

Chapitre 15

Calcul différentiel

<https://www.jean-philippe-preaux.fr>

TABLE DES MATIÈRES

1. Dérivation de fonctions vectorielles.	1
2. Fonctions de $\mathbb{R}^n$ dans $\mathbb{R}$	6
2.1. Rappels : Fonctions de classe $\mathcal{C}^1$ de $\mathbb{R}^2$ dans $\mathbb{R}$	6
2.2. Dérivées partielles de fonctions de $\mathbb{R}^n$ dans $\mathbb{R}$	9
2.3. Fonctions de classe $\mathcal{C}^1$ de $\mathbb{R}^n$ dans $\mathbb{R}$	10
2.4. Opérations sur les fonctions de classe $\mathcal{C}^1$	14
2.5. Dérivées partielles secondes et fonctions de classe $\mathcal{C}^2$	18
2.6. Formule de Taylor d'ordre 2 et recherche d'extremum	19

Dans tout le chapitre I désigne un intervalle d'intérieur non vide, et n un entier ≥ 1 . On considère des applications à variables réelles et valeurs réelles.

Nous développons le calcul différentiel :

– Dans la première partie des applications $f : I \rightarrow \mathbb{R}^n$, c'est-à-dire, en fait, des courbes paramétrées en dimension $n \geq 1$.

– Dans la seconde partie des applications $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

Le calcul différentiel des applications de $\mathbb{R}^n \rightarrow \mathbb{R}^p$ découle sans difficulté de ces deux parties. Nous ne le traiterons cependant pas, afin d'éviter les confusions.

1. DÉRIVATION DE FONCTIONS VECTORIELLES.

DÉFINITION 1. (Dérivée en un point)

Soient $n \in \mathbb{N}^*$, I un intervalle non trivial de $\mathbb{R}$, $f : I \rightarrow \mathbb{R}^n$ et $t_0 \in I$.

On dit que f est dérivable en t_0 si et seulement si le taux d'accroissement en t_0 :

$$\begin{aligned} T_{t_0} f : I \setminus \{t_0\} &\rightarrow \mathbb{R}^n \\ t &\mapsto \frac{1}{t-t_0} \cdot (f(t) - f(t_0)) \end{aligned}$$

admet une limite lorsque $t \rightarrow t_0$. Si cette limite existe, elle est notée $f'(t_0)$ et appelée dérivée de f en t_0 :

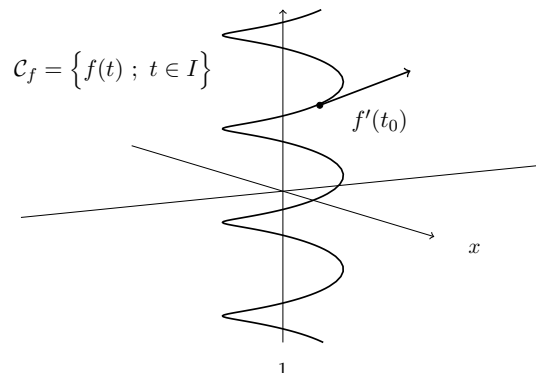
$$f'(t_0) = \lim_{t \rightarrow t_0} T_{t_0} f(t) = \lim_{t \rightarrow t_0} \frac{1}{t-t_0} \cdot (f(t) - f(t_0))$$

On dit que f est dérivable si et seulement si elle est dérivable en tout point de I .

L'application $f' = Df = \frac{df}{dx} : I \rightarrow \mathbb{R}^n$ est alors appelée dérivée de f .

On note $\mathcal{D}(I, \mathbb{R}^n)$ l'ensemble des applications dérivables de I dans $\mathbb{R}^n$.

Remarque. Si $n = 2$ ou $n = 3$, la dérivée s'interprète comme le vecteur vitesse de la courbe paramétrée $t \mapsto f(t)$. Sa norme euclidienne est la vitesse instantanée.



PROPOSITION 1. (Développement limité d'ordre 1)

Sous les mêmes hypothèses, f est dérivable en $t_0 \in I$ si et seulement si f admet un développement limité à l'ordre 1 en t_0 , c'est à dire :

$$f(t) = f(t_0) + (t - t_0) \cdot V + o_{t \rightarrow t_0}(t - t_0)$$

avec $V \in \mathbb{R}^n$ et $o_{t \rightarrow t_0}(t - t_0) = (t - t_0) \cdot h(t)$ où $h : I \rightarrow \mathbb{R}^n$ et $\lim_{t \rightarrow t_0} h(t) = 0_{\mathbb{R}^n}$.

On a alors $f'(t_0) = V$

Démonstration. On a :

$$f(t) = f(t_0) + (t - t_0) \cdot V + o(t - t_0) \iff \frac{1}{t - t_0} \cdot (f(t) - f(t_0)) = V + \frac{1}{t - t_0} \cdot o(t - t_0) \xrightarrow{t \rightarrow t_0} V$$

Remarque. Dans ce cas la courbe paramétrée admet au point $f(t_0)$ la droite tangente de vecteur directeur $V = f'(t_0)$.

On a le résultat fondamental suivant.

PROPOSITION 2. (Dérivabilité des coordonnées)

Soit :

$$\begin{aligned} f : I &\longrightarrow \mathbb{R}^n \\ t &\longmapsto (f_1(t), \dots, f_n(t)) \end{aligned}$$

f est dérivable en t_0 si et seulement si pour tout $i \in \llbracket 1, n \rrbracket$, f_i est dérivable en t_0 . On a alors :

$$f'(t_0) = (f'_1(t_0), \dots, f'_n(t_0))$$

Démonstration. En dimension finie la convergence d'une fonction équivaut à la convergence de ses fonctions coordonnées dans la base canonique :

$$\begin{aligned} T_{t_0} f(t) &= \frac{1}{t - t_0} \cdot ((f_1(t), \dots, f_n(t)) - (f_1(t_0), \dots, f_n(t_0))) \\ &= \left(\frac{f_1(t) - f_1(t_0)}{t - t_0}, \dots, \frac{f_n(t) - f_n(t_0)}{t - t_0} \right) \xrightarrow{t \rightarrow t_0} (f'_1(t_0), \dots, f'_n(t_0)) \end{aligned}$$

Remarque. En particulier, si pour tout $t \in I$, $f'(t) = 0_{\mathbb{R}^n}$ alors f est constante sur l'intervalle I , puisque toutes ses fonctions coordonnées sont des applications à valeurs réelles et de dérivée nulle.

THÉORÈME 3. (Dérivable $\implies$ continue)

Toute application $f : I \rightarrow \mathbb{R}^n$ dérivable est continue.

Démonstration. Découle de la proposition 2 et du fait qu'en dimensions finie, la continuité d'une application équivaut à celles de ses coordonnées dans la base canonique.

PROPRIÉTÉ 4. (Combinaison linéaire)

L'ensemble $\mathcal{D}(I, \mathbb{R}^n)$ des applications dérivables de I dans $\mathbb{R}^n$ est un sous-espace vectoriel de $\mathcal{C}^0(I, \mathbb{R}^n)$, et $f \mapsto f'$ est une application linéaire de $\mathcal{D}(I, \mathbb{R}^n)$ sur $\mathcal{F}(I, \mathbb{R}^n)$ c'est à dire :

si f et g sont dérivables alors pour tout $(\lambda, \mu) \in \mathbb{R}^2$, $(\lambda f + \mu g)$ est dérivable de dérivée

$$(\lambda f + \mu g)' = \lambda f' + \mu g'.$$

Démonstration. Découle de la proposition 2 et de la propriété analogue sur la dérivée scalaire.

PROPRIÉTÉ 5. (Composition à droite)

Soit $g \in \mathcal{D}(I, \mathbb{R})$, tel que $g(I) \subset J$, et $f \in \mathcal{D}(J, \mathbb{R}^n)$, alors $f \circ g \in \mathcal{D}(I, \mathbb{R}^n)$ et

$$(f \circ g)' = g' \cdot (f' \circ g)$$

Démonstration. Découle de la proposition 2 et de la propriété analogue sur la dérivée scalaire :

$$(f \circ g)' = (f_1 \circ g, \dots, f_n \circ g)' = (g' \cdot (f'_1 \circ g), \dots, g' \cdot (f'_n \circ g)) = g' \cdot (f'_1 \circ g, \dots, f'_n \circ g) = g' \cdot (f' \circ g)$$

PROPRIÉTÉ 6. (Composition à gauche par une application linéaire)

Soit $L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ et $f \in \mathcal{D}(I, \mathbb{R}^n)$; alors $L(f) : I \rightarrow \mathbb{R}^p$ est dérivable et :

$$L(f)' = L(f')$$

Démonstration. Par définition et par linéarité de L :

$$\frac{1}{t-t_0} \cdot (L(f)(t) - L(f)(t_0)) = L\left(\frac{1}{t-t_0} \cdot (f(t) - f(t_0))\right) \xrightarrow{t \rightarrow t_0} L(f'(t_0))$$

puisque une application linéaire en dimension finie est continue. ■

PROPRIÉTÉ 7. (Composition à gauche par une application bilinéaire)

Soient des entiers $n, p, q \geq 1$, une application bilinéaire $\varphi : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^q$, et deux applications dérivables définies sur $I : f \in \mathcal{D}(I, \mathbb{R}^n)$ et $g \in \mathcal{D}(I, \mathbb{R}^p)$. Alors l'application :

$$\begin{aligned} \varphi(f, g) : I &\rightarrow \mathbb{R}^q \\ x &\mapsto \varphi(f(x), g(x)) \end{aligned}$$

est dérivable sur I et :

$$(\varphi(f, g))' = \varphi(f', g) + \varphi(f, g').$$

Démonstration. (Non exigible.) L'argument ne présente pas de difficulté :

$$\begin{aligned} T_{t_0}(\varphi(f, g))(t) &= \frac{1}{t-t_0} \cdot [\varphi(f(t), g(t)) - \varphi(f(t_0), g(t_0))] \\ &= \frac{1}{t-t_0} \cdot [\varphi(f(t), g(t)) - \varphi(f(t_0), g(t)) + \varphi(f(t_0), g(t)) - \varphi(f(t_0), g(t_0))] \\ &= \frac{1}{t-t_0} \cdot [\varphi(f(t) - f(t_0), g(t)) + \varphi(f(t_0), g(t) - g(t_0))] \\ &= \varphi\left(\frac{1}{t-t_0} \cdot (f(t) - f(t_0)), g(t)\right) + \varphi\left(f(t_0), \frac{1}{t-t_0} \cdot (g(t) - g(t_0))\right) \end{aligned}$$

(par bilinéarité)

$$T_{t_0}(\varphi(f, g))(t) \xrightarrow{t \rightarrow t_0} \varphi(f'(t_0), g(t_0)) + \varphi(f(t_0), g'(t_0))$$

puisque φ est continue (car bilinéaire en dimension finie), et f, g sont dérivables en t_0 et donc aussi continues. ■

On en déduit les corollaires importants :

COROLLAIRE 8.

