}\right)+(x-2)^2\varepsilon(x). \end{aligned}$$

◁

PROPOSITION 2.3 (composition): Soit f (g resp.) avec un développement limité d'ordre n en a ($b = f(a)$ resp.). Alors $g \circ f$ admet un développement limité d'ordre n (au moins égal à 1) en a avec

$$P_{g \circ f, a, n} = T_n \left[P_{g, f(a), n} \left(P_{f, a, n} - f(a) \right) \right].$$

DÉMONSTRATION. Vu que $f(a) = b$ et f dérivable, on a $f(a+h) = b + hF(h)$ pour une fonction F bornée⁶ au voisinage de 0 et donc

$$(f(a+h) - f(a))^n = h^n F_n(h),$$

avec F_n bornée au voisinage de 0. Par ailleurs, on a pour tout entier j et toute constante A

$$(A + h^n \varepsilon(h))^j = A^j + h^n \varepsilon(h)$$

et donc pour tout polynôme P on a

$$P(A + h^n \varepsilon(h)) = P(A) + h^n \varepsilon(h).$$

Ainsi

$$\begin{aligned} g \circ f(a+h) &= g[f(a+h)] \\ &= g[b + (f(a+h) - b)] \\ &= P_{g, b, n} [f(a+h) - b] + (f(a+h) - b)^n \varepsilon(h) \\ &= P_{g, b, n} [P_{f, a, n}(h) - b + h^n \varepsilon(h)] + h^n F_n(h) \varepsilon(h) \\ &= P_{g, b, n} [P_{f, a, n}(h) - b] + h^n \varepsilon(h) + h^n \varepsilon(h) \\ &= T_n [P_{g, b, n} \circ [P_{f, a, n} - b]](h) + h^n \varepsilon(h) \end{aligned}$$

On a utilisé que si p_0 est une fonction constante, alors $p_0(f(h)) = p_0$, alors que le monôme h^k de degré au moins 1 composé avec une fonction du type $h^n \varepsilon(h)$ est du type $h^n \varepsilon(h)$. □

△ REMARQUE 2.4: Si sa démonstration est un peu délicate, l'énoncé précédent indique la marche à suivre très clairement : pour un développement limité à l'ordre n de la composée de deux fonctions ayant un développement limité au même ordre, il suffit de considérer la composée des deux polynômes de Maclaurin, puis de tronquer ce polynôme à l'ordre n . ▽

▷ EXEMPLES 2.6:

6. Pour f de classe \mathcal{C}^1 , son développement limité permet justement d'écrire $f(a+h) = f(a) + f'(a)h + h\varepsilon(h) = f(a) + h(f'(a) + \varepsilon(h))$. On peut aussi utiliser l'intégrale (à supposer que ses résultats soient disponibles!) $f(a+h) = f(a) + \int_a^{a+h} f'(k)dk = f(a) + h \int_0^1 f'(a+ht)dt$.

2.6.1 Soit $\alpha \neq 0$ et g avec le développement limité $g(h) = P_g(h) + h^n \varepsilon(h)$. Alors

$$g(\alpha h) = P_g(\alpha h) + (\alpha h)^n \varepsilon(h) = P_g(\alpha h) + h^n \varepsilon(h)$$

On a de manière analogue

$$g(h^2) = P_g(h^2) + (h^2)^n \varepsilon(h^2) = P_g(h^2) + h^{2n} \varepsilon(h).$$

Cet exemple est particulier, vu que d'un développement limité à l'ordre n pour g on trouve un développement limité à l'ordre $2n$ pour la composée $x \mapsto g(x^2)$.

2.6.2 À l'ordre $n = 4$, avec $f(h) = h + h^2$ et $g(u) = (1 + u)^{-1}$, on obtient

$$\begin{aligned} (1 + h + h^2)^{-1} &= (1 - u)^{-1} \Big|_{u=-h-h^2} \\ &= (1 + u + u^2 + u^3 + u^4 + u^4 \varepsilon(u)) \Big|_{u=-h-h^2} \\ &= 1 - (h + h^2) + (h + h^2)^2 - (h + h^2)^3 + (h + h^2)^4 + h^4 \varepsilon(u) \\ &= 1 - h - h^2 + h^2 + 2h^3 + h^4 - h^3 + 3h^4 + h^4 + h^4 \varepsilon(u) \\ &= 1 - h + h^3 + 5h^4 + h^4 \varepsilon(u). \end{aligned}$$

2.6.3 La fonction $e^{\cos x}$ est paire : son polynôme de Maclaurin à l'ordre 3 coïncide avec son polynôme de Maclaurin à l'ordre 2 en $x = 0$, ainsi

$$\begin{aligned} e^{\cos h} &= e^{1-h^2/2+h^2\varepsilon(h)} = e e^{-h^2/2+h^2\varepsilon(h)} \\ &= e \left[1 + (-h^2/2 + h^2\varepsilon(h)) + (-h^2/2 + h^2\varepsilon(h))^2/2 + h^2\varepsilon(h) \right] \\ &= e \left[1 - h^2/2 + h^2\varepsilon(h) \right] \\ &= e - e/2h^2 + h^2\varepsilon(h) \end{aligned}$$

et donc $e^{\cos h} = e - e/2h^2 + h^2\varepsilon(h)$ à l'ordre 3 vu la parité de la fonction. \triangleleft

L'énoncé suivant est un cas particulier de la formule de composition : si $g : u \mapsto 1/u$, alors $f^{-1}(x) = g \circ f(x)$.

PROPOSITION 2.4: Soit f ayant un développement limité à l'ordre n en 0 avec $f_0 = f(0) \neq 0$. Alors l'inverse $f^{-1} : x \rightarrow 1/f(x)$ a un développement limité à l'ordre n obtenu par troncature à l'ordre n du développement

$$\frac{1}{f(h)} = f_0^{-1} + \sum_{k=1}^n f_0^{-1} [f_0^{-1}(f_0 - P_{f,n}(h))]^k + h^n \varepsilon(h). \quad (2.2)$$

DÉMONSTRATION. Pour simplifier, posons $b_0 = f_0^{-1}$. On écrit

$$\begin{aligned} \frac{1}{f(h)} &= b_0 [1 - b_0(f_0 - f(h))]^{-1} \\ &= b_0 + \sum_{k=1}^n b_0^{1+k} [f_0 - P_{f,n}(h) - h^n \varepsilon(h)]^k + h^n \varepsilon(h) \\ &= b_0 + \sum_{k=1}^n b_0^{1+k} [f_0 - P_{f,n}(h)]^k + h^n \varepsilon(h) \\ &= f_0^{-1} + T_n \left[\sum_{k=1}^n b_0^{1+k} [f_0 - P_{f,n}(h)]^k \right] + h^n \varepsilon(h). \end{aligned}$$

Le polynôme $f_0 - P_{f,n}$ est nul en $t = 0$ de sorte que le terme $[P_{f,n}(h) - f_0]^n$ est de la forme $h^n R_n(h)$ avec R_n borné au voisinage de $h = 0$ et que le terme $[P_{f,n}(h) - f_0]^n \varepsilon(h)$ soit de la forme $h^n \varepsilon(h)$.

Ainsi, dans la deuxième ligne, les $h^n \varepsilon(h)$ dans les termes puissance de $k = 1, \dots, n$, et tous les produits où ils interviennent, peuvent être mis dans le reste $h^n \varepsilon(h)$ global, ce qui donne le développement de la troisième ligne, auquel il faut appliquer la troncature à l'ordre n . \square

▷ EXEMPLE 2.7: le développement de tang en $x = 0$, s'obtient en divisant le développement de sin par celui de cos. La fonction tang étant impaire, son développement limité à l'ordre 3 est aussi un développement à l'ordre 4

$$\begin{aligned} \text{tang } h &= \frac{\sin h}{\cos h} \\ &= \frac{h - h^3/6 + h^3 \varepsilon(h)}{1 - h^2/2 + h^3 \varepsilon(h)} \\ &= (h - h^3/6 + h^3 \varepsilon(h))(1 + h^2/2 + h^3 \varepsilon(h)) \\ &= h + (-1/6 + 1/2)h^3 + h^3 \varepsilon(h) \\ &= h + 1/3 h^3 + h^4 \varepsilon(h) \end{aligned}$$

On montre

$$\text{tang } h = h + \frac{1}{3}h^3 + \frac{2}{15}h^5 + \frac{17}{315}h^7 + \frac{62}{2835}h^9 + \frac{1382}{155925}h^{11} + \frac{21844}{6081075}h^{13} + \dots$$

une formule beaucoup plus complexe que celle du développement limité à l'origine de sa fonction réciproque arctan donné par l'énoncé suivant. \triangleleft

PROPOSITION 2.5 (intégration): Soit f admettant un développement limité d'ordre n en a . Alors la primitive $F(x) = \int_a^x f(t) dt$ (définie sur un voisinage de a) admet un développement limité d'ordre $n + 1$ en a tel que $P_{F,a,n+1}(x) = \int_0^x P_{f,a,n}(t) dt$

DÉMONSTRATION. Intégrons le développement limité $f(x) = P_{f,a,n}(x) + r(x)$ avec $r(x) = (x-a)^n \varepsilon(x-a)$ pour obtenir

$$F(x) = \int_a^x f(t) dt = \int_a^x P_{f,a,n}(t) dt + R(x), \quad (2.3)$$

où le deuxième terme intégral est un polynôme. Soit $\varepsilon > 0$ et $\alpha > 0$ tel que $|r(t)| \leq \varepsilon |t-a|^n$ pour $|t-a| \leq \alpha$. On a alors l'estimation sur le reste R

$$|R(x)| = \left| \int_a^x r(t) dt \right| \leq \varepsilon \left| \int_a^x |t-a|^n dt \right| \leq \frac{\varepsilon |x-a|^{n+1}}{(n+1)},$$

ce qui prouve que $R(x) = (x-a)^{n+1} \varepsilon(x)$ et donc que (2.3) est un développement limité à l'ordre $n+1$ en a . \square

TABLE II.2 – Autres développements limités de base en $x = x_0$ (II)

$f(x_0 + h)$	$P_{n,f}(h) + h^n \varepsilon(h)$
$\frac{1}{1+h}$	$1 - h + h^2 + \dots + (-1)^n h^n + h^n \varepsilon(h)$
$\ln(1+h)$	$h - \frac{h^2}{2} + \frac{h^3}{3} - \dots + (-1)^n \frac{h^{n+1}}{n+1} + h^{n+1} \varepsilon(h)$
$\frac{1}{1+h^2}$	$1 - h^2 + h^4 + \dots + (-1)^n h^{2n} + h^{2n+1} \varepsilon(h)$
$\text{arctg}(1+h)$	$h - \frac{h^3}{3} + \frac{h^5}{5} + \dots + (-1)^n \frac{h^{2n+1}}{2n+1} + h^{2n+2} \varepsilon(h)$

▷ EXEMPLES 2.8:

2.8.1 Dans le tableau II.2, la ligne pour \sin peut être déduite par intégration de celle pour \cos et inversement vu les relations

$$\sin x = \int_0^x \cos(u) du \quad \text{et} \quad \cos x = - \int_0^x \sin t dt.$$

2.8.2 les développements limités de $\ln(1+h)$ et $\arctan(h)$ sont obtenus par intégration de ceux de $(1+h)^{-1}$ et $(1+h^2)^{-1}$ resp..

2.8.3 Quand on intègre une fonction et un développement limité, ne pas oublier de tenir compte des constantes d'intégration!

2.8.4 Le développement limité d'une fonction de classe \mathcal{C}^n à l'ordre n peut être dérivé, donnant un développement limité à l'ordre $n-1$ de la dérivée f' : le

polynôme $P_{f',a,n-1}$ du développement limité de la dérivée est le dérivé du polynôme, i. e. $P'_{f',a,n} = P_{f',a,n-1}$. Il existe des développements limités qu'on ne peut dériver : il s'agit de développement limité de fonctions non de classe \mathcal{C}^n pour lesquelles on ne peut appliquer la formule de Taylor-Young. \triangleleft

THÉOREME 2.5 (Fonction réciproque): Soit f admettant un développement limité à l'ordre $n \geq 1$ en a avec $f'(a) \neq 0$. Alors la fonction f admet une fonction réciproque définie au voisinage de $b = f(a)$ et possédant un développement limité à l'ordre n .

DÉMONSTRATION. **A revoir!** Supposons que g admette un développement limité en $b = f(a)$ à l'ordre n . Admettant un développement limité à l'ordre $n \geq 1$, la fonction f est dérivable et, vu que $f'(a) \neq 0$, admet une fonction réciproque g dérivable définie au voisinage de $b = f(a)$. On va déterminer les coefficients du développement limité $g(b+k) = \sum_{j=0}^n g_j k^j + k^n \varepsilon(k)$ en écrivant le développement limité de la composée $g \circ f = Id$. Posons $\tilde{P}_{f,n} = P_{f,n} - f_0$ le polynôme de Taylor de f à l'ordre n privé de son terme constant $f_0 = f(a)$.

$$\begin{aligned} a+h &= g \circ f(a+h) = g(f(a) + \tilde{P}_{f,n}(h) + h^n \varepsilon(h)) \\ &= a + g_1 \tilde{P}_{f,n}(h) + g_2 [\tilde{P}_{f,n}(h)]^2 + g_3 [\tilde{P}_{f,n}(h)]^3 + \dots + g_n [\tilde{P}_{f,n}(h)]^n + h^n \varepsilon(h) \\ &= a + g_1 (f_1 h + f_2 h^2 + \dots) + g_2 (f_1 h + f_2 h^2 + \dots)^2 + \dots \end{aligned}$$

d'où, par unicité des développements limités, les équations sur les coefficients

$$\begin{aligned} 1 &= g_1 f_1 \\ 0 &= g_1 f_2 + g_2 f_1^2 \\ 0 &= g_1 f_3 + 2g_2 f_1 f_2 + g_3 f_1^3 \\ &\dots \\ 0 &= P_k(g_1, \dots, g_{k-1}, f_1, \dots, f_{k-1}) + g_k f_1^k, \end{aligned}$$

où les P_k sont des polynômes à coefficients entiers en les variables $g_1, \dots, g_{k-1}, f_1, \dots, f_{k-1}$. Vu que f_1 est non nul, la première équation détermine $g_1 = f_1^{-1}$ qui est non nul, puis la seconde équation détermine $g_2 = -g_1 f_2 f_1^{-2}$, puis la troisième $g_3 = -(g_1 f_3 + g_2 f_1 f_2) f_1^{-3}$ et successivement g_4, \dots, g_n .