- *Produit d'une fonction scalaire par une fonction vectorielle :*

Si $\lambda : I \rightarrow \mathbb{R}$ et $g : I \rightarrow \mathbb{R}^n$ sont dérivables, alors $\lambda g : I \rightarrow \mathbb{R}^n$ est dérivable et :

$$(\lambda g)' = \lambda' g + \lambda g'$$

- *Produit scalaire de fonctions vectorielles :*

Si $f, g : I \rightarrow \mathbb{R}^n$ sont dérivables, alors $\langle f | g \rangle$ est dérivable et :

$$\langle f | g \rangle' = \langle f' | g \rangle + \langle f | g' \rangle$$

où $\langle \cdot | \cdot \rangle$ désigne un produit scalaire sur $\mathbb{R}^n$.

- *Déterminant de fonctions à valeurs dans $\mathbb{R}^2$:*

Si $f, g : I \rightarrow \mathbb{R}^2$, alors $\det(f, g)$ est dérivable et :

$$(\det(f, g))' = \det(f', g) + \det(f, g')$$

Démonstration. Toutes sont des composées à gauche par une application bilinéaire. ■

Exercice 1.

Soient $f : I \rightarrow \mathbb{R}^n$ dérivable. On munit $\mathbb{R}^n$ de sa structure euclidienne canonique.

1. Montrer que $(\|f\|^2)' = 2\langle f' | f \rangle$
2. Montrer que si $\forall t \in I, f(t)$ est unitaire alors $f(t) \perp f'(t)$.

Résolution.

PROPRIÉTÉ 9. (Composition à gauche par une application multilinéaire)

Soient $(n+1)$ entiers $p_1, \dots, p_n, q \geq 1$, soit une application n -linéaire $M : \mathbb{R}^{p_1} \times \dots \times \mathbb{R}^{p_n} \longrightarrow \mathbb{R}^q$, et n applications dérivables $f_1, \dots, f_n$ où pour tout $i \in \llbracket 1, n \rrbracket$, $f_i : I \longrightarrow \mathbb{R}^{p_i}$; alors l'application :

$$\begin{aligned} M(f_1, \dots, f_n) : I &\longrightarrow \mathbb{R}^q \\ x &\longmapsto M(f_1(x), \dots, f_n(x)) \end{aligned}$$

est dérivable et :

$$M(f_1, \dots, f_n)' = M(f_1', \dots, f_n) + \dots + M(f_1, \dots, f_n')$$

Démonstration. Non exigible. Procède de la même façon que pour la composition par une application bilinéaire. ■

COROLLAIRE 10. (Déterminant de fonctions à valeurs dans $\mathbb{R}^n$)

Si pour tout $i \in \llbracket 1, n \rrbracket$, $f_i : I \longrightarrow \mathbb{R}^n$ est dérivable, alors $\det(f_1, \dots, f_n)$ est dérivable et :

$$\det(f_1, \dots, f_n)' = \det(f_1', \dots, f_n) + \dots + \det(f_1, \dots, f_n')$$

Démonstration. Le déterminant de n vecteurs de $\mathbb{R}^n$ est une forme n -linéaire. ■

DÉFINITION 2. (Fonctions de classe $\mathcal{C}^p$)

- Soit $f : I \longrightarrow \mathbb{R}^n$; on définit par récurrence sur $p \in \mathbb{N}$ la dérivée p -ième $f^{(p)}$ de f en posant $f^{(0)} = f$, et, pour tout $p \in \mathbb{N}$, si $f^{(p)}$ existe et est dérivable sur I , alors $f^{(p+1)} = (f^{(p)})'$.
On peut noter aussi $f^{(p)} = D^p f = \frac{d^p f}{dx^p}$.
- L'ensemble des fonctions p fois dérivables sur I à valeurs dans $\mathbb{R}^n$ est noté $\mathcal{D}^p(I, \mathbb{R}^n)$.
- L'ensemble des fonctions p fois dérivables sur I à valeurs dans $\mathbb{R}^n$ dont la dérivée p -ième $f^{(p)}$ est continue est noté $\mathcal{C}^p(I, \mathbb{R}^n)$; une telle application est dite de classe $\mathcal{C}^p$.
- L'ensemble des fonctions p fois dérivables sur I à valeurs dans $\mathbb{R}^n$ pour tout $p \in \mathbb{N}$ est noté $\mathcal{C}^\infty(I, \mathbb{R}^n)$; une telle application est dite de classe $\mathcal{C}^\infty$.
- On a alors les inclusions $\mathcal{C}^\infty(I, \mathbb{R}^n) \subset \mathcal{C}^{p+1}(I, \mathbb{R}^n) \subset \mathcal{D}^{p+1}(I, \mathbb{R}^n) \subset \mathcal{C}^p(I, \mathbb{R}^n)$.

Remarque. On déduit aisément des propriétés précédentes des propriétés analogues pour les fonctions p fois dérivables sur I (seules les formules peuvent se révéler plus compliquées).

On pourrait poursuivre la généralisation en intégrant des fonctions vectorielles, en évoquant les formules de Taylor, etc. Démontrons à titre d'exemple l'inégalité des accroissements finis dans ce contexte.

Exercice 2. Inégalité des accroissements finis.

Soit $f \in \mathcal{C}^1(I, \mathbb{R}^n)$. On munit $\mathbb{R}^n$ de sa structure euclidienne canonique.

On suppose qu'il existe $M \in \mathbb{R}^+$ tel que $\forall t \in I, \|f'(t)\| \leq M$. On se donne $a < b$ deux réels de I .

Soit $\varphi : \begin{cases} I \longrightarrow \mathbb{R} \\ t \longmapsto \langle f(b) - f(a) \mid f(t) \rangle \end{cases}$

1. Montrer que φ est dérivable sur I et que pour tout $t \in I$, $|\varphi'(t)| \leq M \|f(b) - f(a)\|$.
2. En déduire (inégalité des accroissements finis pour la norme euclidienne)

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

Résolution.

2. FONCTIONS DE $\mathbb{R}^n$ DANS $\mathbb{R}$ 2.1. Rappels : Fonctions de classe $\mathcal{C}^1$ de $\mathbb{R}^2$ dans $\mathbb{R}$.

On munit $\mathbb{R}^3$ de sa structure euclidienne canonique, et on le représente dans l'espace muni d'un repère orthonormé direct.

Pour une fonction réelle de 2 variables réelles :

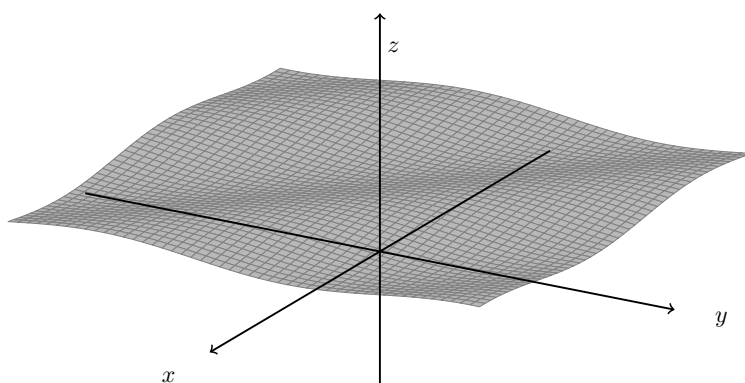
$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto f(x, y) \end{aligned}$$

Son domaine de définition $\mathcal{D}_f$ est l'ensemble des $(x, y) \in \mathbb{R}^2$ où $f(x, y)$ est bien définie.

Le graphe de f est le sous-ensemble de $\mathbb{R}^3$:

$$\mathcal{G}_f = \left\{ (x, y, z) \in \mathbb{R}^3 \mid (x, y) \in \mathcal{D}_f \text{ et } z = f(x, y) \right\}$$

sa représentation dans l'espace est la surface représentative de f .



Pour étudier f , le plus simple est de se ramener à des fonctions d'une seule variable réelle, en considérant les fonctions partielles :

Soient $f : \mathcal{D}_f \longrightarrow \mathbb{R}$ et $(x_0, y_0) \in \mathcal{D}_f$;

- La première fonction partielle en (x_0, y_0) est :

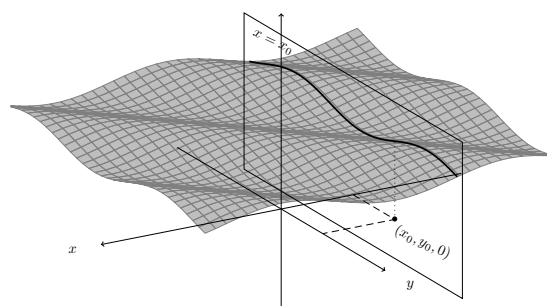
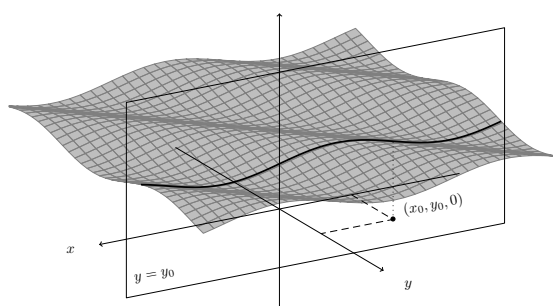
$$f_x : x \longmapsto f(x, y_0)$$

- La seconde fonction partielle en (x_0, y_0) est :

$$f_y : y \longmapsto f(x_0, y)$$

Remarque.