Montrons l'existence d'un développement limité à l'ordre n pour l'inverse g de f . Pour simplifier, supposons $a = 0$ et $b = f(a) = 0$. On a $f(h) = P_{n,f}(h) + h^n \varepsilon(h)$ et, pour le polynôme Q_n uniquement déterminé ci-dessus, on pose $R_n(k) = g(k) - Q_n(k)$. Alors l'égalité $h = g \circ f(h)$ au voisinage de $h = 0$ permet d'écrire $h = Q_n(f(h)) + R_n(f(h))$, soit

$$h = Q_n(P_{n,f} + h^n \varepsilon(h)) + R_n(f(h)) = h + h^n \varepsilon(h) + R_n(f(h))$$

On obtient ainsi $R(f(h)) = h^n \varepsilon(h)$, d'où $R(k) = g(k)^n \varepsilon(g(k)) = k^n \varepsilon(k)$ vu que $|g(k)| \leq C|k|$ au voisinage de 0 (pour une certaine constance C). \square

▷ EXEMPLE 2.9: $(e^{t^2} - 1)/t$, $\tan t$, $t \mapsto f(t) = t + \alpha t^n + t^n \varepsilon(t)$ a sa réciproque avec développement limité $t \mapsto g(t) = t - \alpha t^n + t^n \varepsilon(t)$ \triangleleft

▷ EXEMPLE 2.10: Le calcul du développement limité de la fonction \tan

$$\tan(x) = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \frac{17}{315}x^7 + \frac{62}{2835}x^9 + \frac{1382}{155925}x^{11} + \dots$$

peut-être fait de diverses manières, en voici 7 explicitées pour obtenir des développements limités à l'ordre 5 : pour certains, il n'est pas nécessaire de faire des développements limités à l'ordre 5 pour des calculs intermédiaires ou inversement introduire des développements limités d'ordre 6 pour avoir le résultat final à l'ordre 5 (cf. exemples 2 et 6)

1. $\tan = \sin / \cos$

2. $\tan = \frac{1 - \cos 2x}{\sin(2x)}$

3. $\tan = \sqrt{-1 + 1/\cos^2(x)}$

4. $\tan' = \cos^{-2}$

5. $\tan' = 1 + \tan^2$

6. $-(\ln(\cos))'$

7. $\tan \circ \arctg = Id$

◁

2.5 Calculs de limites

L'usage de développements limités permet l'évaluation de limites, comme par exemple

$$\frac{\sin x}{x} = \frac{x - x^3/6 + x^3\varepsilon(x)}{x} = 1 - x^2/6 + x^2\varepsilon(x) \xrightarrow{x \rightarrow +0} 1$$

$$\frac{1 - \cos x}{(1 - e^x)^2} = \frac{x^2/2 + x^2\varepsilon(x)}{(-x + x\varepsilon(x))^2} = \frac{1/2 + \varepsilon(x)}{(-1 + \varepsilon(x))^2} \xrightarrow{x \rightarrow 0} 1$$

Le deuxième calcul de limite ne peut se faire directement par la règle de l'Hôpital⁷ vu que les dérivées du dénominateur et du numérateur sont toutes deux nulles en $x = 0$.

Par de simples changements de variable (par ex. $h = 1/x$), un développement limité en 0 se transforme en développement dit *asymptotique* au voisinage de $+\infty$ et permet d'examiner le comportement au voisinage de $+\infty$ d'une fonction $F(x)$, comme par exemple le changement de variable $h = 1/\sqrt{x}$ (x, h tous deux positifs)

7. Guillaume François Antoine de L'Hôpital, 1661 – 2 février 1704.

(x grand si et seulement si h voisin de 0^+) pour les développements

$$\begin{aligned}\sqrt{x+1} - \sqrt{x} &= \sqrt{1+h^{-2}} - \sqrt{h^{-2}} \\ &= h^{-1}(\sqrt{1+h^2} - 1) \\ &= h^{-1}(1 + h^2/2 + h^2\varepsilon(h) - 1) \\ &= \frac{h}{2} + h\varepsilon(h) \\ &= \frac{1}{2\sqrt{x}}(1 + \varepsilon(x)) \xrightarrow{x \rightarrow +\infty} 0.\end{aligned}$$

Cet exemple peut être traité plus rapidement sans recours à un développement limité, simplement en utilisant la quantité conjuguée :

$$\sqrt{x+1} - \sqrt{x} = \frac{1}{\sqrt{x+1} + \sqrt{x}} \xrightarrow{x \rightarrow +\infty} 0$$

Mais pour des recherches de limites voisines comme par exemple pour $(1+x)^a - x^a$, l'usage des développements limités est vraiment opportun.

2.6 Étude locale et asymptotique de courbes

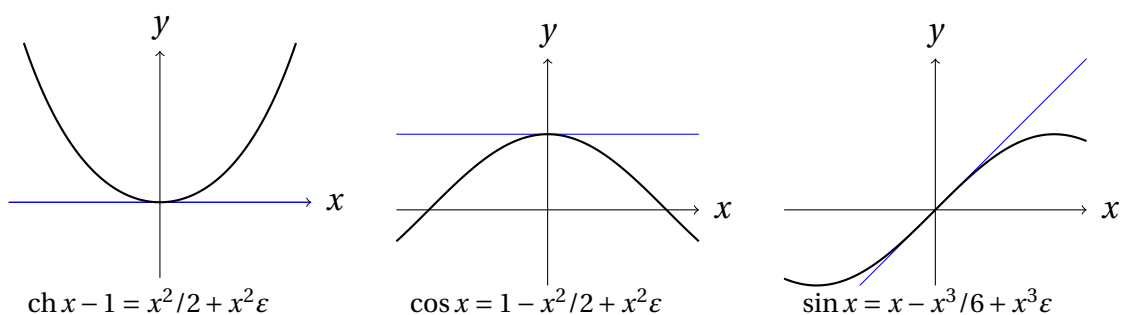


FIGURE II.2 – Vision graphique d'un développement limité $f(x) = f_0 + f_1x + f_2x^2 + \dots$ en $x = 0$: la tangente $\{Y = f_0 + f_1X\}$ approche le graphe $\{y = f(x)\}$ de la fonction f localement au voisinage de $x = 0$: celle-ci passe au-dessous, passe au-dessus ou traverse la tangente, suivant le signe du terme dominant du développement limité (f_2x^2 si $f_2 \neq 0$, f_3x^3 si $f_2 = 0$ et $f_3 \neq 0$), mis à part les termes de degré 0 et 1 paramétrant la tangente.

Un développement limité permet de préciser l'allure locale du graphe de fonctions. Soit f avec un développement limité en 0 d'ordre 2 :

$$f(x) = f_0 + f_1x + f_2x^2 + x^2\varepsilon(x).$$

La somme $f_0 + f_1x$ correspond à l'équation $Y = f_0 + f_1X$ de la tangente au graphe en $x = 0$. Si $f_2 \neq 0$, le développement limité permet de préciser la position relative de la

courbe vis-à-vis de la tangente : au-dessous si $f_2 < 0$, au-dessus si $f_2 > 0$. Si $f_2 = 0$, le développement à l'ordre 3, si le coefficient f_3 n'est pas nul, permet d'affirmer que le graphe de la fonction traverse sa tangente au point $(0, f(0))$. Comme exemples, on peut considérer $\sin, \sinh, \tan, \tanh, \operatorname{argth}, \dots$ au point $(0, 0)$.

Un développement limité permet parfois l'étude du comportement asymptotique de graphes.

Donnons l'exemple de $f(x) = \sqrt{x^2 + 3x + 4}$ définie sur \mathbb{R} . En posant $u = 1/x$, on

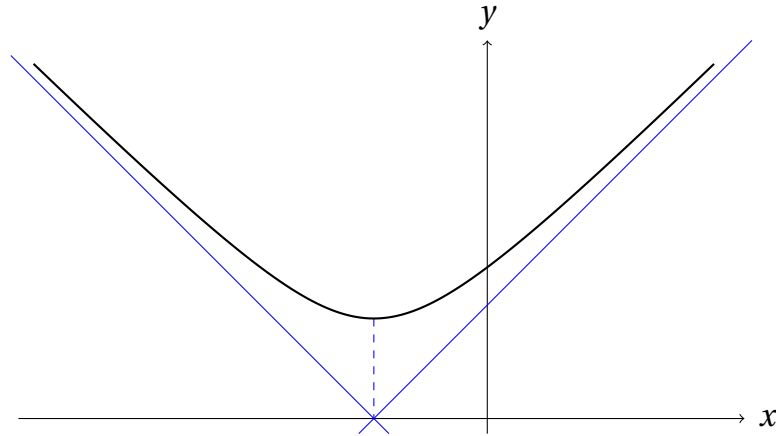


FIGURE II.3 – Le graphe de la fonction $x \in \mathbb{R} \mapsto \sqrt{x^2 + 3x + 4} \in \mathbb{R}$, ses deux asymptotes et son point de minimum.

a

$$\begin{aligned}
 f(x) &= x\sqrt{1 + 3x^{-1} + 4x^{-2}} \\
 &= x\sqrt{1 + 3u + 4u^2} \\
 &= x\left(1 + \frac{1}{2}(3u + 4u^2) - \frac{1}{8}(3u + 4u^2)^2 + u^2\varepsilon(u)\right) \\
 &= x\left(1 + \frac{3u}{2} + \left(2 - \frac{9}{8}\right)u^2 + u^2\varepsilon(u)\right) \\
 &= x\left(1 + \frac{3u}{2} + \frac{7}{8}u^2 + u^2\varepsilon(u)\right) \\
 &= x + \frac{3}{2} + \frac{7}{8}\frac{1}{x} + x^{-1}\varepsilon(x)
 \end{aligned}$$

Pour x grand, le graphe est voisin de la droite $Y = X + 3/2$, légèrement au dessus. Une étude analogue conclut à ce que, pour x au voisinage de $-\infty$, on ait

$$f(x) = -x - \frac{3}{2} - \frac{7}{8}\frac{1}{x} + x^{-1}\varepsilon(x)$$

et donc le graphe asymptote à la droite $Y = -X - 3/2$, légèrement au-dessus de la droite (ici x est négatif, vu qu'on fait l'étude asymptotique pour $x \sim -\infty$). Ces observations, avec en plus le marquage de quelques points, permettent un tracé approximatif du graphe de la fonction $x \in \mathbb{R} \mapsto \sqrt{x^2 + 3x + 4}$.

Chapitre 3

Intégrales généralisées

Dans ce chapitre, les fonctions considérées sont à valeurs réelles ou complexes. L'intégration sur des domaines non bornés apparaît naturellement en probabilités (et donc en physique). Par exemple, la loi normale est une distribution de probabilités de densité $(2\pi)^{-1/2}e^{-x^2/2}$, dont le calcul des moments (moyenne, variance,...) impliquent les intégrales $\int_{-\infty}^{+\infty} x^k e^{-x^2/2} dx$ pour $k \in \mathbb{N}$. La loi de Cauchy (souvent appelée *loi de Lorentz*) a pour densité $\pi^{-1} \frac{1}{1+x^2}$, sans que la moyenne $\int_{-\infty}^{+\infty} \frac{xdx}{\pi(1+x^2)}$ ne soit définie.

La construction de l'intégrale, ses diverses propriétés et ses méthodes de calcul (intégrations par parties, changement de variables, familles de primitives comme celles des fractions rationnelles notamment) sont supposées connues. L'usage d'intégrations par parties ou le changement de variables permettent diverses formulations de la même intégrale (cf. (3.3)).

3.1 Intégrale généralisée convergente

L'intégrale (de Riemann) $\int_a^b f(t)dt$ est définie pour toute fonction réelle étagée (ou en escalier) f telle que f est constante égale à $m_i \in \mathbb{R}$ sur les intervalles $]t_i, t_{i+1}[$ d'une subdivision finie $a = t_0 \leq t_1 \leq \dots \leq t_n = b$ de l'intervalle $I = [a, b]$ par

$$\int_a^b f(t)dt = \sum_{i=0}^{n-1} m_i(t_{i+1} - t_i).$$

Cette intégrale de fonctions en escalier est prolongée à des fonctions intégrables f au sens suivant : étant données les bornes A_f et B_f définies par

$$A_f = \sup_{u \leq f} \int_a^b u(t)dt, \quad B_f = \inf_{U \geq f} \int_a^b U(t)dt \quad (3.1)$$

où les fonctions u et U sont des fonctions étagées, la fonction f est dite *intégrable* si les bornes A_f et B_f sont égales, son intégrable étant définie comme cette valeur

commune, notée de manière classique suivant

$$\int_a^b f(t) dt = A_f = B_f.$$

Toute fonction continue ou continue par morceaux sur un intervalle borné est intégrable¹ : dans la pratique, c'est ce type de fonctions qui apparaissent. Dans la définition de l'intégrale l'intervalle $[a, b]$ est bornée et la fonction définie sur $[a, b]$ est bornée. Ce chapitre est consacré à la généralisation de cette intégrale.

DÉFINITION 3.1: *Soit I un intervalle d'intérieur non vide. Une fonction f est dite localement intégrable sur I si elle est intégrable sur tout segment $[a, b]$ inclus dans I .*

△ REMARQUES 3.1:

1. Pour un intervalle I semi-ouvert (par ex. du type $[a, b[$), la fonction est localement intégrable si elle est intégrable sur tous les intervalles $[a, t]$ avec $a \leq t < b$. Si I est un intervalle fermé, une fonction f localement intégrable sur I est une fonction intégrable sur tout l'intervalle $[a, b]$: l'intégrale $\int_a^b f(t) dt$ n'est pas à considérer comme intégrale généralisée. Dans les exemples suivants

$$\int_0^1 t^{-\alpha} dt, \quad \int_1^{\infty} t^{-\alpha} dt, \quad \int_0^1 \ln t dt, \quad \int_0^1 \frac{dt}{\sqrt{t(1-t)}},$$

c'est l'intégrabilité aux voisinages de l'extrémité 0 , $+\infty$, 0 , 0 et 1 resp. qui sera sujette à cette éventuelle intégrabilité généralisée.

2. Les définitions précédentes (et ultérieures) énoncées pour des fonctions à valeurs réelles valent la plupart du temps aussi pour les fonctions f à valeurs complexes, après avoir considéré cette fonction f comme équivalente à la donnée de 2 fonctions à valeurs réelles, soit sa partie réelle $\Re f$ et sa partie imaginaire $\Im f$. ▽

DÉFINITION 3.2: *Soit $I =]a, b[$ un intervalle ouvert d'extrémités a et b (éventuellement $+\infty$ ou $-\infty$), soit $t_0 \in I$ et $f : I \rightarrow \mathbb{R}$ localement intégrable sur I . L'intégrale $\int_a^b f(t) dt$ sera dite intégrale généralisée convergente (et la fonction intégrable de manière généralisée) si les limites*

$$\lim_{X \rightarrow a^+} \int_X^{x_0} f(t) dt, \quad \lim_{Y \rightarrow b^-} \int_{x_0}^Y f(t) dt$$

existent. Dans ce cas, on notera $\int_a^b f(t) dt$ la limite, dite intégrale généralisée (ou impropre).

1. L'intégrale $\int_a^b f(t) dt$ est aussi bien définie pour une fonction réglée, i. e. f est limite uniforme d'une suite (f_n) de fonctions en escalier : $\|f - f_n\|_{\infty} = \sup_{t \in [a, b]} |f(t) - f_n(t)| \rightarrow 0$ lorsque $n \rightarrow +\infty$.