- La courbe représentative de la première fonction partielle en (x_0, y_0) s'obtient en intersectant la surface représentative de f avec le plan vertical $y = y_0$:
- La courbe représentative de la seconde fonction partielle en (x_0, y_0) s'obtient en intersectant la surface représentative de f avec le plan vertical $x = x_0$:
- Courbe des première et seconde fonctions partielles en (x_0, y_0) :



Soit (x_0, y_0) un point dans l'intérieur du domaine de définition $\mathcal{D}_f$ de f ;

- Si en (x_0, y_0) , la première fonction partielle est dérivable en x_0 , alors on appelle première dérivée partielle (ou dérivée partielle par rapport à x) en (x_0, y_0) le réel :

$$\frac{\partial f}{\partial x}(x_0, y_0) = f'_x(x_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}$$

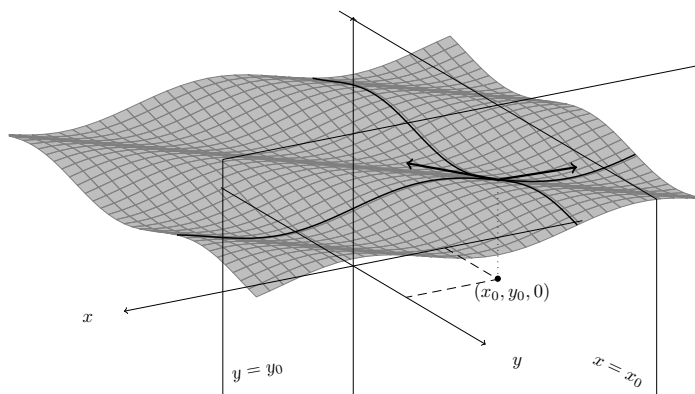
- Si en (x_0, y_0) , la seconde fonction partielle est dérivable en y_0 , alors on appelle seconde dérivée partielle (ou dérivée partielle par rapport à y) en (x_0, y_0) le réel :

$$\frac{\partial f}{\partial y}(x_0, y_0) = f'_y(y_0) = \lim_{h \rightarrow 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h}$$

On dit alors que f admet des dérivées partielles en (x_0, y_0) .

Remarque. On a représenté la surface représentative, et :

- Dans le plan $y = y_0$, la courbe de $z = f_x(x)$ et le vecteur : $(1, 0, \frac{\partial f}{\partial x}(x_0, y_0))$.
- Dans le plan $x = x_0$, la courbe de $z = f_y(y)$ et le vecteur : $(0, 1, \frac{\partial f}{\partial y}(x_0, y_0))$.



On définit en tout point (x_0, y_0) de D où f admet des dérivées partielles : la première fonction dérivée partielle :

$$\begin{aligned} \frac{\partial f}{\partial x} : D &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto \frac{\partial f}{\partial x}(x, y) \end{aligned}$$

et, la seconde fonction dérivée partielle :

$$\begin{aligned} \frac{\partial f}{\partial y} : D &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto \frac{\partial f}{\partial y}(x, y) \end{aligned}$$

Ce sont des fonctions réelles de deux variables réelles.

On dit que f est de classe $\mathcal{C}^1$ sur D si ces deux fonctions sont continues sur D .

Une application de classe $\mathcal{C}^1$ au voisinage d'un point (x_0, y_0) , admet un développement de Taylor d'ordre 1 en ce point :

Lorsque f est de classe $\mathcal{C}^1$ sur un ouvert contenant (x_0, y_0) , on a le développement de Taylor à l'ordre 1 ;

$$f(x_0 + h, y_0 + k) \underset{(0,0)}{=} f(x_0, y_0) + h \times \frac{\partial f}{\partial x}(x_0, y_0) + k \times \frac{\partial f}{\partial y}(x_0, y_0) + o(\|(h, k)\|)$$

Le développement de Taylor s'écrit plus simplement à l'aide du vecteur gradient :

Lorsque f admet des dérivées partielles en (x_0, y_0) , son vecteur gradient en (x_0, y_0) est le vecteur $\nabla f(x_0, y_0) \in \mathbb{R}^2$:

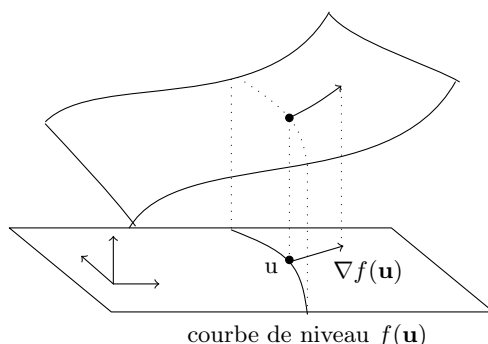
$$\nabla f(x_0, y_0) = \left(\frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0) \right)$$

Si f est de classe $\mathcal{C}^1$ sur un ouvert contenant (x_0, y_0) , son développement de Taylor s'écrit alors :

$$f(x_0 + h, y_0 + k) \underset{(0,0)}{=} f(x_0, y_0) + \langle \nabla f(x_0, y_0) | (h, k) \rangle + o(\|(h, k)\|)$$

Remarque. Le vecteur gradient est un vecteur de $\mathbb{R}^2$ qu'on a coutume de représenter sur le plan $z = 0$. Il vient rapidement que le vecteur gradient a les propriétés suivantes :

- Il donne la direction de plus grande pente sur la surface représentative.
- Il est normal à la tangente au point (x_0, y_0) de la courbe de niveau $f(x_0, y_0)$.



(Ici on a noté $\mathbf{u} = (x_0, y_0)$.)

où la courbe de niveau $f(x_0, y_0)$ est défini comme courbe représentative de :

$$\{(x, y, 0) \in \mathbb{R}^3 \mid f(x, y) = f(x_0, y_0)\}$$

- En un extremum local le gradient s'annule : par exemple pour :

$$f : (x, y) \mapsto \sqrt{1 - (x^2 + y^2)}$$

les deux dérivées partielles :

$$\frac{\partial f}{\partial x}(x, y) = -\frac{x}{\sqrt{1 - (x^2 + y^2)}}$$

$$\frac{\partial f}{\partial y}(x, y) = -\frac{y}{\sqrt{1 - (x^2 + y^2)}}$$

s'annulent en $(0, 0)$.

Mais c'est une condition nécessaire non suffisante : le gradient peut être nul en un point (x_0, y_0) sans que la fonction n'y admette un extremum ; par exemple pour :

$$g : (x, y) \mapsto x^2 - y^2$$

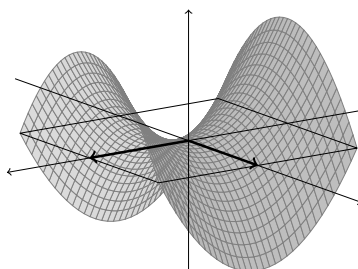
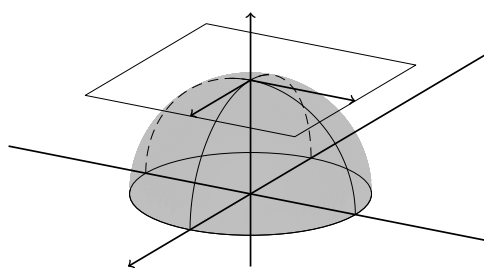
les deux dérivées partielles :

$$\frac{\partial g}{\partial x}(x, y) = 2x \quad ; \quad \frac{\partial g}{\partial y}(x, y) = -2y$$

s'annulent en $(0, 0)$ qui n'est pas un extremum car :

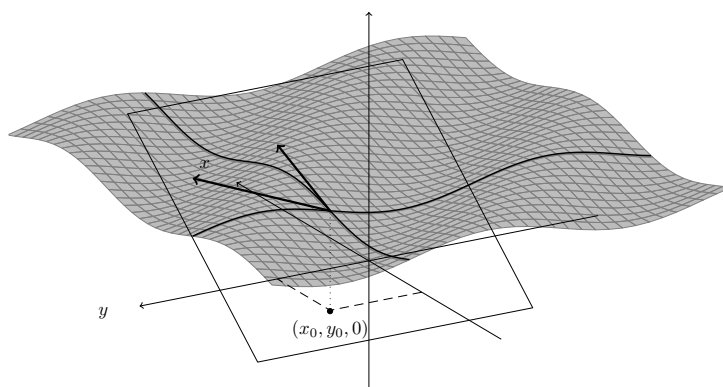
$$g(x, 0) = x^2 > g(0, 0) > -y^2 = g(0, y)$$

pour $x, y \neq 0$.



Lorsque f est de classe $\mathcal{C}^1$ sur un ouvert contenant (x_0, y_0) , sa surface représentative admet un plan tangent au point $M_0(x_0, y_0, f(x_0, y_0))$. C'est un sous-espace affine de dimension 2.

Les deux vecteurs $(1, 0, \frac{\partial f}{\partial x}(x_0, y_0))$ et $(0, 1, \frac{\partial f}{\partial y}(x_0, y_0))$ sont des vecteurs directeurs du plan tangent à la surface représentative au point M_0 .



Ainsi le plan tangent en $M_0(x_0, y_0, f(x_0, y_0))$ a pour équation :

$$z = f(x_0, y_0) + (x - x_0) \times \frac{\partial f}{\partial x}(x_0, y_0) + (y - y_0) \times \frac{\partial f}{\partial y}(x_0, y_0)$$

2.2. Dérivées partielles de fonctions de $\mathbb{R}^n$ dans $\mathbb{R}$.

On suppose $\mathbb{R}^n$ muni de sa structure euclidienne canonique ; on note $\langle . | . \rangle$ son produit scalaire et $\|.\|$ la norme associée.

DÉFINITION 3. (Dérivée selon une direction)

Soit $A \subset \mathbb{R}^n$ et $f : A \longrightarrow \mathbb{R}$, pour tout point a dans l'intérieur de A et pour tout vecteur $v \in \mathbb{R}^n$, on appelle dérivée en a de f selon le vecteur v , si elle existe :

$$D_v f(a) = \lim_{t \rightarrow 0} \frac{1}{t} \cdot (f(a + t \cdot v) - f(a))$$

c'est à dire la dérivée en 0 de la fonction de la variable réelle :

$$\begin{array}{ccc} \mathbb{R} & \longrightarrow & \mathbb{R} \\ t & \longmapsto & f(a + t \cdot v) \end{array}$$

DÉFINITION 4. (Fonctions partielles ; dérivées partielles)

Soit U un ouvert de $\mathbb{R}^n$, $f : U \longrightarrow \mathbb{R}$ et $a = (a_1, \dots, a_n) \in U$.

- La i -ème application partielle de f en a est l'application de la variable réelle :

$$t \longmapsto f(a_1, \dots, a_i + t, \dots, a_n)$$

- La i -ème dérivée partielle de f en a est, si elle existe, la dérivée en 0 de la i -ème application partielle de f en a :

$$\frac{\partial f}{\partial x_i}(a) = \lim_{t \rightarrow 0} \frac{1}{t} \cdot (f(a_1, \dots, a_i + t, \dots, a_n) - f(a_1, \dots, a_i, \dots, a_n))$$

c'est à dire la dérivée en a de f selon le i -ème vecteur e_i de la base canonique de $\mathbb{R}^n$. On la note $\frac{\partial f}{\partial x_i}(a)$ ou encore $\partial_i f(a)$.

Exemple. Lorsque f est constante, toutes ses dérivées partielles sont nulles.

Remarques.

- En notant $(e_1, \dots, e_n)$ la base canonique, la i -ème dérivée partielle en a est la dérivée en a selon le vecteur e_i :

$$\partial_i f(a) = D_{e_i} f(a)$$

- Lorsque les variables de f sont notées $x, y, z, \dots$ on note les dérivées partielles :

$$\frac{\partial f}{\partial x}(a), \quad \frac{\partial f}{\partial y}(a), \quad \frac{\partial f}{\partial z}(a)$$

- Les dérivées partielles sont des fonctions de n variables réelles à valeurs dans $\mathbb{R}$.

DÉFINITION 5. (Vecteur gradient)

Sous les mêmes hypothèses, si f admet des dérivées partielles en $a \in U$, on note :

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a) \right) \in \mathbb{R}^n$$

le vecteur gradient de f en a .

Remarque. On a alors pour tout $i \in \llbracket 1, n \rrbracket$:

$$\frac{\partial f}{\partial x_i}(a) = D_{e_i} f(a) = \langle \nabla f(a) \mid e_i \rangle$$

où $e_i = (0, \dots, 1, \dots, 0)$ désigne le i -ème vecteur de la base canonique.

Remarque. Comme déjà remarqué dans un exercice traité au chapitre "Espaces normés", une fonction peut admettre des dérivées partielles sans être continue. Par exemple :

$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto \begin{cases} \frac{xy}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{si } (x, y) = (0, 0) \end{cases} \end{aligned}$$

L'application n'est pas continue en $(0, 0)$; en effet, soit $(x_n, y_n) = \left(\frac{1}{n}, \frac{1}{n}\right) \xrightarrow{n \rightarrow +\infty} (0, 0)$ alors :

$$f(x_n, y_n) = \frac{\frac{1}{n^2}}{\frac{2}{n^2}} = \frac{1}{2} \xrightarrow{n \rightarrow +\infty} \frac{1}{2} \neq 0$$

Pourtant elle admet des dérivées partielles en $(0, 0)$:

$$\frac{f(t, 0) - f(0, 0)}{t} = 0 \xrightarrow{t \rightarrow 0} 0 = \frac{\partial f}{\partial x}(0, 0) \qquad \frac{f(0, t) - f(0, 0)}{t} = 0 \xrightarrow{t \rightarrow 0} 0 = \frac{\partial f}{\partial y}(0, 0)$$

2.3. Fonctions de classe $\mathcal{C}^1$ de $\mathbb{R}^n$ dans $\mathbb{R}$.

DÉFINITION 6. (Fonction de classe $\mathcal{C}^1$)

Sous les mêmes hypothèses, on dit que f est de classe $\mathcal{C}^1$ si f admet des dérivées partielles en tout points $a \in U$ et si toutes les fonction dérivées partielles sont continues sur U .

L'ensemble des applications de classe $\mathcal{C}^1$ de U dans $\mathbb{R}$ est noté $\mathcal{C}^1(U, \mathbb{R})$.

Exemple. Soit :

$$\begin{aligned} f : \mathbb{R}^3 &\longrightarrow \mathbb{R} \\ (x, y, z) &\longmapsto x^2 y - 5z \end{aligned}$$

f est de classe $\mathcal{C}^1$ et admet pour dérivées partielles :

$$\frac{\partial f}{\partial x}(x, y, z) = 2xy, \quad \frac{\partial f}{\partial y}(x, y, z) = x^2, \quad \frac{\partial f}{\partial z}(x, y, z) = -5$$

en effet les dérivées partielles sont continues car polynomiales.

THÉORÈME 11. (Développement de Taylor d'ordre 1)

Soit $f \in \mathcal{C}^1(U, \mathbb{R})$ et soit $a \in U$. Alors pour tout $h = (h_1, \dots, h_n) \in \mathbb{R}^n$ tel que $a + h \in U$, on a :

$$f(a + h) \underset{h \rightarrow 0_{\mathbb{R}^n}}{=} f(a) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a) \times h_i + o(\|h\|) \underset{h \rightarrow 0_{\mathbb{R}^n}}{=} f(a) + \langle \nabla f(a) \mid h \rangle + o(\|h\|)$$

Démonstration. Admis. S'obtient par exemple en appliquant n fois l'égalité des accroissements finis aux fonctions partielles. ■

DÉFINITION 7. (Différentielle de f en a)

Soit $f \in \mathcal{C}^1(U, \mathbb{R})$ et $a \in U$; on appelle différentielle de f en a , l'application :

$$\begin{aligned} df(a) : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ h = (h_1, \dots, h_n) &\longmapsto df(a) \cdot h = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a) \times h_i = \langle \nabla f(a) \mid h \rangle \end{aligned}$$

C'est une forme linéaire. En particulier la formule de Taylor à l'ordre 1 s'écrit :

$$f(a+h) \underset{h \rightarrow 0_{\mathbb{R}^n}}{=} f(a) + df(a) \cdot h + o(\|h\|)$$

Remarques.

- Lorsque $f \in \mathcal{C}^1(U, \mathbb{R})$, on peut définir aussi la différentielle de f :

$$\begin{aligned} df : U &\longrightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}) \\ a &\longmapsto df(a) \end{aligned}$$

- Lorsque f est de classe $\mathcal{C}^1$; pour $v \in \mathbb{R}^n$ et $t \in \mathbb{R}$ suffisamment proche de 0 :

$$f(a+t \cdot v) - f(a) = df(a) \cdot (t \cdot v) + o(\|t \cdot v\|) = t \cdot df(a) \cdot v + o(|t|)$$

et donc :

$$D_v f(a) = \lim_{t \rightarrow 0} \frac{1}{t} \cdot (f(a+t \cdot v) - f(a)) = df(a) \cdot v = \langle \nabla f(a) \mid v \rangle$$

Ainsi on a le résultat suivant qui interprète $df(a) \cdot v$:

PROPRIÉTÉ 12. (Lien différentielle/dérivée directionnelle)

Si $f \in \mathcal{C}^1(U, \mathbb{R})$ alors pour tout $a \in U$ et $v \in \mathbb{R}^n$, f admet une dérivée en a selon le vecteur v et :

$$D_v f(a) = df(a) \cdot v = \langle \nabla f(a) \mid v \rangle$$

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

$$\forall (x, y) \in \mathbb{R}^2, \quad f(x, y) = \begin{cases} \frac{x^3 + y^3}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{sinon} \end{cases}$$

1. On suppose f de classe $\mathcal{C}^1$; calculer $df(0, 0)$.
2. Pour $(a, b) \in \mathbb{R}^2 \setminus \{(0, 0)\}$, calculer à l'aide de la définition $D_{(a,b)} f(0, 0)$.
3. Conclure.

Résolution.

On a l'interprétation suivante du vecteur gradient :

PROPRIÉTÉ 13. (Le gradient est la direction de plus grande pente)

Soit $f \in \mathcal{C}^1(U, \mathbb{R})$ et $a \in U$; soit $v \in \mathbb{R}^n$ un vecteur unitaire, $\|v\| = 1$; alors

$$df(a) \cdot v \leq \|\nabla f(a)\|$$

et si $\nabla f(a) \neq 0_{\mathbb{R}^n}$:

$$df(a) \cdot \frac{1}{\|\nabla f(a)\|} \cdot \nabla f(a) = \|\nabla f(a)\|$$

Ainsi $\nabla f(a)$ est localement la direction de plus grande pente.