△ REMARQUES 3.2:

1. On a donc, indépendamment du choix du point x_0 intérieur à l'intervalle $[a, b]$ (d'après la relation de Chasles)

$$\int_a^b f(t)dt = \lim_{X \rightarrow a^+} \int_X^{x_0} f(t)dt + \lim_{Y \rightarrow b^-} \int_{x_0}^Y f(t)dt$$

2. Il se peut que l'intégrale $\int_0^\infty f(t)dt$ soit convergente, sans que la fonction ne converge vers 0, ni que f soit bornée lorsque $t \rightarrow \infty$: la fonction f telle que $f(t) = 1 - 2^k|t - k|$ sur l'intervalle $I_k = [k - 2^{-k}, k + 2^{-k}]$ pour $k > 0$ et zéro ailleurs, n'a pas de limite lorsque $t \rightarrow \infty$, alors que la fonction g avec $g(t) = kf(t)$ sur I_k pour $k \geq 0$ n'est pas bornée. Les fonctions f et g sont cependant intégrables.
3. Au cas où a est fini et $f : [a, x_0] \rightarrow \mathbb{R}$ est Riemann intégrable sur l'intervalle borné $[a, x_0]$, l'intégrale $\int_x^{x_0} f(t)dt$ converge vers $\int_a^{x_0} f(t)dt$ lorsque $x \rightarrow a^+$: l'intégrale généralisée est bien définie pour une fonction intégrable sur un intervalle compact $[a, x_0]$ et les deux intégrales (Riemann et généralisée) sont égales.
4. Une intégrale peut être cataloguée comme généralisée, bien qu'elle ne le soit pas. Les intégrales suivantes

$$\int_0^1 \frac{\sin t}{t} dt, \quad \int_0^1 \frac{t}{e^t - 1} dt$$

concernent des fonctions continues sur $[0, 1]$.

5. Le théorème de changement de variable pour calculer les intégrales ou la méthode d'intégration par parties subsistent pour les intégrales généralisées, sous réserve d'existence de chacun des termes intervenant dans le calcul. ▽

▷ EXEMPLES 3.1:

3.1.1 L'intégrale $\int_0^Y e^{-t} dt = 1 - e^{-Y}$ converge vers 1 lorsque $Y \rightarrow +\infty$. On a donc la convergence de l'intégrale $\int_0^{+\infty} e^{-t} dt$ avec $\int_0^{+\infty} e^{-t} dt = 1$.

3.1.2 $\int_X^Y dt/(1+t^2) = [\arctg t]_X^Y = \arctg Y - \arctg X \rightarrow \pi$ si $X \rightarrow -\infty$ et $Y \rightarrow +\infty$.

3.1.3 $\int_X^1 dt/\sqrt{t} = [2\sqrt{t}]_X^1 = 2[1 - \sqrt{X}] \rightarrow 2$ lorsque $X \rightarrow 0^+$. Par ailleurs $\int_1^Y dt/\sqrt{t} = 2[\sqrt{Y} - 1]$ tend vers $+\infty$ lorsque $Y \rightarrow +\infty$, l'intégrale $\int_1^\infty dt/\sqrt{t}$ n'est pas convergente, pas plus que $\int_0^\infty dt/\sqrt{t}$.

3.1.4 $\int_1^Y dt/t = [\ln t]_1^Y = \ln Y$ tend vers $+\infty$ quand $Y \rightarrow +\infty$: l'intégrale $\int_1^{+\infty} dt/t$ ne converge pas.

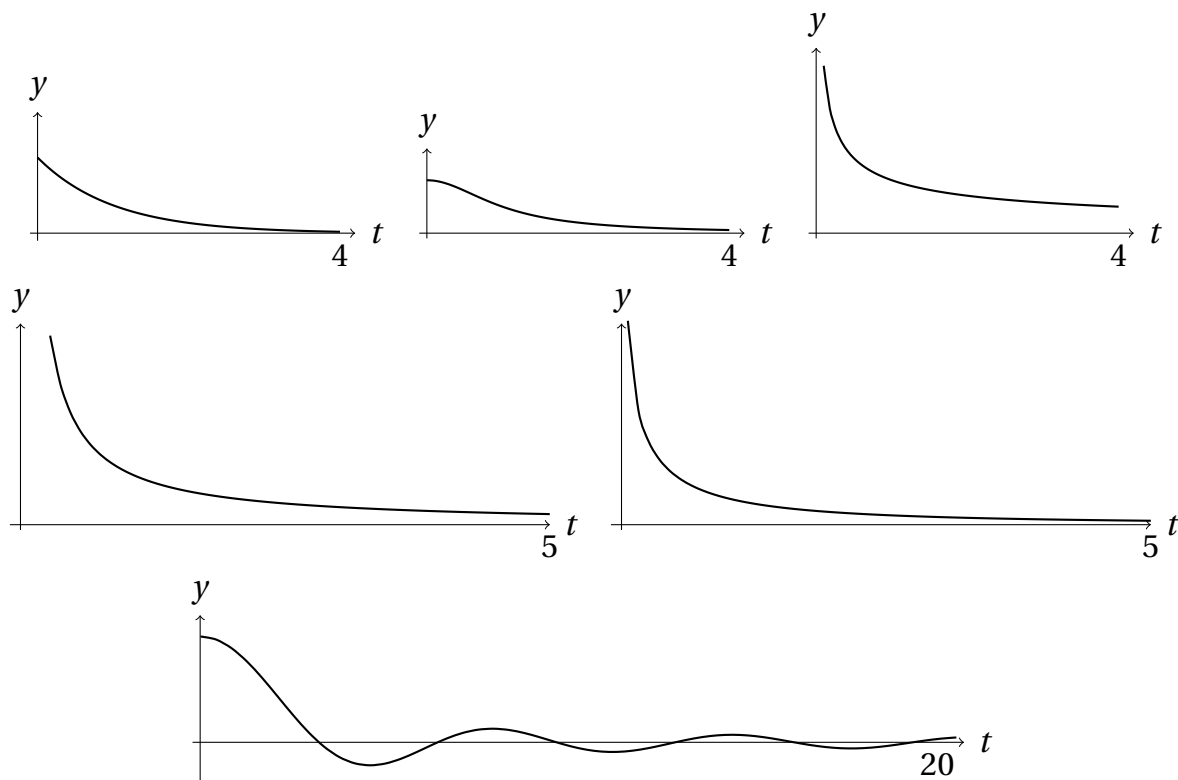


FIGURE III.1 – Les graphes des fonctions des exemples 3.1 : e^{-t} , $1/(1+t^2)$, $1/\sqrt{t}$, $1/t$, $1/((1+t)\sqrt{t})$ et $\sin(t)/t$.

3.1.5 Vu $2 \int ((1+t)\sqrt{t})^{-1} dt = \arctan \sqrt{t}$, on a

$$\begin{aligned} \int_0^{\infty} \frac{dt}{(1+t)\sqrt{t}} &= \lim_{X \rightarrow 0^+} \int_X^1 \frac{dt}{(1+t)\sqrt{t}} + \lim_{Y \rightarrow +\infty} \int_1^Y \frac{dt}{(1+t)\sqrt{t}} \\ &= \lim_{X \rightarrow 0^+} \left[\frac{\pi}{2} - 2 \arctan \sqrt{X} \right] - \lim_{Y \rightarrow +\infty} \left[2 \arctan \sqrt{Y} - \frac{\pi}{2} \right] = \pi. \end{aligned}$$

3.1.6 En intégrant par parties, et vu que $(1 - \cos t)/t^2$ est prolongeable par continuité en $t = 0$ (avec valeur $1/2$), on a lorsque $Y \rightarrow +\infty$

$$\int_0^Y \frac{\sin t}{t} dt = \left[\frac{1 - \cos t}{t} \right]_0^Y + \int_0^Y \frac{1 - \cos t}{t^2} dt \xrightarrow{Y \rightarrow +\infty} \int_0^{+\infty} \frac{1 - \cos t}{t^2} dt.$$

On montrera dans les exemples 3.2 que le dernier terme est absolument convergent (cf. définition 3.3 ci-dessous) : le théorème 3.4 ci-dessous assure alors que l'intégrale $\int_0^{+\infty} \frac{\sin t}{t} dt$ est semi-convergente. \triangleleft

Soit $f : [a, b[\rightarrow \mathbb{R}$ localement intégrable et $c \in]a, b[$. Alors l'intégrale $\int_c^b f(t) dt$ est convergente si et seulement si l'intégrale $\int_a^b f(t) dt$ l'est. Cela résulte immédia-

tement de la relation de Chasles² pour les intégrales :

$$\int_a^Y f(t) dt = \int_a^c f(t) dt + \int_c^Y f(t) dt.$$

Si $I = \mathbb{R}$, il ne suffit pas que $\int_{-x}^x f(t) dt$ converge lorsque $x \rightarrow +\infty$ pour que l'intégrale $\int_{-\infty}^{+\infty} f(t) dt$ soit déclarée convergente. Par exemple, si f est la fonction impaire $f : t \in \mathbb{R} \mapsto t^3$, les intégrales $\int_{-X}^X t^3 dt$ sont nulles, sans que l'intégrale généralisée $\int_{-\infty}^{+\infty} t^3 dt$ ne converge : $\int_0^Y t^3 dt = Y^4/4 \rightarrow +\infty$ et $\int_X^0 t^3 dt = -X^4/4 \rightarrow -\infty$ lorsque $X \rightarrow -\infty$.

THÉORÈME 3.1 (Intégrales de Riemann): *Soient α un réel.*

*L'intégrale $\int_1^{+\infty} t^{-\alpha} dt$ est convergente si et seulement si $\alpha > 1$.
L'intégrale $\int_0^1 t^{-\alpha} dt$ est convergente si et seulement si $\alpha < 1$.*

DÉMONSTRATION. On utilise les primitives

$$\int t^\alpha dt = \frac{t^{\alpha+1}}{\alpha+1} \text{ si } \alpha \neq -1, \quad \int t^{-1} dt = \ln|t|$$

pour déterminer les limites $\int_X^Y t^{-\alpha} dt$ quand $X \rightarrow 0^+$ et quand $Y \rightarrow +\infty$. □

On a aussi les intégrales dites de Bertrand

LEMME 3.1: *L'intégrale $\int_e^{+\infty} t^{-\alpha} \ln^{-\beta} t dt$ converge si et seulement si $\alpha > 1$ ou ($\alpha = 1$ et $\beta > 1$);*

L'intégrale $\int_0^{e^{-1}} t^{-\gamma} |\ln(t)|^{-\beta} dt$ converge si et seulement si $\gamma < 1$ ou ($\gamma = 1$ et $\beta > 1$).

3.2 Fonctions positives et convergence absolue

THÉORÈME 3.2: *Soit $f : [a, b[\rightarrow \mathbb{R}$, positive et localement intégrable. Alors l'intégrale $\int_a^b f(t) dt$ est convergente si et seulement si la fonction $X \in [a, b[\mapsto \int_a^X f(t) dt$ est bornée sur $[a, b[$.*

DÉMONSTRATION. Soit la fonction F définie par $F(X) = \int_a^X f(t) dt$ pour $X \in [a, b[$. Vu la positivité de f , la fonction F est croissante : la fonction F est majorée si et seulement si l'ensemble de ses valeurs est une partie majorée, si et seulement si cette partie a une borne supérieure qui est la limite $\lim_{x \rightarrow b^-} F(x)$. □

On a le résultat de comparaison

THÉORÈME 3.3: *Soient f et g deux fonctions positives, définies sur un intervalle $I = [a, b[$ et telles que $f(t) \leq g(t)$ pour tout $t \in I$.*

2. Michel Chasles, 15 novembre 1793, Épernon (Eure-et-Loir) – 18 décembre 1880, Paris.

1. Si l'intégrale $\int_a^b g(t) dt$ est convergente, alors il en est de même pour $\int_a^b f(t) dt$.
2. Si l'intégrale $\int_a^b f(t) dt$ ne converge pas, alors il en est de même pour $\int_a^b g(t) dt$.

DÉMONSTRATION. La relation de comparaison entre f et g sur l'intervalle $I =]a, b[$ induit une comparaison sur les intervalles

$$\int_a^X f(t) dt \leq \int_a^X g(t) dt, \quad X \in [a, b[.$$

Le théorème est démontré en usant du caractère borné (ou non borné) mis en avant dans le théorème précédent 3.2. \square

▷ EXEMPLES 3.2:

- 3.2.1 On a l'inégalité $e^{-t^2} \leq e^{-t}$ pour $t \geq 1$. En résulte la convergence de l'intégrale $\int_0^{+\infty} e^{-t^2} dt$.
- 3.2.2 Vu les inégalités $0 < \sin t \leq t$ pour $t \in]0, 1]$, l'intégrale $\int_0^1 dt/(\sin t)$ ne converge pas, à l'instar de l'intégrale $\int_0^1 dt/t$
- 3.2.3 La fonction définie par $f(t) = \sin^2 t/t^2$ sur $[0, +\infty[$ (avec prolongement par continuité en $t = 0$) est majorée par t^{-2} : l'intégrale $\int_0^{+\infty} \sin^2(t)t^{-2} dt$ est convergente comme $\int_1^{+\infty} t^{-2} dt$ vu le théorème 3.1. \triangleleft

DÉFINITION 3.3: Soit $f : t \in [t_0, b[\mapsto f(t) \in \mathbb{R}$ localement intégrable. L'intégrale $\int_{t_0}^b f(t) dt$ est dite absolument convergente si l'intégrale $\int_{t_0}^b |f(t)| dt$ est une intégrale généralisée convergente.

THÉORÈME 3.4: Soit $f : [x_0, b[\mapsto \mathbb{C}$ une fonction localement intégrable sur $[x_0, b[$ telle que l'intégrale $\int_{x_0}^b f(t) dt$ est absolument convergente. Alors l'intégrale $\int_{x_0}^b f(t) dt$ est convergente et

$$\left| \int_{x_0}^b f(t) dt \right| \leq \int_{x_0}^b |f(t)| dt.$$

△ REMARQUE 3.3: La réciproque est erronée : une intégrale convergente n'est pas nécessairement absolument convergente : un contre-exemple est donné par l'intégrale $\int_1^{+\infty} \frac{\sin x}{x} dx$. ∇

DÉMONSTRATION. Commençons par le cas f à valeurs réelles. La fonction f est décomposable en la somme $f = f_+ - f_-$ avec

$$f_+ = \sup(f, 0), \quad f_- = -\inf(f, 0)$$

deux fonctions positives telles que $f = f_+ - f_-$. Les fonctions f_+, f_- sont localement intégrables si f l'est et

$$\int_{x_0}^X f(t) dt = \int_{x_0}^X f_+(t) dt - \int_{x_0}^X f_-(t) dt.$$

Les fonctions (positives) f_+ et f_- sont majorées par $|f|$, donc avec une intégrale $\int_{x_0}^{+\infty} f_{\pm}(t) dt$ semi-convergente. Il y a donc aussi convergence pour l'intégrale généralisée $\int_{x_0}^{+\infty} [f_+(t) - f_-(t)] dt$. L'inégalité finale se déduit de la même majoration sur les intervalles fermés $[x_0, X]$.