Démonstration. On applique Cauchy-Schwartz :

$$df(a) \cdot v \leq |\langle \nabla f(a) | v \rangle| \leq \|\nabla f(a)\| \times \|v\| = \|\nabla f(a)\|$$

Si $\nabla f(a) \neq 0_{\mathbb{R}^n}$:

$$df(a) \cdot \frac{1}{\|\nabla f(a)\|} \cdot \nabla f(a) = \frac{1}{\|\nabla f(a)\|} \langle \nabla f(a) | \nabla f(a) \rangle = \frac{1}{\|\nabla f(a)\|} \times \|\nabla f(a)\|^2 = \|\nabla f(a)\|$$

Remarque. Précisons que $\nabla f(a)$ est localement la direction de plus grande pente : soit $v \in \mathbb{R}^n$ le vecteur unitaire positivement colinéaire à $\nabla f(a)$ et $v \in \mathbb{R}^n$ un vecteur unitaire quelconque. On a alors pour $t \in \mathbb{R}_+$ suffisamment proche de 0^+ :

$$\begin{aligned} f(a + t \cdot \nabla) &= f(a) + t \times df(a) \cdot \nabla + o(|t|) \\ f(a + t \cdot v) &= f(a) + t \times df(a) \cdot v + o(|t|) \\ \implies f(a + t \cdot \nabla) - f(a + t \cdot v) &= t \times \underbrace{(df(a) \cdot \nabla - df(a) \cdot v)}_{\geq 0} + o(|t|) \end{aligned}$$

et donc pour t suffisamment proche de 0^+ , quelque soit la direction v , $f(a + t \cdot \nabla) \geq f(a + t \cdot v)$.

Être de classe $\mathcal{C}^1$ assure de la continuité.

COROLLAIRE 14. ($\mathcal{C}^1 \implies \mathcal{C}^0$)

Si f est de classe $\mathcal{C}^1$ alors f est continue.

Démonstration. En tout $a \in U$ et pour toute suite $h_p = (h_p^1, \dots, h_p^n) \xrightarrow{p \rightarrow +\infty} 0_{\mathbb{R}^n}$ on applique le développement limité :

$$f(a + h_p) - f(a) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a) \times h_p^i + o(\|h_p\|) \xrightarrow{p \rightarrow +\infty} 0$$

et donc $f(x) \xrightarrow{x \rightarrow a} f(a)$ d'après la caractérisation séquentielle de la limite : f est continue en a .

Remarque. Reprenons l'exemple vu plus haut d'une application qui admet des dérivées partielles sans être continue.

$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto \begin{cases} \frac{xy}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{si } (x, y) = (0, 0) \end{cases} \end{aligned}$$

ses dérivées partielles sont :

$$\frac{\partial f}{\partial x}(x, y) = \begin{cases} 0 & \text{si } (x, y) = (0, 0) \\ \frac{y^3 - x^2 y}{(x^2 + y^2)^2} & \text{si } (x, y) \neq (0, 0) \end{cases} \quad \frac{\partial f}{\partial y}(x, y) = \begin{cases} 0 & \text{si } (x, y) = (0, 0) \\ \frac{x^3 - x y^2}{(x^2 + y^2)^2} & \text{si } (x, y) \neq (0, 0) \end{cases}$$

et $\frac{\partial f}{\partial x}(0, y) = \frac{1}{y}$ n'a pas de limite quand $y \rightarrow 0$. Ainsi, comme attendu, f n'est pas de classe $\mathcal{C}^1$.

PROPRIÉTÉ 15. (Dérivation le long d'un arc de U)

Soit $f \in \mathcal{C}^1(U, \mathbb{R})$ et $\gamma : I \rightarrow U$ une application dérivable définie sur un intervalle I et à valeurs dans U . Alors :

$$\begin{aligned} f \circ \gamma : I &\longrightarrow \mathbb{R} \\ t &\longmapsto f \circ \gamma(t) \end{aligned}$$

est dérivable et pour tout $t_0 \in I$, en notant $a = \gamma(t_0)$:

$$(f \circ \gamma)'(t_0) = df(a) \cdot \gamma'(t_0)$$

De plus, si γ est de classe $\mathcal{C}^1$, alors $f \circ \gamma$ est de classe $\mathcal{C}^1$.

Remarque. C'est une dérivation le long de l'arc γ . Pour un arc rectiligne : $\gamma(t) = t \cdot v$, avec $v \in \mathbb{R}^n$, $\gamma'(t) = v$ et on retrouve la dérivée en $a = \gamma(t_0)$ selon le vecteur v :

$$(f \circ \gamma)'(t_0) = df(a) \cdot v = D_v f(a)$$

Démonstration. Soit $a = \gamma(t_0)$; notons pour tout $h \in \mathbb{R}^n$, $h = (h_1, \dots, h_n)$; notamment $\gamma(t) = (\gamma_1(t), \dots, \gamma_n(t)) \in U \subset \mathbb{R}^n$. Considérons le développement d'ordre 1 en a de f :

$$f(a + h) = f(a) + \sum_{i=1}^n h_i \times \partial_i f(a) + \|h\| \times \varepsilon(h)$$

avec $\varepsilon(h) \xrightarrow{h \rightarrow 0_{\mathbb{R}^n}} 0$. En particulier pour $h = \gamma(t) - \gamma(t_0)$:

$$\begin{aligned} f(\gamma(t)) &= f(\gamma(t_0)) + \sum_{i=1}^n (\gamma_i(t) - \gamma_i(t_0)) \times \partial_i f(a) + \|\gamma(t) - \gamma(t_0)\| \times \varepsilon(\gamma(t) - \gamma(t_0)) \\ \implies \frac{f(\gamma(t)) - f(\gamma(t_0))}{t - t_0} &= \sum_{i=1}^n \frac{\gamma_i(t) - \gamma_i(t_0)}{t - t_0} \times \partial_i f(a) + \left\| \frac{\gamma(t) - \gamma(t_0)}{t - t_0} \right\| \times \varepsilon(\gamma(t) - \gamma(t_0)) \end{aligned}$$

Puisque γ est dérivable en t_0 ,

$$\forall i \in \llbracket 1, n \rrbracket, \frac{\gamma_i(t) - \gamma_i(t_0)}{t - t_0} \xrightarrow{t \rightarrow t_0} \gamma'_i(t_0) \implies \sum_{i=1}^n \frac{\gamma_i(t) - \gamma_i(t_0)}{t - t_0} \times \partial_i f(a) \xrightarrow{t \rightarrow t_0} \sum_{i=1}^n \gamma'_i(t_0) \times \partial_i f(a)$$

et la norme étant continue :

$$\left\| \frac{\gamma(t) - \gamma(t_0)}{t - t_0} \right\| \xrightarrow{t \rightarrow t_0} \|\gamma'(t_0)\|$$

γ étant dérivable en t_0 , elle est aussi continue en t_0 , donc $\gamma(t) \xrightarrow{t \rightarrow t_0} \gamma(t_0)$ et donc :

$$\varepsilon(\gamma(t) - \gamma(t_0)) \xrightarrow{t \rightarrow t_0} 0$$

Ainsi :

$$\frac{f(\gamma(t)) - f(\gamma(t_0))}{t - t_0} \xrightarrow{t \rightarrow t_0} \sum_{i=1}^n \gamma'_i(t_0) \times \partial_i f(a) \quad \text{c'est à dire :} \quad (f \circ \gamma)'(t_0) = df(a) \cdot \gamma'(t_0)$$

Si f et γ sont de classe $\mathcal{C}^1$ toutes les fonctions γ'_i , $\partial_i f$ et γ sont continues, et donc de :

$$(f \circ \gamma)'(t_0) = \sum_{i=1}^n \gamma'_i(t_0) \times \partial_i f \circ \gamma(t_0)$$

découle que $(f \circ \gamma)'$ est continue, c'est à dire que $(f \circ \gamma)$ est de classe $\mathcal{C}^1$. ■

Exemple. Une grandeur physique qui dépend du temps et de l'espace peut se modéliser par

$$\begin{aligned} f : \mathbb{R}^3 \times \mathbb{R} &\longrightarrow \mathbb{R} \\ (x, y, z, t) &\longmapsto f(x, y, z, t) \end{aligned}$$

La position au cours du temps d'un point $M(t)$ dans l'espace est donnée par $\gamma : t \longmapsto (x(t), y(t), z(t))$. Ainsi $g : t \longmapsto f(x(t), y(t), z(t), t)$ donne l'évolution au cours du temps de la grandeur f au point $M(t)$ et au temps t . Si f est de classe $\mathcal{C}^1$ et γ est dérivable, alors g est dérivable et on peut écrire :

$$g'(t) = x'(t) \frac{\partial f}{\partial x}(\gamma(t), t) + y'(t) \frac{\partial f}{\partial y}(\gamma(t), t) + z'(t) \frac{\partial f}{\partial z}(\gamma(t), t) + \frac{\partial f}{\partial t}(\gamma(t), t)$$

ce qu'on écrit plus simplement en omettant le point où les dérivées partielles sont évaluées :

$$\frac{\partial}{\partial t} f(x(t), y(t), z(t), t) = x'(t) \frac{\partial f}{\partial x} + y'(t) \frac{\partial f}{\partial y} + z'(t) \frac{\partial f}{\partial z} + \frac{\partial f}{\partial t}$$

THÉORÈME 16. (Caractérisation des applications constantes sur un convexe)

Soit U un ouvert convexe non vide et $f : U \longrightarrow \mathbb{R}$; alors f est constante sur U si et seulement si f est de classe $\mathcal{C}^1$ et $df = 0$.

Démonstration. Si f est constante, il est clair par définition que f admet en tout point des dérivées partielles toutes nulles, et donc f est de classe $\mathcal{C}^1$ et $df = 0$.