Si f est à valeurs complexes, il suffit de remarquer $|\Re f| \leq |f|$ et $|\Im f| \leq |f|$: si la fonction $|f|$ à valeurs réelles positives a une intégrale généralisée convergente, il en est de même par comparaison pour $|\Re f|$ et $|\Im f|$, et donc par ce théorème qui vient d'être montré pour des fonctions à valeurs réelles, la semi-convergence de leurs intégrales, et donc aussi par définition, la convergence généralisée de l'intégrale de $f = \Re f + i\Im f$ sur $[a, b]$. \square

\triangle REMARQUE 3.4: On dit que l'intégrale $\int_0^{+\infty} f(t) dt$ est semi-convergente si l'intégrale n'est pas absolument convergente, *i. e.* l'intégrale $\int_0^{+\infty} |f(t)| dt$ n'est pas convergente. ∇

3.3 Intégrales de Gauß, Dirichlet et Fresnel.

Le calcul d'intégrale particulière fait appel à divers développements du calcul différentiel et intégral, par exemple les développements limités ou asymptotique comme lorsque $t \rightarrow +\infty$. Ainsi l'intégrant

$$\frac{\sin t}{t + \cos t} = \frac{\sin t}{t} \frac{1}{1 + \frac{\cos t}{t}} = \frac{\sin t}{t} \left(1 - \frac{\cos t}{t} + t^{-1} \varepsilon(t) \right) = \frac{\sin t}{t} - \frac{\sin t \cos t}{t^2} + t^{-2} \varepsilon(t),$$

est la somme d'un terme d'intégrale semi-convergente et d'un terme absolument convergent.

3.3.1 Intégrale de Gauß

Le théorème ci-dessous donne explicitement la transformée de Fourier de la gaussienne³. En particulier, on a $\int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi}$.

THÉORÈME 3.5: *L'intégrale gaussienne $t \mapsto \int_{-\infty}^{+\infty} e^{-t^2/2} e^{itx}$ est absolument convergente et obéit à la formule*

$$\int_{-\infty}^{+\infty} e^{itx} e^{-t^2/2} dt = \sqrt{2\pi} e^{-x^2/2}, \quad x \in \mathbb{R}.$$

3. Carl Friedrich Gauss 30 avril 1777, Brunswick, Basse-Saxe – 23 février 1855, Göttingen, Basse-Saxe, Allemagne.

△ REMARQUE 3.5: La formule précédente exprime la transformée de Fourier de la gaussienne : en général, la transformée de Fourier d'une fonction $f : \mathbb{R} \rightarrow \mathbb{C}$ est la fonction, notée \widehat{f} , définie sur \mathbb{R} et à valeurs complexes telle que

$$\widehat{f}(x) = \int_{\mathbb{R}} e^{ixt} f(t) dt, \quad x \in \mathbb{R}.$$

La définition de la transformée de Fourier présuppose des hypothèses convenables d'intégrabilité pour la fonction f . Il y a peu de fonctions f pour lesquelles le calcul de la transformée de Fourier \widehat{f} peut se conclure en termes de fonctions classiques : par exemple, si la gaussienne $t \in \mathbb{R} \mapsto e^{-t^2} \in \mathbb{R}$ a une transformée de Fourier simple, il n'en est pas de même de la fonction $t \mapsto \chi_{[a,b]}(t)e^{-t^2}$ à support borné $[a, b]$. ▽

DÉMONSTRATION. Soit F définie par $F(x) = \int_{-\infty}^{+\infty} e^{itx} e^{-t^2/2} dt$. La fonction F est bien définie, l'intégrale étant (à x fixé) absolument convergence. On montre, et il sera admis, que la fonction F est continue et dérivable indéfiniment sous le signe d'intégration.

$$F'(x) = i \int_{-\infty}^{+\infty} t e^{itx} e^{-t^2/2} dt = -i \left[e^{itx} e^{-t^2/2} \right]_{-\infty}^{+\infty} + i \int_{-\infty}^{+\infty} i x e^{itx} e^{-t^2/2} dt = -x F(x)$$

La fonction F vérifie l'équation différentielle $y'(x) + xy(x) = 0$ et est donc de la forme $F(x) = Ge^{-x^2/2}$ avec la constante $G = \int_{-\infty}^{+\infty} e^{-t^2/2} dt$ à calculer.

Le calcul de cette constante intégrale passe classiquement par l'intégration de la fonction $(x, y) \in \mathbb{R}^2 \mapsto e^{-(x^2+y^2)/2}$ sur le plan $\mathbb{R}_{x,y}^2$, mais il existe une preuve basée sur l'étude d'intégrales de fonctions d'une variable réelle (dépendant d'un paramètre réel).

Soient les fonctions Ψ et Φ définies sur \mathbb{R}^+ par

$$\Psi(x) = \left(\int_0^x e^{-s^2} ds \right)^2, \quad \Phi(x) = \int_0^1 \frac{e^{-x^2(1+u^2)}}{1+u^2} du$$

avec pour dérivées (la justification de la dérivation sous le signe d'intégration pour Φ est admise)

$$\begin{aligned} \Psi'(x) &= 2e^{-x^2} \int_0^x e^{-s^2} ds, \\ \Phi'(x) &= -2x \int_0^1 e^{-x^2(1+u^2)} du = -2e^{-x^2} \int_0^x e^{-v^2} dv \end{aligned}$$

Par le changement de variable $v = xu$ dans Φ' , on a donc établi $\Phi'(x) = -\Psi'(x)$, ainsi

$$\Psi(x) + \Phi(x) = \Psi(0) + \Phi(0) = 0 + \int_0^1 \frac{du}{1+u^2} = \left[\arctan u \right]_0^1 = \frac{\pi}{4}, \quad x \geq 0$$

et la conclusion

$$\pi/4 = \Psi(+\infty) + \Phi(+\infty) = \left(\int_0^{+\infty} e^{-s^2} ds \right)^2$$

après avoir remarqué

$$0 \leq \Phi(+\infty) \leq \lim_{x \rightarrow +\infty} e^{-x^2} \int_0^1 (1+u^2)^{-1} du = 0. \quad \square$$

On a donc $\int_{-\infty}^{+\infty} e^{-s^2} ds = \sqrt{\pi}$, soit $\int_{-\infty}^{+\infty} e^{-t^2/2} dt = \sqrt{2\pi}$.

3.3.2 Intégrale de Dirichlet

Le théorème suivant utilise aussi le théorème de convergence absolue 3.4 pour établir la semi-convergence de l'intégrale $\int_0^{+\infty} t^{-1} \sin t dt$ (dont le calcul de la valeur exacte est délicat).

THÉORÈME 3.6: *L'intégrale $\int_0^{+\infty} \frac{\sin t}{t} dt$ est semi-convergente sans être absolument convergente. On a*

$$\int_0^{+\infty} \frac{\sin t}{t} dt = \int_0^{+\infty} \left[\frac{\sin t}{t} \right]^2 dt = \pi/2,$$

et

$$\lim_{n \rightarrow +\infty} \sum_{k=1}^n \frac{\sin(kx)}{k} = -\frac{x}{2} + \int_0^{+\infty} \frac{\sin t}{t} dt, \quad x \in]0, 2\pi[.$$

DÉMONSTRATION. Effectuant une intégration par parties, on obtient

$$\int_0^X \frac{\sin t}{t} dt = \left[\frac{1 - \cos t}{t} \right]_0^X + \int_0^X \frac{1 - \cos t}{t^2} dt = \frac{1 - \cos X}{X} + \int_0^X \frac{1 - \cos t}{t^2} dt$$

l'intégrale dans le dernier terme étant convergente : la convergence de la première est donc établie, avec

$$\int_0^X \frac{\sin t}{t} dt = \int_0^X \frac{1 - \cos t}{t^2} dt.$$

Pour la non convergence absolue affirmée par le théorème, vu l'inégalité $|\sin t| \geq |\sin^2 t|$ et l'égalité $\sin^2 t = (1 - \cos(2t))/2$, on a

$$\int_1^X \left| \frac{\sin t}{t} \right| dt \geq \int_1^X \left| \frac{\sin^2 t}{t} \right| dt = \int_1^X \frac{dt}{2t} - \int_1^X \frac{\cos 2t}{2t} dt.$$

L'intégrale $\int_1^{+\infty} (2t)^{-1} dt$ ne converge pas, alors que une intégration par parties montre la convergence de $\int_0^{+\infty} (\cos 2t)/(2t) dt$.

Pour calculer la valeur de cette intégrale semi-convergente, introduisons la fonction F définie par

$$F(x) = \int_0^{+\infty} e^{-tx} \frac{\sin t}{t} dt, \quad x \geq 0$$

La suite est basée sur des manipulations dont la justification fait appel à des résultats au-delà de ce cours : on les admettra.

Le calcul d'une primitive

$$\begin{aligned}\int e^{-tx} \sin(t) dt &= \Im m \left[\int e^{-(x-i)t} dt \right] = \Im m \left[\frac{e^{-(x-i)t}}{-x+i} \right] \\ &= \frac{e^{-xt}}{1+x^2} \Im m (e^{it}(-x-i)) = \frac{e^{-xt}}{1+x^2} (-x \sin t - \cos t)\end{aligned}$$

suivi d'une intégration par parties donne

$$\int_0^X e^{-tx} \frac{\sin(t)}{t} dt = \left[\frac{e^{-xt}(-x \sin t - \cos t) + 1}{(x^2 + 1)t} \right]_0^X + \int_0^X \frac{e^{-xt}(-x \sin t - \cos t) + 1}{(x^2 + 1)t^2} dt$$

où au voisinage de $t = 0$ et pour x borné, on a

$$\begin{aligned}e^{-xt}(-x \sin t - \cos t) + 1 &= \left(1 - xt + \frac{(xt)^2}{2}\right) \left(-xt - 1 + \frac{t^2}{2}\right) + 1 + t^2 \varepsilon(t) \\ &= \frac{(1+x^2)t^2}{2} + t^2 \varepsilon(t)\end{aligned}$$

ce qui permet d'écrire

$$F(x) = \int_0^{+\infty} \frac{e^{-xt}(-x \sin t - \cos t) + 1}{(x^2 + 1)t^2} dt.$$

La fonction F est continue sur \mathbb{R}^+ avec majoration

$$|F(x)| \leq \int_0^{+\infty} e^{-xt} dt = \frac{1}{x}, \quad x > 0 \quad (3.2)$$

et dérivable sur \mathbb{R}^{*+} avec dérivée

$$F'(x) = - \int_0^{+\infty} e^{-tx} \sin(t) dt = - \Im m \left[\int_0^{+\infty} e^{-(x-i)t} dt \right] = - \Im m \left[\frac{1}{x-i} \right] = \frac{-1}{1+x^2}$$

Ainsi $F(x) = C - \operatorname{arctg} x : F(+\infty) = C - \pi/2 = 0$ vu (3.2), soit finalement $F(x) = \pi/2 - \operatorname{arctg} x$ et par continuité en $x = 0^+$ $F(0^+) = \pi/2$, soit $\int_0^{+\infty} \sin t (dt/t) = \pi/2$.

En complément, une intégration par parties donne l'identité

$$\int_0^{+\infty} \left(\frac{\sin t}{t}\right)^2 dt = \left[-\frac{\sin^2 t}{t}\right]_0^{+\infty} + \int_0^{+\infty} \frac{2 \sin t \cos t}{t} dt = \int_0^{+\infty} \frac{\sin 2t}{t} dt = \int_0^{+\infty} \frac{\sin t}{t} dt$$

Soit $x \in (0, 2\pi)$ (le résultat est faux pour $x = 0$!). Vu que

$$\begin{aligned}\sum_{k=1}^n \cos(kt) &= \Re e \left[\sum_{k=1}^n e^{ikt} \right] = \Re e \left[e^{it} \frac{1 - e^{itn}}{1 - e^{it}} \right] = \Re e \left[e^{i(n+1)t/2} \frac{e^{-itn/2} - e^{int/2}}{e^{-it/2} - e^{it/2}} \right] \\ &= \frac{\cos[(n+1)t/2] \sin(nt/2)}{\sin(t/2)} = \frac{\sin[(2n+1)t/2] - \sin(t/2)}{2 \sin(t/2)} \\ &= \frac{\sin[(2n+1)t/2]}{2 \sin(t/2)} - \frac{1}{2}\end{aligned}$$

où on a utilisé $\sin a \cos b = (\sin(a+b) + \sin(a-b))/2$, d'où par suite

$$\begin{aligned} \sum_{k=1}^n \frac{\sin(kx)}{k} &= \int_0^x \sum_{k=1}^n \cos(kt) dt = -\frac{x}{2} + \int_0^x \frac{\sin((2n+1)t/2)}{2 \sin(t/2)} dt \\ &= -\frac{x}{2} + \int_0^x \left[\frac{1}{2 \sin(t/2)} - \frac{1}{t} \right] \sin((2n+1)t/2) dt + \int_0^x \frac{\sin((2n+1)t/2)}{t} dt \\ &= -\frac{x}{2} + \varepsilon(n) + \int_0^{(2n+1)x/2} \frac{\sin u}{u} du \end{aligned}$$

La première intégrale (notée $\varepsilon(n)$) tend vers 0 par le lemme de Riemann-Lebesgue

LEMME 3.2: Soit $f : [a, b] \rightarrow \mathbb{R}$ continue. La suite $\left(\int_a^b f(t) e^{int} dt \right)_{n \geq 0}$ converge vers 0 lorsque $n \rightarrow \infty$.

Démonstration du lemme. Une intégration par partie permet de le montrer pour f dérivable ou en escalier. On approche f continue par f_ε dérivable (ou en escalier) telle que $\sup_{t \in [a, b]} |f(t) - f_\varepsilon(t)| dt \leq \varepsilon$ de telle sorte que

$$\limsup_n \left| \int_a^b f(t) e^{int} dt \right| \leq \int_a^b |f(t) - f_\varepsilon(t)| dt + \lim_n \left| \int_a^b e^{int} f_\varepsilon(t) dt \right| \leq \varepsilon(b-a)$$

pour tout $\varepsilon > 0$, ce qui achève la démonstration. \square

La seconde intégrale a été transformée en $\int_0^{(n+1/2)x} \frac{\sin t}{t} dt$ après changement de variable linéaire $u = (n+1/2)t$, ce qui permet de montrer

$$\lim_{n \rightarrow +\infty} \sum_{k=1}^n \frac{\sin(kx)}{k} = -\frac{x}{2} + \int_0^{+\infty} \frac{\sin t}{t} dt. \quad \square$$

Par une intégration par partie et un changement de variable (linéaire : $u = 2v$), on montre par ailleurs

$$\int_0^{+\infty} \frac{\sin u}{u} du = \int_0^{+\infty} \frac{1 - \cos u}{u^2} du = \int_0^{+\infty} \left[\frac{\sin v}{v} \right]^2 dv. \quad (3.3)$$

3.3.3 Intégrale de Fresnel

L'intégrale de Fresnel $\int_0^{+\infty} e^{it^2} dt$ est semi-convergente : elle n'est pas absolument convergente, mais elle est convergente comme un changement de variable suivie d'intégration par parties permettent de l'établir

$$\int_1^X e^{ix^2} dx = \frac{1}{2} \int_1^{X^2} \frac{e^{iu}}{\sqrt{u}} du = \frac{1}{2} \left[\frac{-ie^{iu}}{\sqrt{u}} \right]_1^{X^2} - \frac{1}{4} \int_1^{X^2} \frac{e^{iu}}{u^{3/2}} du,$$

Les parties réelle et imaginaire de l'intégrale de Fresnel sont aussi semi-convergentes.

THÉORÈME 3.7: *Les intégrales de Fresnel*

$$\int_0^{+\infty} \sin t^2 dt, \quad \int_0^{+\infty} \cos t^2 dt$$

sont semi-convergentes de même valeur $\sqrt{2\pi}/4$.

DÉMONSTRATION. Soit la fonction F définie par

$$F(t) = \int_0^{+\infty} \frac{e^{-(u^2+i)t^2}}{u^2+i} du,$$

qui est bornée, avec intégrant majoré par la fonction intégrable $(u^4+1)^{-1/2}$ et continue sur \mathbb{R} (admis). Les valeurs $F(t)$ sont majorées par

$$|F(t)| \leq \int_0^{+\infty} e^{-u^2 t^2} du = \frac{1}{t} \int_0^{+\infty} e^{-v^2} dv, \quad t > 0,$$

de telle sorte que $F(+\infty) = 0$. La fonction F est de classe \mathcal{C}^1 sur \mathbb{R}^{+*} , avec dérivée

$$F'(t) = -2te^{-it^2} \int_0^{+\infty} e^{-u^2 t^2} du = -2e^{-it^2} \int_0^{+\infty} e^{-v^2} dv = -e^{-it^2} \sqrt{\pi}.$$

Ainsi

$$\sqrt{\pi} \int_0^{+\infty} e^{-it^2} dt = \int_0^{+\infty} -F'(t) dt = F(0) - F(+\infty) = \int_0^{+\infty} \frac{du}{u^2+i}$$

Cette dernière intégrale est calculable en décomposant en éléments simples des fractions rationnelles. Remarquons

$$\int_0^{+\infty} \frac{du}{u^2+i} = \int_0^{+\infty} \frac{u^2-i}{u^4+1} du = \int_0^{+\infty} \frac{u^2 du}{u^4+1} - i \int_0^{+\infty} \frac{du}{u^4+1} = (1-i) \frac{\pi}{2\sqrt{2}}$$

où les deux dernières intégrales sont égales (faire le changement de variable $u = 1/v$). Vu $1+u^4 = (u^2 + \sqrt{2}u + 1)(u^2 - \sqrt{2}u + 1)$, on obtient

$$\frac{1}{1+u^4} = \frac{1}{2\sqrt{2}} \left[\frac{u + \sqrt{2}}{u^2 + \sqrt{2}u + 1} + \frac{-u + \sqrt{2}}{u^2 - \sqrt{2}u + 1} \right]. \quad (3.4)$$

et

$$\begin{aligned} \frac{u + \sqrt{2}}{u^2 + \sqrt{2}u + 1} &= \frac{u + \sqrt{2}}{\left(u + \frac{\sqrt{2}}{2}\right)^2 + \frac{1}{2}} = \frac{1}{2} \frac{[2u + \sqrt{2}] + \sqrt{2}}{\left(u + \frac{\sqrt{2}}{2}\right)^2 + \frac{1}{2}} \\ &= \frac{1}{2} \frac{2u + \sqrt{2}}{\left(u + \frac{\sqrt{2}}{2}\right)^2 + \frac{1}{2}} + \frac{\sqrt{2}/2}{\left(u + \frac{\sqrt{2}}{2}\right)^2 + \frac{1}{2}} \\ &= \frac{1}{2} \frac{2u + \sqrt{2}}{\left(u + \frac{\sqrt{2}}{2}\right)^2 + \frac{1}{2}} + \frac{\sqrt{2}}{(\sqrt{2}u + 1)^2 + 1} \end{aligned}$$

puis, si $I(u)$ est le contenu du crochet de (3.4),

$$\begin{aligned} \int_{-X}^X I(u) du &= \left[\frac{1}{2} \log \left[\left(u + \frac{\sqrt{2}}{2} \right)^2 + \frac{1}{2} \right] - \frac{1}{2} \log \left[\left(-u + \frac{\sqrt{2}}{2} \right)^2 + \frac{1}{2} \right] \right]_{-X}^{-X} \\ &\quad + \left[\operatorname{arctg}[\sqrt{2}u + 1] - \operatorname{arctg}[-\sqrt{2}u + 1] \right]_{-X}^X \\ &= 2\pi + \mathcal{O}(X^{-1}), \quad \text{lorsque } |X| \rightarrow +\infty \end{aligned}$$

où on a utilisé

$$\begin{aligned} \ln(|X + A|^2 + B) &= \ln(X^2 + 2AX + A^2 + B) = \ln(X^2) + \ln(1 + 2AX^{-1} + (A^2 + B)X^{-2}) \\ &= \ln(X^2) + \mathcal{O}(|X|^{-1}) \quad \text{si } |X| \rightarrow +\infty. \end{aligned}$$

Ainsi

$$\int_0^{\infty} \frac{du}{1+u^4} = \frac{\pi}{2\sqrt{2}}. \quad \square$$

ce qui achève le calcul des intégrales de Fresnel.

3.4 Fonctions Γ et B

DÉFINITION 3.4: La fonction Γ (*\ga.ma*) est définie pour $z \in \{\Re z > 0\}$ et la fonction B (*be.ta*) pour $x, y > 0$ suivant

$$\begin{aligned} \Gamma(z) &= \int_0^{+\infty} t^{z-1} e^{-t} dt, \quad \Re z > 0, \\ B(x, y) &= \int_0^1 t^{x-1} (1-t)^{y-1} dt, \quad \Re x, \Re y > 0. \end{aligned}$$

On vérifie les convergences de la première intégrale pour z réel positif non nul : en $t = 0$, l'intégrand est majoré par l'intégrand de Riemann $\int_0 t^{z-1} dt$ convergent si $z > 0$; en $t = +\infty$, l'intégrand est majoré par $Me^{-t/2}$ avec $M = \sup_{t \geq 0} [t^{\Re z - 1} e^{-t/2}]$. Il y a donc convergence de l'intégrale avec z réel positif. La convergence de l'intégrale vaut aussi pour $z \in \{z \in \mathbb{C} | \Re z > 0\}$ en utilisant simplement le théorème 3.4 sur les intégrales absolument convergentes.

Pour l'intégrale définissant la fonction B , il y a convergence pour $\Re x > 0, \Re y > 0$ avec $|t^x| = t^{\Re x}, |(1-t)^y| = (1-t)^{\Re y}$ où $t \in [0, 1]$, suivant la comparaison avec les intégrales de Riemann (cf. théorème 3.1).

On a les valeurs particulières

$$\Gamma(1) = 1, \quad \Gamma\left(\frac{1}{2}\right) = 2 \int_0^{+\infty} e^{-u^2} du = \sqrt{\pi}$$

et l'équation fonctionnelle

LEMME 3.3: *La fonction γ vérifie la relation fonctionnelle*

$$\Gamma(z+1) = z\Gamma(z), \quad \Re z > 0$$

et a comme valeurs particulières $\Gamma(1/2) = \sqrt{\pi}$, d'où résultent les formules

$$\Gamma(n+1) = n!, \quad \Gamma\left(n + \frac{1}{2}\right) = \frac{(2n)!\sqrt{\pi}}{4^n n!}, \quad n \in \mathbb{N}.$$

DÉMONSTRATION. L'équation fonctionnelle provient d'une intégration par parties

$$\Gamma(z+1) = \int_0^{+\infty} t^z e^{-t} dt = [-t^z e^{-t}]_0^{+\infty} + \int_0^{+\infty} z t^{z-1} e^{-t} dt = z\Gamma(z).$$

Les valeurs de Γ sur \mathbb{N} et $\mathbb{N} + \frac{1}{2}$ résultent de cette équation fonctionnelle. \square

Il y a de multiples formules impliquant les fonction Γ et B . Leur démonstration (admise ici) nécessitent les outils de la théorie des fonctions dérivables d'une variable complexe. Citons la relation, dite formule des compléments,

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}, \quad \Re z \in (0, 1)$$

qui permet de prolonger la fonction Γ au delà de l'axe imaginaire $\{\Re z = 0\}$. Les fonctions Γ et B sont reliées par la formule

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}, \quad \Re x > 0, \Re y > 0.$$

Euler a établi la représentation

$$\Gamma(z) = \lim_{n \rightarrow \infty} \frac{n!(n+1)^z}{z(z+1)\dots(z+n)}$$

alors que Weierstrass⁴ a montré la formule en produit infini

$$\Gamma(z) = \frac{e^{-\gamma z}}{z} \prod_{n=1}^{+\infty} \left(1 + \frac{z}{n}\right)^{-1} e^{z/n}, \quad \Re z > 0,$$

où γ est la constante d'Euler-Mascheroni⁵. Cette formule permet aussi de prolonger la fonction Γ de manière analytique à tout le plan complexe avec des pôles simples en $z \in -\mathbb{N}$. Terminons par la formule de Stirling⁶

$$\Gamma(z+1) \simeq_{z \rightarrow +\infty} \sqrt{2\pi z} \left(\frac{z}{e}\right)^z.$$

4. Karl Weierstraß, 31 octobre 1815, Ostenfelde, Westphalie, – 19 février 1897, Berlin.

5. Lorenzo Mascheroni, 13 mai 1750, Bergame — 14 juillet 1800, Paris.

6. James Stirling, mai 1692, Garden, Stirling – 5 décembre 1770, Édimbourg.

Chapitre 4

Optimisation de fonctions en plusieurs variables

Dans la suite, on note les boules (ouverte et fermée resp.) centrées en $\mathbf{x} \in \mathbb{R}^m$ et de rayon $r \geq 0$ suivant

$$B(\mathbf{x}, r) = \{\mathbf{y} \in \mathbb{R}^m \mid \|\mathbf{y} - \mathbf{x}\| < r\}, \quad \overline{B(\mathbf{x}, r)} = \{\mathbf{y} \in \mathbb{R}^m \mid \|\mathbf{y} - \mathbf{x}\| \leq r\},$$

La distance $\|\cdot\|_2$ est la distance euclidienne $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^m x_i^2}$ pour $\mathbf{x} = (x_1, \dots, x_m) \in \mathbb{R}^m$.

Une partie Ω de \mathbb{R}^m est dite *ouverte* si pour tout point $\mathbf{m} \in \Omega$, il existe un $\alpha > 0$ (éventuellement très petit) tel que la boule ouverte $B(\mathbf{m}, \alpha)$ soit incluse dans Ω .

DÉFINITION 4.1: Soit X une partie de \mathbb{R}^m et U une fonction $U : X \rightarrow \mathbb{R}$. Le point $x_* \in X$ est un minimum (resp. maximum) local en $x_* \in X$ s'il existe $\alpha > 0$ tel que $U(x) \geq U(x_*)$ ($U(x) \leq U(x_*)$ resp.) pour tout $x \in B(x_*, \alpha)$.

On emploiera le terme d'*optimum* pour un minimum ou un maximum. Un optimum est dit strict si les inégalités sont strictes; l'optimum est dit global si l'inégalité vaut pour tout $x \in X$. Sans ce caractère local, l'optimum sera dit implicitement *global*.

Il faut veiller à ne pas confondre point de minimum et valeur minimum (voire borne inférieure) de la fonction :

$$\operatorname{argmin}_{x \in X} U(x) = \{x_* \in X, U(x_*) = U_*\}, \quad U_* = \min_{x \in X} [U(x)] = \min_X U$$

La définition de U_* vaut si U admet un minimum, *i. e.* si la partie $\operatorname{argmin}_{x \in X} U(x)$ n'est pas vide.

Si la fonction est minorée, alors la borne inférieure $U_* = \inf_{x \in X} [U(x)]$ est bien définie, mais pas nécessairement atteinte par un $x_* \in X$. Suivant les cas, on pourra s'intéresser seulement à la valeur minimum ou au point de minimum.

Il y a des relations entre les optima de U et $-U$. Si x_* est un minimum de U , alors x_* est un maximum de $-U$:

$$\min_{x \in X} [U(x)] = -\max_{x \in X} [-U(x)], \quad \operatorname{argmin}_{x \in X} U(x) = \operatorname{argmax}_{x \in X} -U(x),$$

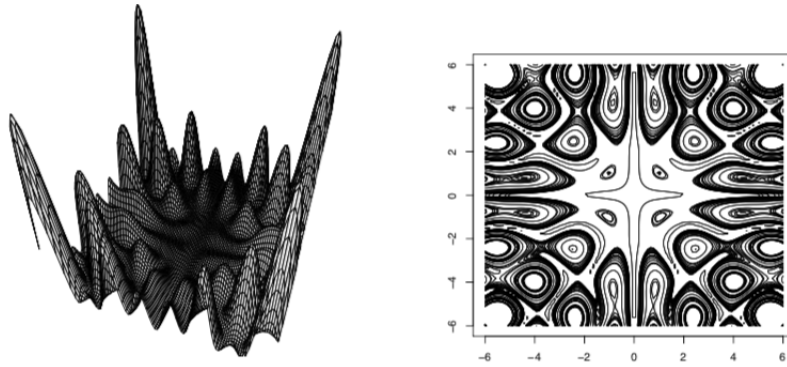


FIGURE IV.1 – Les lignes de niveau et la surface du graphe pour la fonction $U(x, y) = (x \sin(2y) + y \sin(2x))^2 \operatorname{ch}(x \sin(x)/10) + (x \cos y - y \sin x)^2 \operatorname{ch}(y \cos(2y)/10)$.

ainsi suffit-il de traiter les problèmes de minimisation, leur traduction pour des maxima n'étant en général pas évoquée explicitement. On se limitera dans la suite à ne considérer le plus souvent que des minima.

La représentation graphique des variations d'une fonction, en particulier ses optima, se fait avec des le tracé des lignes de niveau (introduites dans le [chapitre 1](#)) ou la surface du graphe : la figure [IV.1](#) exhibe une fonction avec beaucoup d'optima locaux, illustrant ainsi la difficulté de localisation d'optimum global. Néanmoins, la figure [IV.2](#) exhibe dans un voisinage du point critique de U le portrait des lignes de niveau : d'une part un ensemble de cercles qui est la trace de la partie quadratique de la fonction U , signe d'un point d'optimum, d'autre part deux droites qui s'intersectent transversalement avec des arcs d'hyperboles remplissant les secteurs du complémentaire de ces deux droites : ces éléments de figure se retrouvent dans les figures [IV.3](#) et [IV.4](#) : ces remarques heuristiques et graphiques sont confortées par le théorème [4.3](#) et avec d'autres considérations qui ne seront pas développées ici.