Réciproquement, supposons f de classe $\mathcal{C}^1$ et de différentielle partout nulle; soient a, b deux points quelconques de U ouvert convexe; alors l'application $\gamma_{a,b} : t \in [0, 1] \longmapsto t \cdot a + (1 - t) \cdot b$ est à valeurs dans U . D'après la propriété 15

$$\begin{aligned} f \circ \gamma_{a,b} : [0, 1] &\longrightarrow \mathbb{R} \\ t &\longmapsto f(t \cdot a + (1 - t) \cdot b) \end{aligned}$$

est dérivable et de dérivée partout nulle. Ainsi :

$$f(a) - f(b) = \int_0^1 \frac{d}{dt} (f(t \cdot a + (1 - t) \cdot b)) dt = \int_0^1 0 dt = 0 \implies f(b) = f(a)$$

Puisque c'est vrai pour tout $(a, b) \in U^2$, f est constante. ■

Remarques.

- C'est clairement faux si U n'est qu'ouvert : par exemple la fonction f définie sur $\mathbb{R}^* \times \mathbb{R}$ par $f(x, y) = 1$ si $x > 0$ et -1 si $x < 0$ est de classe $\mathcal{C}^1$ et de différentielle partout nulle.

• Par contre on peut assouplir l'hypothèse sur U en supposant que U est ouvert et qu'entre deux points de U il existe un arc dérivable dans U qui les relie.

2.4. Opérations sur les fonctions de classe $\mathcal{C}^1$.

PROPRIÉTÉ 17. (Combinaison linéaire d'applications de classe $\mathcal{C}^1$)

Soient $f, g \in \mathcal{C}^1(U, \mathbb{R})$ et $\lambda \in \mathbb{R}$; alors $\lambda \cdot f + g \in \mathcal{C}^1(U, \mathbb{R})$.

Pour tout $(\lambda, \mu) \in \mathbb{R}^2$ et tout $a \in U$:

$$\begin{aligned}\forall i \in \llbracket 1, n \rrbracket, \quad \frac{\partial}{\partial x_i}(\lambda \cdot f + \mu \cdot g)(a) &= \lambda \cdot \frac{\partial f}{\partial x_i}(a) + \mu \cdot \frac{\partial g}{\partial x_i}(a) \\ \nabla(\lambda \cdot f + \mu \cdot g)(a) &= \lambda \cdot \nabla f(a) + \mu \cdot \nabla g(a) \\ d(\lambda \cdot f + \mu \cdot g)(a) &= \lambda \cdot df(a) + \mu \cdot dg(a)\end{aligned}$$

Démonstration. Par définition et combinaison linéaire de dérivées scalaires, $\lambda \cdot f + g$ admet des dérivées partielles en tout point de U et $\partial_i(\lambda \cdot f + g) = \lambda \cdot \partial_i f + \partial_i g$. Par combinaison linéaire d'applications continues, toutes les dérivées partielles de $\lambda \cdot f + g$ sont continues. ■

PROPRIÉTÉ 18. (Produit d'applications de classe $\mathcal{C}^1$)

Si f et g sont deux fonctions de $\mathcal{C}^1(U, \mathbb{R})$, alors leur produit fg est aussi dans $\mathcal{C}^1(U, \mathbb{R})$.

Pour tout $a \in U$:

$$\begin{aligned}\forall i \in \llbracket 1, n \rrbracket, \quad \frac{\partial}{\partial x_i}(f \times g)(a) &= \frac{\partial f}{\partial x_i}(a) \times g(a) + f(a) \times \frac{\partial g}{\partial x_i}(a) \\ \nabla(f \times g)(a) &= g(a) \cdot \nabla f(a) + f(a) \cdot \nabla g(a) \\ d(f \times g)(a) &= g(a) \cdot df(a) + f(a) \cdot dg(a)\end{aligned}$$

Démonstration. Par définition et dérivation de produits de fonctions scalaires, fg admet des dérivées partielles et :

$$\partial_i(fg) = \partial_i f \times g + f \times \partial_i g$$

or f et g étant de classe $\mathcal{C}^1$, toutes les application $\partial_i f$, $\partial_i g$, f et g sont continues, et donc $\partial_i(fg)$ est continue par somme et produit d'applications continues. ■

PROPOSITION 19. (Les polynômes sont de classe $\mathcal{C}^1$)

Si $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ est polynomiale en ses variables, alors f est de classe $\mathcal{C}^1$.

Démonstration. Notons $x_1, \dots, x_n$ les variables ; pour tout $i \in \llbracket 1, n \rrbracket$, la i -ème projection $\pi_i : (x_1, \dots, x_i, \dots, x_n) \mapsto x_i$ admet des dérivées partielles toutes continues. Par produit et combinaison linéaire, toute fonction polynôme est donc de classe $\mathcal{C}^1$. ■

PROPRIÉTÉ 20. (Composition à gauche)

Soient $f \in \mathcal{C}^1(U, \mathbb{R})$ et $\varphi : I \rightarrow \mathbb{R}$ une application de classe $\mathcal{C}^1$ où I est un intervalle d'intérieur non vide vérifiant $f(U) \subset I \subset \mathbb{R}$. Alors $\varphi \circ f : U \rightarrow \mathbb{R}$ est de classe $\mathcal{C}^1$.

Pour tout $a \in U$:

$$\begin{aligned}\forall i \in \llbracket 1, n \rrbracket, \quad \frac{\partial}{\partial x_i}(\varphi \circ f)(a) &= (\varphi' \circ f(a)) \times \frac{\partial f}{\partial x_i}(a) \\ \nabla(\varphi \circ f)(a) &= (\varphi' \circ f(a)) \cdot \nabla f(a) \\ d(\varphi \circ f)(a) &= (\varphi' \circ f(a)) \cdot df(a)\end{aligned}$$

Démonstration. On a alors en $a = (a_1, \dots, a_n) \in U$:

$$g \circ f(a) = g \circ f(a_1, \dots, a_i, \dots, a_n)$$

et pour tout $i \in \llbracket 1, n \rrbracket$:

$$t \mapsto g \circ f(a_1, \dots, a_i + t, \dots, a_n)$$

est dérivable en 0 de dérivée $g' \circ f(a) \times \frac{\partial f}{\partial x_i}(a)$ par dérivation d'une composée de fonctions scalaires ; g et f étant de classe $\mathcal{C}^1$, $a \mapsto g' \circ f(a) \times \frac{\partial f}{\partial x_i}(a)$ est continue. ■

Exercice 4. Sur quel domaine ouvert la fonction suivante est-elle de classe $\mathcal{C}^1$? Calculer ses dérivées partielles :

$$f : (x, y) \mapsto \ln(x^2 + xy + y^2)$$

Résolution.

COROLLAIRE 21. (Inverse)

Si $f \in \mathcal{C}^1(U, \mathbb{R})$ ne s'annule pas sur U , alors $1/f \in \mathcal{C}^1(U, \mathbb{R})$ et :

$$\begin{aligned}\forall i \in \llbracket 1, n \rrbracket, \quad \frac{\partial}{\partial x_i} \left(\frac{1}{f} \right) (a) &= -\frac{1}{f(a)^2} \times \frac{\partial f}{\partial x_i} (a) \\ \nabla \left(\frac{1}{f} \right) (a) &= -\frac{1}{f(a)^2} \cdot \nabla f(a) \\ d \left(\frac{1}{f} \right) (a) &= -\frac{1}{f(a)^2} \cdot df(a)\end{aligned}$$

Démonstration. Découle de la propriété précédente en prenant pour φ la fonction inverse. ■

Remarque. Notamment, les fonctions rationnelles, c'est à dire quotient de deux polynômes, sont de classe $\mathcal{C}^1$ partout où elles sont définies.

Le théorème qui suit, appelé règle de la chaîne, est très important. Il généralise la dérivation d'une composée à droite, et permet notamment de lier les dérivées partielles avant et après changement de variables.

THÉORÈME 22. (Règle de dérivation en chaîne)

Soient U un ouvert non vide de $\mathbb{R}^n$, et V un ouvert non vide de $\mathbb{R}^p$; soient deux applications :

$$\begin{array}{ccc} u : U & \longrightarrow & \mathbb{R}^p \\ (x_1, \dots, x_n) & \longmapsto & (u_1(x_1, \dots, x_n), \dots, u_p(x_1, \dots, x_n)) \end{array} \quad \begin{array}{ccc} f : V & \longrightarrow & \mathbb{R} \\ (u_1, \dots, u_p) & \longmapsto & f(u_1, \dots, u_p) \end{array}$$

On suppose que $u(U) \subset V$, de sorte que $f \circ u : U \longrightarrow \mathbb{R}$ soit définie, que les fonctions $u_1, \dots, u_p : U \longrightarrow \mathbb{R}$ soient de classe $\mathcal{C}^1$ et que f soit de classe $\mathcal{C}^1$. Alors l'application $f \circ u$:

$$\begin{array}{ccc} f \circ u : U & \longrightarrow & \mathbb{R} \\ (x_1, \dots, x_n) & \longmapsto & f(u_1(x_1, \dots, x_n), \dots, u_p(x_1, \dots, x_n)) \end{array}$$

est de classe $\mathcal{C}^1$, et en notant $b = u(a)$, on a pour tout $i \in \llbracket 1, n \rrbracket$:

$$\frac{\partial}{\partial x_i} (f \circ u)(a) = \sum_{k=1}^p \frac{\partial f}{\partial u_k}(b) \times \frac{\partial u_k}{\partial x_i}(a)$$

Démonstration. Soit $a = (a_1, \dots, a_n) \in U$ et $i \in \llbracket 1, n \rrbracket$. On considère l'application :

$$\gamma_i : t \longmapsto (u_1(a_1, \dots, t, \dots, a_n), \dots, u_p(a_1, \dots, t, \dots, a_n))$$