▷ EXEMPLES 4.1:

4.1.1 $X = \{\mathbf{m} \in \mathbb{R}^2 \mid \|\mathbf{m}\|_2 < 1\}$ et $U(\mathbf{m}) = -\|\mathbf{m}\|_2^2$: $\operatorname{argmin}_X U = \emptyset$, $\operatorname{argmax}_X U = \{(0, 0)\}$.

4.1.2 $X = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 \leq 1\}$ et $U(x, y) = -x^2 - y^2$: $\operatorname{argmin}_X U = \{x^2 + y^2 = 1\}$ avec $U_* = -1$, $\operatorname{argmax}_X U = \{(0, 0)\}$ avec $U^* = 0$.

4.1.3 $X = \mathbb{R}^2$ et $U(x, y) = -x^2 - y^2$: $\operatorname{argmax}_X U = \{(0, 0)\}$, $\operatorname{argmin}_X U = \emptyset$ et $U_* = -\infty$, $U^* = 0$,

4.1.4 $X = \mathbb{R}^2$ et $U(x, y) = x^2 - y^2$: $\operatorname{argmin}_X U = \operatorname{argmax}_X U = \emptyset$,

4.1.5 $X = \{x \in \mathbb{R}, 0 \leq x \leq 1\}$ et $U(x, y) = (x(1-x))^2$, $U_* = 0$, $\operatorname{argmin}_X U = \{0, 1\}$, $U^* = 1/16$, $\operatorname{argmax}_X U = \{1/2\}$

4.1.6 $X = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 \leq 1\}$ et $U(x, y) = \sin x \cos y$, $\operatorname{argmin} U$ et $\operatorname{argmax} U$ sont inclus dans le bord $\{x^2 + y^2 = 1\}$: ce sont des optima de $\theta \in [0, 2\pi] \mapsto$

$\sin(\cos\theta) \cos(\sin\theta)$ ($0, \pi$ sont deux points critiques).

◁

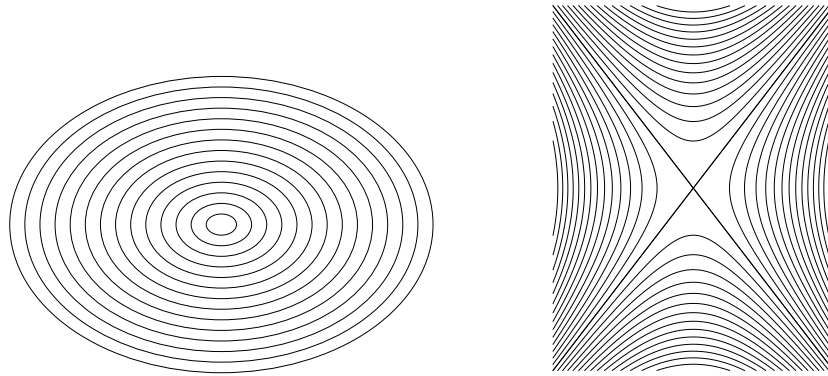


FIGURE IV.2 – Les lignes de niveau de $U(x, y) = x^2 + y^2$ et $U(x, y) = x^2 - y^2$.

Si X est une partie bornée et fermée¹ du plan \mathbb{R}^2 , une fonction U continue sur X est minorée et majorée : elle atteint ses valeurs de minimum et de maximum. Une partie n'est pas fermée bornée soit parce qu'elle est non bornée (par exemple \mathbb{R}^2 tout entier ou un demi-plan $\{(x, y) | y \geq 0\}$), soit parce qu'elle est ouverte avec un bord non vide dans \mathbb{R}^2 (par exemple le disque ouvert $\{(x, y) \in \mathbb{R}^2 | x^2 + y^2 < 1\}$ à l'opposé du disque fermé $\{(x, y) \in \mathbb{R}^2 | x^2 + y^2 \leq 1\}$). Les conditions d'optimalité ci-dessous sont applicables à des optima de fonctions définies sur des ouverts de \mathbb{R}^m .

4.1 La condition d'Euler-Fermat

Le résultat suivant est attribué à Fermat² et Euler³ a mis du temps à se formuler, à une époque où le calcul différentiel n'était pas bien formalisé comme il a commencé à l'être avec Newton⁴ et Leibniz⁵.

THÉORÈME 4.1 (Euler-Fermat): Soit Ω ouvert de \mathbb{R}^m , $U : \mathbf{x} \in \Omega \mapsto U(\mathbf{x}) \in \mathbb{R}$ une fonction différentiable et $\mathbf{x}_* \in \Omega$ un optimum local de U . Alors les dérivées partielles $\partial_{x_i} U(\mathbf{x}_*)$, $i = 1, \dots, m$ sont nulles.

1. Une partie X de \mathbb{R}^2 est fermée si elle contient son *bord*. De manière équivalente, la partie X est fermée si son complémentaire dans \mathbb{R}^2 est un ouvert. Une partie fermée bornée est un compact et réciproquement. Un résultat général attribué à Weierstraß, indique qu'une fonction continue sur un compact est bornée et atteint ses valeurs d'optimum.

2. Pierre de Fermat, \approx 1605, Beaumont-de-Lomagne, Tarn-et-Garonne – 12 janvier 1665, Castres.

3. Leonhard Euler, 15 avril 1707, Bâle Suisse – le 7 septembre 1783, Saint-Pétersbourg, Empire russe.

4. Isaac Newton, 25 décembre 1642, Lincolnshire, Angleterre – 20 mars 1727, Londres, Grande-Bretagne.

5. Gottfried Wilhelm Leibniz, 1er juillet 1646, Leipzig (électorat de Saxe) – 14 novembre 1716 Hanovre (électorat de Hanovre).

Ce théorème a diverses formulations de cette annulation, toutes équivalentes avec l'annulation des dérivées partielles : annulation du gradient $\nabla U(\mathbf{x}_*)$, annulation des dérivées directionnelles, annulation de la différentielle.

▷ EXEMPLES 4.2:

4.2.1 $x \in \mathbb{R} \mapsto U(x) = x^2$ a $x_* = 0$ comme minimum (unique, global, strict) : $U'(x_*) = 0$.

4.2.2 $x \in \mathbb{R} \mapsto U(x) = x^3$ a $x_* = 0$ comme zéro du gradient (*i. e.* sa dérivée) $U'(x) = 3x^2$, mais n'est pas un optimum.

4.2.3 $\mathbf{m} \in \mathbb{R}^2 \mapsto U(\mathbf{m}) = x^2 - y^2$ a $\mathbf{m}_* = (0, 0)$ comme zéro du gradient $\nabla U(\mathbf{m}) = (2x, -2y)$, mais n'est pas un optimum

4.2.4 $\mathbf{m} \in \mathbb{R}^2 \mapsto U(\mathbf{m}) = -x^2 - y^2$ a $\mathbf{m}_* = 0$ comme zéro du gradient $\nabla U(\mathbf{m}) = (-2x, -2y)$, c'est un maximum (global)

4.2.5 $\mathbf{m} \in \mathbb{R}^2 \mapsto U(\mathbf{m}) = x^2 + y^4$ a $(0, 0)$ comme unique zéro du gradient $\nabla U(\mathbf{m}) = (2x, 4y^3)$, qui est l'unique minimum global

4.2.6 $\mathbf{m} \in \mathbb{R}^2 \mapsto U(\mathbf{m}) = x^2 + 4xy + 4y^2 = (x + 2y)^2$ a la droite $x + 2y$ comme lieu des zéros du gradient $\nabla U(\mathbf{m}) = 2(x + 2y)(1, 4)$, qui est aussi le lieu des minima de U

4.2.7 Les exemples précédents sont implicitement avec $\Omega = \mathbb{R}^2$. Si on prend $X = \overline{B(0, 1)}$ et U définie par $U(\mathbf{m}) = x$, on a $\nabla U(\mathbf{m}) = (1, 0)$ et le gradient ∇U n'est pas nul aux points d'optimum

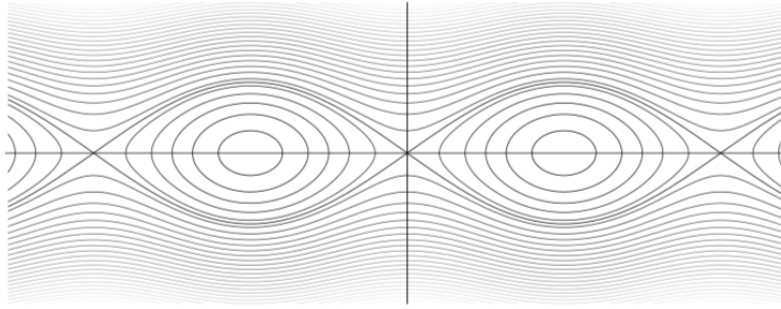
$$\begin{aligned} \operatorname{argmin}_{\mathbf{m} \in X} U(\mathbf{m}) &= (-1, 0), & \min_{\mathbf{m} \in X} U(\mathbf{m}) &= -1, \\ \operatorname{argmax}_{\mathbf{m} \in X} U(\mathbf{m}) &= (1, 0), & \max_{\mathbf{m} \in X} U(\mathbf{m}) &= 1. \end{aligned}$$

Sur la boule ouverte $B(0, 1)$, la fonction U est bornée, mais n'atteint ni sa borne inf ni sa borne sup. Ainsi cette observation n'est pas en contradiction avec le théorème 4.1.

4.2.8 Avec $X = \mathbb{R}$ et $U(x) = |x|$, l'origine est l'unique point de minimum, mais la fonction U n'est pas dérivable à l'origine, et donc pas de contradiction avec le point de minimum du théorème 4.1. ◀

DÉMONSTRATION. Soit $\alpha > 0$ suffisamment petit tel que la boule ouverte $B(\mathbf{m}_*, \alpha)$ soit incluse dans Ω . Soit \mathbf{v} un vecteur unitaire. L'application $t \in]-\alpha, +\alpha[\mapsto U(\mathbf{m}_* + t\mathbf{v}) \in \mathbb{R}$ est bien définie. Supposons que \mathbf{m}_* est un minimum local. Alors, pour $i = 1, \dots, n$, \mathbf{e}_i le i -ème vecteur de la base canonique

$$U(\mathbf{m}_* + t\mathbf{e}_i) \geq U(\mathbf{m}_*), \quad U(\mathbf{m}_* - t\mathbf{e}_i) \geq U(\mathbf{m}_*), \quad t \in]0, \alpha[$$

FIGURE IV.3 – Les lignes de niveau de la fonction $U(\mathbf{m}) = \cos x + y^2$.

et, en divisant par $t > 0$

$$\frac{U(\mathbf{m}_* + te_i) - U(\mathbf{m}_*)}{t} \geq 0, \quad \frac{U(\mathbf{m}_* - te_i) - U(\mathbf{m}_*)}{t} \geq 0, \quad t \in]0, \alpha[$$

puis faisant tendre vers 0^+ , on obtient

$$\frac{\partial U}{\partial x_i}(\mathbf{m}_*) \geq 0, \quad -\frac{\partial U}{\partial x_i}(\mathbf{m}_*) \geq 0,$$

soit

$$\frac{\partial U}{\partial x_i}(\mathbf{m}_*) \leq 0 \leq \frac{\partial U}{\partial x_i}(\mathbf{m}_*),$$

d'où il résulte la nullité de la dérivée partielle $\partial_{x_i} U(\mathbf{m}_*) = 0$, et ce pour tout $i = 1, \dots, n$. Le gradient $\nabla U(\mathbf{m}_*) = (\partial_{x_1} U(\mathbf{m}_*), \dots, \partial_{x_n} U(\mathbf{m}_*))$ est donc nul. \square

\triangle REMARQUES 4.1:

1. Il s'agit d'une condition nécessaire : les zéros du gradient sont des candidats à être des minima ou des maxima.
2. Cette condition nécessaire incite à chercher des zéros du gradient de la fonction U à minimiser : ce sont les *points critiques*⁶ au sens de la définition suivante ∇

DÉFINITION 4.2: Un point \mathbf{m}_* est dit point critique de la fonction U dérivable, s'il annule le gradient ∇U .

4.2 Optimisation d'une fonction d'une variable

La recherche de points \mathbf{m}_* de minimum (ou maximum) d'une fonction $U :]a, b[\rightarrow \mathbb{R}$ est basée sur la notion de point critique (*i. e.* $U'(\mathbf{m}_*) = 0$) et le développement de Taylor à l'ordre 2, déjà étudié au [chapitre 2](#) et que nous rappelons ici pour une fonction d'une variable.

6. L'appellation point *stationnaire* ou *d'équilibre* est aussi utilisée.

THÉORÈME 4.2 (Formule de Taylor à l'ordre 2): Soit $U :]a, b[\rightarrow \mathbb{R}$ de classe \mathcal{C}^∞ . Alors, pour $x \in]a, b[$, on a la formule dite de Taylor

$$U(\mathbf{m}_* + h) = U(\mathbf{m}_*) + U'(\mathbf{m}_*)h + \frac{U''(\mathbf{m}_*)}{2}h^2 + h^2\varepsilon(h), \quad h \text{ petit}$$

avec $\varepsilon(h) \rightarrow 0$ lorsque $h \rightarrow 0$.

Si \mathbf{m}_* est un point critique de U , alors on a

$$U(\mathbf{m}_* + h) = U(\mathbf{m}_*) + \left[\frac{U''(\mathbf{m}_*)}{2} + \varepsilon(h) \right] h^2.$$

Si $U''(\mathbf{m}_*)$ n'est pas nul, alors pour h assez petit non nul tel que $|\varepsilon(h)| < |U''(\mathbf{m}_*)|/4$, la différence $U(\mathbf{m}_* + h) - U(\mathbf{m}_*)$ est du même signe que $U''(\mathbf{m}_*)$.

1. si $U''(\mathbf{m}_*) > 0$, alors \mathbf{m}_* est un minimum local strict,
2. si $U''(\mathbf{m}_*) < 0$, alors \mathbf{m}_* est un maximum local strict,
3. si $U''(\mathbf{m}_*) = 0$, alors on ne peut rien dire et il faut approfondir l'étude, par exemple en écrivant un développement de Taylor à un ordre supérieur.

4.3 Optima : condition du second ordre en plusieurs variables

Avec les notations de Monge⁷

$$p = \partial_x U, \quad q = \partial_y U, \quad r = \partial_{x^2}^2 U, \quad s = \partial_{xy}^2 U, \quad t = \partial_{y^2}^2 U,$$

on a le gradient et la matrice hessienne

$$\nabla U = (p, q), \quad \text{Hess } U = \begin{pmatrix} r & s \\ s & t \end{pmatrix},$$

où l'évaluation des dérivées est faite implicitement (suivant le contexte) en $\mathbf{m} = (x, y)$. Le théorème suivant donne une condition suffisante d'optimum local \mathbf{m}_* pour une fonction de deux variables

THÉORÈME 4.3: Soit Ω ouvert de \mathbb{R}^2 , $U : \Omega \rightarrow \mathbb{R}$ de classe \mathcal{C}^2 et \mathbf{m}_* un point critique de U .