Puisque a est un point intérieur de U et $u(a)$ est un point intérieur de V , pour t suffisamment proche de a_i , $(a_1, \dots, t, \dots, a_n) \in U$ et $u(a_1, \dots, t, \dots, a_n) \in V$. Puisque pour tout $k \in \llbracket 1, p \rrbracket$, la fonction u_k est de classe $\mathcal{C}^1$, il découle par définition que toutes les fonctions $\gamma_{i,k} : t \longmapsto u_k(a_1, \dots, t, \dots, a_n)$ sont de classe $\mathcal{C}^1$; autrement dit la fonction γ_i est de classe $\mathcal{C}^1$. D'après la propriété 15 la fonction :

$$f \circ \gamma_i : t \longmapsto f((u_1(a_1, \dots, t, \dots, a_n), \dots, u_p(a_1, \dots, t, \dots, a_n)))$$

est dérivable en a_i , et sa dérivée, qui par définition vaut, $\partial_i(f \circ u)(a)$ est égale à :

$$(f \circ \gamma_i)'(a_i) = \partial_i(f \circ u)(a) = df(u(a)) \cdot \gamma_i'(a_i) = \sum_{k=1}^p \partial_k f(u(a)) \times \gamma_{i,k}'(a_i)$$

c'est à dire :

$$\partial_i(f \circ u)(a) = \sum_{k=1}^p \partial_k f(u(a)) \times \partial_i u_k(a_i)$$

soit encore :

$$\frac{\partial}{\partial x_i} (f \circ u)(a) = \sum_{k=1}^p \frac{\partial f}{\partial u_k}(b) \times \frac{\partial u_k}{\partial x_i}(a)$$

De plus les u_k étant toutes continues la fonction u est continue, et puisque f et les u_k sont de classe $\mathcal{C}^1$, par somme et produit de fonctions continues :

$$\partial_i(f \circ u) = \sum_{k=1}^p \partial_k f \circ u \times \partial_i u_k$$

est continue. Autrement dit $f \circ u$ est de classe $\mathcal{C}^1$. ■

Remarque. La règle de la chaîne s'écrit de manière concise en omettant les points d'évaluation :

$$\frac{\partial}{\partial x_i}(f \circ u) = \sum_{k=1}^p \frac{\partial f}{\partial u_k} \times \frac{\partial u_k}{\partial x_i}$$

Il est facile de s'en rappeler, d'autant plus si on représente les dérivées partielles sous la forme matricielle jacobienne, la formule est alors celle d'un produit matriciel :

$$\left(\frac{\partial f \circ u}{\partial x_1} \quad \dots \quad \frac{\partial f \circ u}{\partial x_i} \quad \dots \quad \frac{\partial f \circ u}{\partial x_n} \right) = \left(\frac{\partial f}{\partial u_1} \quad \dots \quad \frac{\partial f}{\partial u_k} \quad \dots \quad \frac{\partial f}{\partial u_p} \right) \times \begin{pmatrix} \frac{\partial u_1}{\partial x_1} & \dots & \frac{\partial u_1}{\partial x_i} & \dots & \frac{\partial u_1}{\partial x_n} \\ \vdots & & \vdots & & \vdots \\ \frac{\partial u_k}{\partial x_1} & \dots & \frac{\partial u_k}{\partial x_i} & \dots & \frac{\partial u_k}{\partial x_n} \\ \vdots & & \vdots & & \vdots \\ \frac{\partial u_p}{\partial x_1} & \dots & \frac{\partial u_p}{\partial x_i} & \dots & \frac{\partial u_p}{\partial x_n} \end{pmatrix}$$

(mais en omettant encore les points d'évaluation, la vraie formule avec les matrices jacobienes est

$$J(f \circ u)(a) = J(f)(u(a)) \times J(u)(a)$$

qui généralise aux fonctions de plusieurs variables la formule de dérivation d'une composée).

Mais attention aux points d'évaluation :

$$\frac{\partial}{\partial x_i}(f \circ u)(a) = \sum_{k=1}^p \frac{\partial f}{\partial u_k}(u(a)) \times \frac{\partial u_k}{\partial x_i}(a)$$

Exemples.

- Considérons les deux fonctions :

$$\begin{aligned} g : U \subset \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 & f : V \subset \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto (u(x, y), v(x, y)) & (u, v) &\longmapsto f(u, v) \end{aligned}$$

où U, V sont des ouverts de $\mathbb{R}^2$, $f \in \mathcal{C}^1(V, \mathbb{R})$, $g \in \mathcal{C}^1(U, \mathbb{R})$ et $g(U) \subset V$.

Soit $a = (x, y) \in U$; $f \circ g$ est dérivable en a , et la règle de la chaîne donne :

$$\begin{aligned} \frac{\partial f \circ g}{\partial x}(x, y) &= \frac{\partial f}{\partial u}(u(x, y), v(x, y)) \times \frac{\partial u}{\partial x}(x, y) + \frac{\partial f}{\partial v}(u(x, y), v(x, y)) \times \frac{\partial v}{\partial x}(x, y) \\ \frac{\partial f \circ g}{\partial y}(x, y) &= \frac{\partial f}{\partial u}(u(x, y), v(x, y)) \times \frac{\partial u}{\partial y}(x, y) + \frac{\partial f}{\partial v}(u(x, y), v(x, y)) \times \frac{\partial v}{\partial y}(x, y) \end{aligned}$$

ce qui s'écrit plus simplement en omettant les points d'évaluation :

$$\begin{aligned} \frac{\partial f \circ g}{\partial x} &= \frac{\partial f}{\partial u} \times \frac{\partial u}{\partial x} + \frac{\partial f}{\partial v} \times \frac{\partial v}{\partial x} \\ \frac{\partial f \circ g}{\partial y} &= \frac{\partial f}{\partial u} \times \frac{\partial u}{\partial y} + \frac{\partial f}{\partial v} \times \frac{\partial v}{\partial y} \end{aligned}$$

- Soient $(a, b, c, d) \in \mathbb{R}^4$, $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ de classe $\mathcal{C}^1$ et :

$$h(x, y) = f(ax + by, cx + dy)$$

h est de classe $\mathcal{C}^1$ sur $\mathbb{R}^2$ puisque les deux fonctions $(x, y) \longmapsto ax + by$ et $(x, y) \longmapsto cx + dy$ sont polynomiales et donc $\mathcal{C}^1$. D'après la règle de la chaîne :

$$\begin{aligned} \frac{\partial h}{\partial x}(x, y) &= a \times \frac{\partial f}{\partial x}(ax + by, cx + dy) + c \times \frac{\partial f}{\partial y}(ax + by, cx + dy) \\ \frac{\partial h}{\partial y}(x, y) &= b \times \frac{\partial f}{\partial x}(ax + by, cx + dy) + d \times \frac{\partial f}{\partial y}(ax + by, cx + dy) \end{aligned}$$

soit en omettant les points d'évaluation :

$$\frac{\partial h}{\partial x} = a \times \frac{\partial f}{\partial x} + c \times \frac{\partial f}{\partial y} \quad \frac{\partial h}{\partial y} = b \times \frac{\partial f}{\partial x} + d \times \frac{\partial f}{\partial y}$$

- Exemple à connaître : coordonnées polaires.

Soit $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ une fonction de classe $\mathcal{C}^1$ et $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ définie par :

$$g(r, \theta) = f(r \cos \theta, r \sin \theta)$$

Puisque $(r, \theta) \longmapsto r \cos \theta$ et $(r, \theta) \longmapsto r \sin \theta$ sont de classe $\mathcal{C}^1$, g est de classe $\mathcal{C}^1$ et :

$$\frac{\partial g}{\partial r} = \cos \theta \frac{\partial f}{\partial x}(x, y) + \sin \theta \frac{\partial f}{\partial y}(x, y) \quad \frac{\partial g}{\partial \theta} = -r \sin \theta \frac{\partial f}{\partial x}(x, y) + r \cos \theta \frac{\partial f}{\partial y}(x, y)$$

avec $x = r \cos \theta$ et $y = r \sin \theta$.

Exercice 5. Déterminer les fonctions $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ de classe $\mathcal{C}^1$ vérifiant :

$$\frac{\partial f}{\partial x} - \frac{\partial f}{\partial y} = 1$$

On pourra effectuer le changement de variables $u = x - y$ et $v = x + y$.

Résolution.

2.5. Dérivées partielles secondes et fonctions de classe $\mathcal{C}^2$.

Dans toute cette partie U désigne un ouvert non vide de $\mathbb{R}^n$ et f une application de U dans $\mathbb{R}$.

Les dérivées partielles de $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$, lorsqu'elles existent sont elles-mêmes des applications de U dans $\mathbb{R}$; à ce titre elles peuvent donc admettre des dérivées partielles; ce sont les dérivées partielles secondes de f .

DÉFINITION 8. (Dérivées partielles secondes)

Soit U un ouvert non vide de $\mathbb{R}^n$ et $f : U \rightarrow \mathbb{R}$ une application admettant des dérivées partielles, $\partial_j f$ pour tout $j \in \llbracket 1, n \rrbracket$. Pour tout $(i, j) \in \llbracket 1, n \rrbracket^2$, lorsqu'elle existe, la fonction $\partial_i(\partial_j f)$ est appelée dérivée partielle selon (i, j) ; on parle alors d'une dérivée partielle d'ordre 2 de f , que l'on note :

$$\partial_{i,j}^2 f \quad \text{ou} \quad \frac{\partial^2 f}{\partial x_i \partial x_j} \quad \text{ou lorsque } i = j : \quad \partial_{i,i}^2 f \quad \text{ou} \quad \frac{\partial^2 f}{\partial x_i^2}$$

Exemple. Lorsque f est constante, elle est admet des dérivées partielles d'ordre 1 toutes nulles, et donc des dérivées partielles d'ordre 2 nulles aussi.