1. Si $rt - s^2 > 0$ et $r > 0$, alors \mathbf{m}_* est un point de minimum local strict.
2. Si $rt - s^2 > 0$ et $r < 0$, alors \mathbf{m}_* est un point de maximum local strict.
3. Si $rt - s^2 < 0$, alors \mathbf{m}_* est un point col (dit encore selle), qui n'est ni un maximum, ni un minimum.
4. Si $rt - s^2 = 0$, on ne peut rien dire.

7. Gaspard Monge, 9 mai 1746, Beaune – 28 juillet 1818, Paris.

DÉMONSTRATION. On va esquisser la démonstration, en traitant d'abord le cas où Hess $U(\mathbf{m}_*)$ est diagonale : Hess $U = \begin{pmatrix} r & 0 \\ 0 & t \end{pmatrix}$. La formule de Taylor à l'ordre 2 au point \mathbf{m}_* indique, si $\mathbf{v} = (h, k)$ est un vecteur de \mathbb{R}^2 ,

$$U(\mathbf{m}_* + \mathbf{v}) = U(\mathbf{m}_*) + \frac{rh^2 + tk^2}{2} + \|\mathbf{v}\|_2^2 \varepsilon(\mathbf{v}), \quad \mathbf{v} \in B(0, \alpha).$$

On continue comme en dimension 1. Si $s = 0$, la condition $rt - s^2 > 0$ et $r > 0$ signifient $r > 0, t > 0$. En prenant $\mu = \min(r, t)/2$ (qui est strictement positif), on a

$$U(\mathbf{m}_* + \mathbf{v}) \geq U(\mathbf{m}_*) + (h^2 + k^2)(\mu + \varepsilon(\mathbf{v})) \quad (4.1)$$

$$= U(\mathbf{m}_*) + \|\mathbf{v}\|_2^2 (\mu + \varepsilon(\mathbf{v})), \quad \mathbf{v} \in B((0, 0), \alpha). \quad (4.2)$$

Ainsi, lorsque $\|\mathbf{v}\|_2^2 = h^2 + k^2$ est assez petit pour que le terme $|\varepsilon(\mathbf{v})|$ soit inférieur à $\mu/2$ de sorte que $\mu + \varepsilon(\mathbf{v}) \geq \mu/2$ et que

$$U(\mathbf{m}_* + \mathbf{v}) \geq U(\mathbf{m}_*) + m\|\mathbf{v}\|_2^2 > U(\mathbf{m}_*) \text{ si } \mathbf{v} \neq (0, 0)$$

i. e. \mathbf{m}_* est un minimum local strict.

Par un raisonnement analogue, les conditions $s = 0, pq - s^2 > 0$ et $p < 0$ impliquent $p < 0, q < 0$ et l'examen de (4.1) montre que \mathbf{m}_* est un maximum local strict.

Si les coefficients diagonaux de la hessienne sont non nuls de signes opposés, on constate que le long des axes le point \mathbf{m}_* est un minimum ou un maximum, suivant le signe des coefficients r, t :

$$U(\mathbf{m}_* + (h, 0)) = U(\mathbf{m}_*) + \frac{h^2 r}{2} + h^2 \varepsilon(h) = U(\mathbf{m}_*) + h^2 \frac{r + 2\varepsilon(h)}{2} \geq U(\mathbf{m}_*) + h^2 r/4,$$

$$U(\mathbf{m}_* + (0, k)) = U(\mathbf{m}_*) + \frac{k^2 t}{2} + k^2 \varepsilon(k) = U(\mathbf{m}_*) + k^2 \frac{t + 2\varepsilon(k)}{2} \leq U(\mathbf{m}_*) + k^2 t/4$$

où les dernières inégalités valent pour h, k assez petits et r, s ont été supposés strictement positif et négatif resp.

Si la hessienne Hess $U(\mathbf{m}_*)$ n'est pas diagonale, on peut considérer le terme quadratique $Q(\mathbf{v})$ dans le développement de Taylor à l'ordre 2

$$U(\mathbf{m}_* + \mathbf{v}) = U(\mathbf{m}_*) + \frac{Q(\mathbf{v})}{2} + \|\mathbf{v}\|_2^2 \varepsilon(\mathbf{v}), \quad \mathbf{v} \in B(0, \alpha).$$

pour l'écrire

$$Q(\mathbf{v}) = rh^2 + 2shk + tk^2 = r \left(h + \frac{sk}{r} \right)^2 + k^2 \frac{rt - s^2}{r}, \quad (4.3)$$

terme qui est toujours > 0 sous les conditions $r > 0, rt - s^2 > 0$. En utilisant des coordonnées polaires

$$(\rho, \theta) \in \mathbb{R}^+ \times [0, 2\pi) \longmapsto (h = \rho \cos \theta, k = \rho \sin \theta) \in \mathbb{R}^2,$$

l'expression (4.3) prend la forme

$$rh^2 + 2shk + tk^2 = \rho^2 \left[r \left(\cos\theta + \sin\theta \frac{s}{r} \right)^2 + \sin^2\theta \frac{rt - s^2}{r} \right] \quad (4.4)$$

dont le second facteur dans le membre de droite est strictement positif : c'est une fonction de θ dans l'intervalle $[0, 2\pi]$, fonction continue qui est donc minorée par $\mu > 0$. Vu que $\rho^2 = \|\mathbf{v}\|_2^2$ on a donc dans (4.3) la minoration $Q(\mathbf{v}) > \mu \|\mathbf{v}\|_2^2$ et on conclut comme précédemment, via l'équation

$$U(\mathbf{m}_* + \mathbf{v}) \geq U(\mathbf{m}_*) + \|\mathbf{v}\|_2^2 (\mu + \varepsilon(\mathbf{v})),$$

au caractère de minimum pour \mathbf{m}_* .

La condition pour un maximum de U peut être établie directement, à moins de préférer le passage par $-U$ dont les minima (resp. maxima) coïncident avec les maxima (resp. minima) de $-U$.

Enfin, dans le cas d'un point selle, on montre qu'il existe des directions \mathbf{v}_\pm telles que $\pm \langle \mathbf{v}_\pm, \text{Hess } U(\mathbf{m}_*) \mathbf{v}_\pm \rangle > 0$. Si on approche \mathbf{m}_* le long d'une droite dans la direction de \mathbf{v}_\pm , le point \mathbf{m}_* sera un minimum ou maximum (le long de cette droite au voisinage de \mathbf{m}_*), si bien que \mathbf{m}_* ne sera ni un minimum ni un maximum (dans un voisinage de \mathbf{m}_*). \square

\triangle REMARQUES 4.2:

1. Si \mathbf{m}_* est un point de minimum de U , on montre que $\langle \mathbf{v}, \text{Hess } U(\mathbf{m}_*) \mathbf{v} \rangle \geq 0$ pour tout $\mathbf{v} \in \mathbb{R}^m$. Cette positivité est donc une condition nécessaire d'optimum, mais pas suffisante. La condition suffisante donnée par le théorème 4.3 pour un minimum local est l'inégalité $\langle \mathbf{v}, \text{Hess } U(\mathbf{m}_*) \mathbf{v} \rangle > 0$ pour tout vecteur \mathbf{v} non nul.
2. Le théorème 4.3 énonce une condition suffisante pour qu'un point critique soit un point de minimum (local). On montre que cette condition (valable en dimension quelconque) consiste pour un minimum en la stricte positivité de la hessienne : $\langle \text{Hess } U(\mathbf{x}_*) \mathbf{v}, \mathbf{v} \rangle > 0$ si \mathbf{v} est un vecteur non nul.
3. La condition $rt - s^2 > 0, r > 0$ peut s'exprimer $\det \text{Hess}(\mathbf{m}_*) > 0$ avec le premier coefficient de la hessienne strictement positif. ∇

\triangleright EXEMPLES 4.3:

4.3.1 Soit U la fonction définie sur \mathbb{R}^2 par $U(\mathbf{m}) = \cos x + y^2$: cette fonction U intervient dans l'étude de la dynamique du pendule. Son gradient et sa hessienne (toujours diagonale) sont

$$\nabla U(\mathbf{m}) = (-\sin x, 2y), \quad \text{Hess } U(\mathbf{m}) = \begin{pmatrix} -\cos x & 0 \\ 0 & 2 \end{pmatrix}.$$

Ses points critiques sont les $C_k = (k\pi, 0)$ avec $k \in \mathbb{Z}$: si k est impair la hessienne est de déterminant strictement positif, ainsi que les deux coefficients diagonaux, on a ainsi un minimum (local et même global, puisque les valeurs des points de minimum sont toutes égales à -1 et on a toujours $\cos x + y^2 \geq -1$); si k est pair, la hessienne est diagonale avec les deux coefficients diagonaux de signe opposé et C_k est un point selle comme annoncé dans le théorème 4.3. Les égalités

$$U(\mathbf{m}) = \begin{cases} 1 - 2 \sin^2 \left(\frac{x - k\pi}{2} \right) + y^2 & \text{si } k \text{ est pair,} \\ -1 + 2 \sin^2 \left(\frac{x - k\pi}{2} \right) + y^2 & \text{si } k \text{ est impair.} \end{cases}$$

donnent des formules locales au voisinage de C_k , confirmant le type de point critique donné par le théorème 4.3 en ce point. \triangleleft

Donnons d'autres exemples : une fonction polynôme de degré 3 et des exemples liés à la méthode des moindres carrés introduite par Gauß.

\triangleright EXEMPLES 4.4:

4.4.1 Soit $U : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par

$$U(x, y) = x^3 + y^3 - 3xy.$$

Le gradient de U et sa hessienne sont

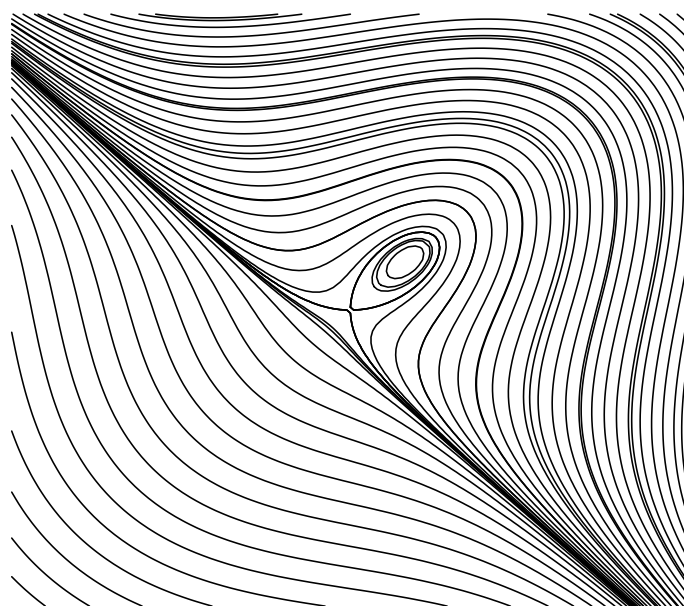


FIGURE IV.4 – Les lignes de niveaux de la fonction $U(\mathbf{m}) = x^3 + y^3 - 3xy$ sur $[-6, 6]^2$.

$$\nabla U(\mathbf{m}) = 3(x^2 - y, y^2 - x), \quad \text{Hess } U(\mathbf{m}) = 3 \begin{pmatrix} 2x & -1 \\ -1 & 2y \end{pmatrix}$$

de déterminant $\det \text{Hess } U(\mathbf{m}) = 9(4xy - 1)$. Les points critiques de U sont $(0, 0)$ et $(1, 1)$. Le premier est un point selle (le terme quadratique au voisinage de $(0, 0)$

étant $-3xy$), le second un minimum local vu que sa matrice hessienne est $3 \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$ à déterminant $pq - r^2 = 27 > 0$ et les coefficients diagonaux sont strictement positifs. Ce minimum local est non global vu que

$$U(x^3 + y^3 - 3xy)|_{x=y} = 2x^3 - 3x^2 \xrightarrow{x \rightarrow -\infty} -\infty.$$

4.4.2 La méthode des moindres carrés permet d'ajuster la valeur de paramètres caractérisant la loi régissant la réaction (exprimée via des variables y) de dispositifs expérimentaux à des données \mathbf{x} en entrées : ces observations expérimentales sont empruntées d'erreurs et soumises à un bruit. La minimisation (dite parfois *régression*) des résidus quadratiques permet d'améliorer l'adéquation de paramètres spécifiant la loi du dispositif.

Un premier exemple correspond à un nuage de valeurs encodé par un n -échantillon $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ issu de mesures indépendantes. On recherche une valeur x_* la plus proche de ce nuage en considérant le minimum du résidu (dit quadratique) $x \in \mathbb{R} \mapsto R(x, \mathbf{x}) = \sum_{j=1}^n (x - x_j)^2$. Ce point de minimum x_* doit être un point critique du résidu $R(x, \mathbf{x})$

$$0 = \nabla_x R(x_*, \mathbf{x}) = 2 \sum_{j=1}^n (x_* - x_j) = 2n \left(x_* - \frac{1}{n} \sum_{j=1}^n x_j \right),$$

soit $x_* = n^{-1} \sum_{j=1}^n x_j$, *i. e.* la moyenne $\bar{\mathbf{x}}$ des valeurs de l'échantillon \mathbf{x} . Ce point critique est un point de minimum *global*. En effet, le résidu $R(x, \mathbf{x})$ est une fonction polynomiale du second degré avec comme coefficient du monôme x^2 le scalaire n (supposé au moins égal à 1). On peut aussi exprimer ce trinôme en plaçant l'origine au point de minimum, *i. e.* en prenant la variable u telle que $x = x_* + u$

$$\begin{aligned} R(x_* + u, \mathbf{x}) &= \sum_{j=1}^n (x_* + u - x_j)^2 \\ &= \sum_{j=1}^n \left[x_*^2 + u^2 + x_j^2 + 2[x_* u - u x_j - x_* x_j] \right] \\ &= n \left[\bar{\mathbf{x}}^2 - \bar{\mathbf{x}}^2 + u^2 \right] \end{aligned}$$

où on a noté les moyennes $\bar{\mathbf{x}} = n^{-1} \sum_{j=1}^n x_j$ et $\bar{\mathbf{x}}^2 = n^{-1} \sum_{j=1}^n x_j^2$.

4.4.3 L'exemple précédent peut être généralisé pour approcher un nuage de points $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ dans \mathbb{R}^m par un point \mathbf{x}_* qui minimise le résidu quadratique

$$R(\mathbf{x}, \mathbf{X}) = \sum_{j=1}^n \|\mathbf{x} - \mathbf{x}_j\|_2^2, \quad \mathbf{x} \in \mathbb{R}^m, \mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n) \in (\mathbb{R}^m)^n,$$

Le calcul précédent peut être adapté à ce cas m -dimensionnel :

$$\begin{aligned} R(\mathbf{x}, \mathbf{X}) &= \sum_{j=1}^n \|\bar{\mathbf{X}} + \mathbf{x} - \bar{\mathbf{X}} - \mathbf{x}_j\|_2^2 \\ &= \sum_{j=1}^n \left[\|\bar{\mathbf{X}}\|_2^2 + \|\mathbf{x} - \bar{\mathbf{X}}\|_2^2 + \|\mathbf{x}_j\|_2^2 \right. \\ &\quad \left. + 2[\langle \bar{\mathbf{X}}, \mathbf{x} - \bar{\mathbf{X}} \rangle - \langle \mathbf{x} - \bar{\mathbf{X}}, \mathbf{x}_j \rangle - \langle \bar{\mathbf{X}}, \mathbf{x}_j \rangle] \right] \\ &= n \left[\|\bar{\mathbf{X}}\|_2^2 - \bar{\mathbf{X}}^2 + \|\mathbf{x} - \bar{\mathbf{X}}\|_2^2 \right] \end{aligned}$$

en ayant noté les moyennes $\bar{\mathbf{X}} = n^{-1} \sum_{j=1}^n \mathbf{x}_j$ et $\overline{\|\mathbf{X}\|_2^2} = n^{-1} \sum_{j=1}^n \|\mathbf{x}_j\|_2^2$. On peut préciser le gradient du résidu $\mathbf{x} \mapsto R(\mathbf{x}, \mathbf{X})$ en calculant la dérivée directionnelle à partir de⁸

$$\begin{aligned} R(\mathbf{x} + t\mathbf{v}, \mathbf{X}) &= \sum_{j=1}^n \|\mathbf{x} - \mathbf{x}_j + t\mathbf{v}\|_2^2 \\ &= \sum_{j=1}^n \left[\|\mathbf{x} - \mathbf{x}_j\|_2^2 + 2t\langle \mathbf{v}, \mathbf{x} - \mathbf{x}_j \rangle + t^2 \|\mathbf{v}\|_2^2 \right] \\ &= \sum_{j=1}^n \|\mathbf{x} - \mathbf{x}_j\|_2^2 + 2t\langle \mathbf{v}, \sum_{j=1}^n [\mathbf{x} - \mathbf{x}_j] \rangle + nt^2 \|\mathbf{v}\|_2^2 \\ &= \sum_{j=1}^n \|\mathbf{x} - \mathbf{x}_j\|_2^2 + 2nt\langle \mathbf{v}, \mathbf{x} - \bar{\mathbf{X}} \rangle + nt^2 \|\mathbf{v}\|_2^2 \end{aligned}$$

où a utilisé la propriété de moyenne

$$\sum_{j=1}^n [\mathbf{x} - \mathbf{x}_j] = n\mathbf{x} - \sum_{j=1}^n \mathbf{x}_j = n\mathbf{x} - n\bar{\mathbf{X}} = n[\mathbf{x} - \bar{\mathbf{X}}]$$

Ainsi

$$\nabla_{\mathbf{v}} R(\mathbf{x}, \mathbf{X}) = 2n\langle \mathbf{x} - \bar{\mathbf{X}}, \mathbf{v} \rangle,$$

indiquant que le gradient est nul en le seul $\mathbf{x} = \bar{\mathbf{X}}$, point de minimum x_* du résidu quadratique $R(\mathbf{x}, \mathbf{X})$.

4.4.4 Un autre exemple est l'ajustement des coefficients (a, b) d'une loi linéaire scalaire du type $Y = aX + b$ qui est approchée par une distribution de points proches de la droite théorique $y = ax + b$, points soumis à des perturbations aléatoires ou des erreurs d'observation. Cet ajustement est réalisé à partir d'un échantillon de quantités mesurées $\mathbf{y} = (y_1, \dots, y_n)$ correspondant à des données $\mathbf{x} = (x_1, \dots, x_n)$ en entrée d'un dispositif expérimental, par exemple l'étalonnage d'un ampèremètre dont le fonctionnement est linéaire (contrôle des variations de y relativement à celles de x , absence de phénomènes d'explosion) et est décrit par une modélisation linéaire. Vu le nombre n (en général élevé) de mesures scalaires en regard des $K = 2$ paramètres disponibles (soit a, b), cette méthode de moindres carrés apporte une solution au problème $J(a_0, b_0) = 0$ surdéterminé (et sans solution exacte) dont la solution donnerait, si elle existait, des paramètres réalisant exactement les relations $y_i = ax_i + b$: elle remplace cette équation surdéterminée par le problème de minimisation $\min_{a,b} J(a, b)$. Cette résolution dite de *régression linéaire*, est typique des problèmes d'ajustement de paramètres. Ici on examine le cas particulier du J somme de *résidus quadratiques* où la fonction J est donnée par

$$(a, b) \in \mathbb{R}^2 \mapsto J(a, b, \mathbf{x}, \mathbf{y}) = \sum_{j=1}^n (ax_j + b - y_j)^2 \in \mathbb{R}$$

qu'on notera simplement $J(a, b)$ (les vecteurs \mathbf{x}, \mathbf{y} sont des données empiriques fixes). La fonction J a pour gradient

$$\begin{aligned} \nabla_{a,b} J(a, b) &= 2 \left(\sum_{j=1}^n \mathbf{x}_j (ax_j + b - y_j) \quad \sum_{j=1}^n (ax_j + b - y_j) \right) \\ &= 2n \left(\overline{ax^2} + b\bar{x} - c_{x,y} \quad \overline{ax} + b - \bar{y} \right) \end{aligned}$$

8. On a utilisé $\|a + b\|_2^2 = \|a\|_2^2 + \|b\|_2^2 + 2\langle a, b \rangle$ avec $\langle a, b \rangle = \sum_{j=1}^m a_j b_j$ dans \mathbb{R}^m .

où on a noté

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{j=1}^n x_j, \quad \overline{\mathbf{x}^2} = \frac{1}{n} \sum_{j=1}^n x_j^2, \quad \bar{\mathbf{y}} = \frac{1}{n} \sum_{j=1}^n y_j, \quad c_{\mathbf{x},\mathbf{y}} = \frac{1}{n} \sum_{j=1}^n x_j y_j.$$

Un point critique (a_*, b_*) de J vérifie ainsi

$$a_* \overline{\mathbf{x}^2} + b_* \bar{\mathbf{x}} - c_{\mathbf{x},\mathbf{y}} = 0, \quad a_* \bar{\mathbf{x}} + b_* - \bar{\mathbf{y}} = 0,$$

une équation linéaire⁹ aisée à résoudre et avec une solution unique

$$a_* = \frac{\bar{\mathbf{x}}\bar{\mathbf{y}} - c_{\mathbf{x},\mathbf{y}}}{\overline{\mathbf{x}^2} - \bar{\mathbf{x}}^2}, \quad b_* = \bar{\mathbf{y}} - a_* \bar{\mathbf{x}}. \quad (4.5)$$

Le dénominateur $\overline{\mathbf{x}^2} - \bar{\mathbf{x}}^2$ est non nul si il y a au moins deux x_j distincts¹⁰, ce qui est expérimentalement justifié! La hessienne de J est constante égale à

$$\text{Hess } J(a, b) = 2n \begin{pmatrix} \overline{\mathbf{x}^2} & \bar{\mathbf{x}} \\ \bar{\mathbf{x}} & 1 \end{pmatrix}$$

de déterminant $4n^2(\overline{\mathbf{x}^2} - \bar{\mathbf{x}}^2)$, qui est strictement positif (d'après le cas d'égalité de Cauchy-Schwarz traité dans la note 10). Le point critique (a_*, b_*) est ainsi un point de minimum local strict.

Pour montrer que ce point de minimum est global, on introduit les variables $u = a - a_*$, $v = b - b_*$, afin d'exprimer la fonction J en les variables (u, v) de manière simple (somme d'un terme quadratique et d'un terme constant, le terme linéaire disparaissant du fait que le point origine est un point critique).

$$\begin{aligned} J(a, b) &= \sum_{j=1}^n (y_j - ax_j - b)^2 \\ &= \sum_{j=1}^n y_j^2 + a^2 \sum_{j=1}^n x_j^2 - 2a \sum_{j=1}^n x_j y_j + nb^2 - 2b \sum_{j=1}^n (y_j - ax_j) \\ &= n \left[\bar{\mathbf{y}}^2 + \overline{\mathbf{x}^2} a^2 - 2ac_{\mathbf{x},\mathbf{y}} + b^2 - 2b(\bar{\mathbf{y}} - a\bar{\mathbf{x}}) \right] \end{aligned}$$

9. Les ajustements par minimisation de résidus quadratiques amènent à résoudre des systèmes linéaires : cela justifie amplement l'usage de ces résidus quadratiques.

10. C'est une application de l'inégalité de Cauchy-Schwarz. Soient \mathbf{h}, \mathbf{k} des vecteurs de l'espace \mathbb{R}^n muni d'un produit scalaire $\langle \mathbf{h}, \mathbf{k} \rangle$ et la norme associée $\|\mathbf{h}\| = \sqrt{\langle \mathbf{h}, \mathbf{h} \rangle}$. Alors l'inégalité de Cauchy-Schwarz

$$-\|\mathbf{h}\|\|\mathbf{k}\| \leq \langle \mathbf{h}, \mathbf{k} \rangle \leq \|\mathbf{h}\|\|\mathbf{k}\|, \quad \mathbf{h}, \mathbf{k} \in \mathbb{R}^n$$

est valable, avec égalité dans la deuxième inégalité si et seulement s'il existe un scalaire λ positif tel que $\mathbf{h} = \lambda \mathbf{k}$. Ainsi, avec les vecteurs $\mathbf{h} = (h_1, \dots, h_n)$ et $(1, \dots, 1)$

$$-\sqrt{n} \sqrt{\sum_{j=1}^n h_j^2} \leq \sum_{j=1}^n h_j \leq \sqrt{n} \sqrt{\sum_{j=1}^n h_j^2},$$

ou encore

$$\left[\frac{\sum_{j=1}^n h_j}{n} \right]^2 \leq \frac{\sum_{j=1}^n h_j^2}{n},$$

soit l'inégalité $\bar{\mathbf{h}}^2 \leq \overline{\mathbf{h}^2}$ avec les notations $\bar{\mathbf{h}} = \frac{1}{n} \sum_{j=1}^n h_j$ et $\overline{\mathbf{h}^2} = \frac{1}{n} \sum_{j=1}^n h_j^2$. S'il y a égalité $\bar{\mathbf{h}}^2 = \overline{\mathbf{h}^2}$, alors \mathbf{h} est colinéaire au vecteur $(1, \dots, 1)$, donc le vecteur \mathbf{h} a toutes ses coordonnées égales.

soit en les variables (u, v)

$$\begin{aligned}
 \frac{1}{n} J(a_* + u, b_* + v) &= \bar{\mathbf{y}}^2 + \bar{\mathbf{x}}^2 (a_* + u)^2 - 2(a_* + u)c_{\mathbf{x}, \mathbf{y}} + (b_* + v)^2 \\
 &\quad - 2(b_* + v)(\bar{\mathbf{y}} - (a_* + u)\bar{\mathbf{x}}) \\
 &= \bar{\mathbf{x}}^2 u^2 + v^2 + 2\bar{\mathbf{x}}uv \\
 &\quad + 2 \left[a_* \bar{\mathbf{x}}^2 u - c_{\mathbf{x}, \mathbf{y}} u + b_* v + b_* \bar{\mathbf{x}} u - (\bar{\mathbf{y}} - a_* \bar{\mathbf{x}}) v \right] \quad (4.6) \\
 &\quad + \bar{\mathbf{y}}^2 + \bar{\mathbf{x}}^2 a_*^2 - 2a_* c_{\mathbf{x}, \mathbf{y}} + b_*^2 - 2b_* (\bar{\mathbf{y}} - a_* \bar{\mathbf{x}}) \\
 &= (\bar{\mathbf{x}}^2 - \bar{\mathbf{x}}^2) u^2 + (\bar{\mathbf{x}} u + v)^2 \\
 &\quad + \bar{\mathbf{y}}^2 + \bar{\mathbf{x}}^2 a_*^2 - 2a_* c_{\mathbf{x}, \mathbf{y}} + b_*^2 - 2b_* (\bar{\mathbf{y}} - a_* \bar{\mathbf{x}}) \\
 &= (\bar{\mathbf{x}}^2 - \bar{\mathbf{x}}^2) u^2 + (\bar{\mathbf{x}} u + v)^2 + \bar{\mathbf{y}}^2 + \bar{\mathbf{x}}^2 a_*^2 - 2a_* c_{\mathbf{x}, \mathbf{y}} - b_*^2. \quad (4.7)
 \end{aligned}$$

Vu que $(u, v) = (0, 0)$ est un point critique, la partie linéaire (4.6) est nulle : cette nullité des coefficients correspondent aux équations de point critique (4.5). La partie quadratique tend vers $+\infty$ lorsque que $\|(u, v)\|_2$ tend vers $+\infty$ vu qu'elle est minorée par $C(u^2 + v^2) = C\|(u, v)\|_2^2$ pour une constante $C > 0$. En effet, l'inégalité $x^2 + (x - y)^2 \geq y^2/2$ permet les minoration

$$x^2 + (x - y)^2 \geq \frac{x^2}{2} + \frac{x^2 + (x - y)^2}{2} \geq \frac{x^2}{2} + \frac{y^2}{4} \geq \frac{x^2 + y^2}{4}.$$

Par des changements homothétiques sur u et v on obtient la minoration du terme quadratique dans (4.7) par $C\|(u, v)\|_2^2$ pour une constante C convenable. \triangleleft

En n variables avec $n \geq 3$, on a une analyse analogue pour des conclusions assez similaires à celles du théorème 4.3, sauf que l'algèbre linéaire y est un peu plus élaborée. La formule de Taylor en un point critique \mathbf{m}_* à l'ordre 2 s'écrit

$$\begin{aligned}
 U(\mathbf{m}_* + \mathbf{v}) &= U(\mathbf{m}_*) + \langle \nabla U(\mathbf{m}_*), \mathbf{v} \rangle + \frac{\langle \mathbf{v}, \text{Hess } U(\mathbf{m}_*) \mathbf{v} \rangle}{2} + \|\mathbf{v}\|_2^2 \mathcal{E}(\mathbf{v}), \\
 &= U(\mathbf{m}_*) + \frac{\langle \mathbf{v}, \text{Hess } U(\mathbf{m}_*) \mathbf{v} \rangle}{2} + \|\mathbf{v}\|_2^2 \mathcal{E}(\mathbf{v}), \quad \mathbf{v} \in B(0, \alpha)
 \end{aligned}$$

et la condition suffisante afin que \mathbf{m}_* soit un point de minimum porte sur la hessienne et sa positivité au sens où il existe $m > 0$ tel que

$$\langle \mathbf{v}, \text{Hess } U(\mathbf{m}_*) \mathbf{v} \rangle \geq m \|\mathbf{v}\|_2^2, \quad \mathbf{v} \in \mathbb{R}^m.$$

En dimension 3, cette condition de positivité est assurée par diverses conditions, par exemple une condition sur 3 déterminants extraits de la hessienne $H = \text{Hess } U(\mathbf{m}_*) = (H_{ij})$

$$H_{11} > 0, \quad \det \begin{pmatrix} H_{11} & H_{12} \\ H_{12} & H_{22} \end{pmatrix} > 0, \quad \det \begin{pmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{pmatrix} = \det H > 0,$$