DÉFINITION 9. (Fonctions de classe $\mathcal{C}^2$)

Avec les mêmes notations; lorsque $f : U \rightarrow \mathbb{R}$ admet des dérivées secondes selon (i, j) pour tout $(i, j) \in \llbracket 1, n \rrbracket^2$ et qu'elles sont toutes continues sur U , on dit que f est de classe $\mathcal{C}^2$. On note $\mathcal{C}^2(U, \mathbb{R})$ l'ensemble des applications de U dans $\mathbb{R}$ de classe $\mathcal{C}^2$.

Remarque. Par définition f est de classe $\mathcal{C}^2$ si et seulement si elle admet des dérivées partielles toutes de classe $\mathcal{C}^1$.

PROPRIÉTÉ 23. ($\mathcal{C}^2 \implies \mathcal{C}^1$)

Si $f : U \rightarrow \mathbb{R}$ est de classe $\mathcal{C}^2$ alors f est de classe $\mathcal{C}^1$.

Démonstration. Par définition si f est de classe $\mathcal{C}^2$ alors toutes ses dérivées d'ordre 1 sont de classe $\mathcal{C}^1$, et en particulier (corollaire 14) elles sont continues; ainsi f est de classe $\mathcal{C}^1$. ■

Remarque. Dit différemment, et avec le corollaire 14 :

$$\mathcal{C}^2(U, \mathbb{R}) \subset \mathcal{C}^1(U, \mathbb{R}) \subset \mathcal{C}^0(U, \mathbb{R}).$$

Puisqu'une dérivée partielle d'ordre 2 $\partial_{i,j}^2 f$ est elle-même une dérivée partielle d'ordre 1 $\partial_{i,j}^2 f = \partial_i \partial_j f$, et puisque lorsque f est de classe $\mathcal{C}^2$ toutes ses dérivées partielles d'ordre 1 sont de classe $\mathcal{C}^1$, toutes les opérations sur les dérivées partielles se transposent aux dérivées partielles secondes. On obtient :

PROPRIÉTÉ 24. (Combinaison linéaire d'applications de classe $\mathcal{C}^2$)

Soient $f, g \in \mathcal{C}^2(U, \mathbb{R})$ et $\lambda \in \mathbb{R}$; alors $\lambda \cdot f + g \in \mathcal{C}^2(U, \mathbb{R})$.

Pour tout $(\lambda, \mu) \in \mathbb{R}^2$ et tout $a \in U$:

$$\forall (i, j) \in \llbracket 1, n \rrbracket^2, \quad \frac{\partial^2}{\partial x_i \partial x_j} (\lambda \cdot f + \mu \cdot g)(a) = \lambda \cdot \frac{\partial^2 f}{\partial x_i \partial x_j}(a) + \mu \cdot \frac{\partial^2 g}{\partial x_i \partial x_j}(a)$$

Démonstration. Découle immédiatement de la propriété analogue à l'ordre 1 (propriété 17). ■

PROPRIÉTÉ 25. (Produit d'applications de classe $\mathcal{C}^2$)

Si f et g sont deux fonctions de $\mathcal{C}^2(U, \mathbb{R})$, alors leur produit fg est aussi dans $\mathcal{C}^2(U, \mathbb{R})$.

Pour tout $a \in U$:

$$\begin{aligned} \forall (i, j) \in \llbracket 1, n \rrbracket^2, \quad \frac{\partial^2}{\partial x_i \partial x_j} (f \times g)(a) &= \frac{\partial^2 f}{\partial x_i \partial x_j}(a) \times g(a) + \frac{\partial f}{\partial x_j}(a) \times \frac{\partial g}{\partial x_i}(a) \\ &\quad + f(a) \times \frac{\partial^2 g}{\partial x_i \partial x_j}(a) + \frac{\partial f}{\partial x_i}(a) \times \frac{\partial g}{\partial x_j}(a) \end{aligned}$$

Démonstration. Découle de la propriété analogue sur les dérivées d'ordre 1 (propriété 18). ■

PROPOSITION 26. (Les polynômes sont de classe $\mathcal{C}^2$)

Si $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ est polynomiale en ses variables, alors f est de classe $\mathcal{C}^2$.

Démonstration. Notons $x_1, \dots, x_n$ les variables; pour tout $i \in \llbracket 1, n \rrbracket$, la i -ème projection $\pi_i : (x_1, \dots, x_i, \dots, x_n) \longmapsto x_i$ admet des dérivées partielles secondes toutes continues. Par produit et combinaison linéaire, toute fonction polynôme est donc de classe $\mathcal{C}^2$. ■

PROPRIÉTÉ 27. (Composition à gauche)

Soient $f \in \mathcal{C}^2(U, \mathbb{R})$ et $\varphi : I \longrightarrow \mathbb{R}$ une application de classe $\mathcal{C}^2$ où I est un intervalle d'intérieur non vide vérifiant $f(U) \subset I \subset \mathbb{R}$. Alors $\varphi \circ f : U \longrightarrow \mathbb{R}$ est de classe $\mathcal{C}^2$.

Démonstration. Découle de la propriété analogue à l'ordre 1 (propriété 27 ainsi que des propriétés 18 et 23). ■

COROLLAIRE 28. (Inverse)

Si $f \in \mathcal{C}^2(U, \mathbb{R})$ ne s'annule pas sur U , alors $1/f \in \mathcal{C}^2(U, \mathbb{R})$.

Démonstration. Découle de la propriété précédente en prenant pour φ la fonction inverse. ■

Remarque. Notamment, les fonctions rationnelles, c'est à dire quotient de deux polynômes, sont de classe $\mathcal{C}^2$ partout où elles sont définies.

Le théorème de Schwarz est un résultat essentiel qui énonce que pour les applications de classes $\mathcal{C}^2$, l'ordre de dérivation des dérivées partielles secondes n'importe pas.

THÉORÈME 29. (De Schwarz)

Soit $f : U \longrightarrow \mathbb{R}$ une application de classe $\mathcal{C}^2$. Alors :

$$\forall (i, j) \in \llbracket 1, n \rrbracket^2, \quad \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

Démonstration. Admis. ■

DÉFINITION 10. (Matrice Hessienne)

Si f admet des dérivées partielles d'ordre 2 en $a \in U$, on appelle matrice Hessienne de f en a , la matrice de $\mathcal{M}_n(\mathbb{R})$ dont les coefficients sont $\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a) \right)_{(i,j) \in \llbracket 1, n \rrbracket^2}$:

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

Remarque. Lorsque f est de classe $\mathcal{C}^2$, d'après le théorème de Schwarz la matrice hessienne $H_f(a)$ est en tout point a symétrique réelle, et donc d'après le théorème spectral, elle est diagonalisable dans $\mathcal{M}_n(\mathbb{R})$.

2.6. Formule de Taylor d'ordre 2 et recherche d'extremum.

Soient $X, Y \in \mathcal{M}_{n,1}(\mathbb{R})$ et $M \in \mathcal{M}_n(\mathbb{R})$:

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad M = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{pmatrix}$$

Alors :

$$(X, Y) \longmapsto X^T Y = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^n x_i y_i = \langle X \mid Y \rangle$$

est le produit scalaire usuel et :

$$(X, Y) \mapsto X^T M Y = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^n \sum_{j=1}^n x_i a_{i,j} y_j$$

est une forme bilinéaire. Lorsque M est symétrique :

$$X^T M Y = (X^T M Y)^T = Y^T M^T X = Y^T M X$$

la forme bilinéaire est symétrique.

THÉORÈME 30. (Formule de Taylor à l'ordre 2)

Soit $f \in \mathcal{C}^2(U, \mathbb{R})$ et $a \in U$. Alors f admet en a le développement de Taylor d'ordre 2 : pour tout $h = (h_1, \dots, h_n) \in \mathbb{R}^n$ suffisamment proche de $0_{\mathbb{R}^n}$:

$$f(a+h) - f(a) \underset{h \rightarrow 0_{\mathbb{R}^n}}{=} df(a) \cdot (h_1, \dots, h_n) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n h_i \times \frac{\partial^2 f}{\partial x_i \partial x_j}(a) \times h_j + o(\|h\|^2)$$

où $o(\|h\|^2) = \|h\|^2 \times \varepsilon(h)$ avec $h : U \rightarrow \mathbb{R}$ et $\lim_{h \rightarrow 0_{\mathbb{R}^n}} \varepsilon(h) = 0$.

En identifiant les vecteur h et $\nabla f(a)$ avec leur matrice des coordonnées dans la base canonique, le développement de Taylor d'ordre 2 s'écrit de manière plus concise :

$$f(a+h) \underset{h \rightarrow 0_{\mathbb{R}^n}}{=} f(a) + \nabla f(a)^T h + \frac{1}{2} h^T H_f(a) h + o(\|h\|^2)$$

Démonstration. Admis. ■

Exemple. Lorsque $n = 2$:

$$\begin{aligned} f(a+h, b+k) &\underset{(0,0)}{=} f(a,b) + h \times \frac{\partial f}{\partial x}(a,b) + k \times \frac{\partial f}{\partial y}(a,b) \\ &\quad + \frac{h^2}{2} \times \frac{\partial^2 f}{\partial x^2}(a,b) + \frac{hk}{2} \times \frac{\partial^2 f}{\partial x \partial y}(a,b) + \frac{kh}{2} \times \frac{\partial^2 f}{\partial y \partial x}(a,b) + \frac{k^2}{2} \times \frac{\partial^2 f}{\partial y^2}(a,b) \\ &\quad + o(h^2 + k^2) \end{aligned}$$

et f étant de classe $\mathcal{C}^2$, avec le théorème de Schwarz :

$$\begin{aligned} f(a+h, b+k) &\underset{(0,0)}{=} f(a,b) + h \times \frac{\partial f}{\partial x}(a,b) + k \times \frac{\partial f}{\partial y}(a,b) \\ &\quad + \frac{h^2}{2} \times \frac{\partial^2 f}{\partial x^2}(a,b) + hk \times \frac{\partial^2 f}{\partial x \partial y}(a,b) + \frac{k^2}{2} \times \frac{\partial^2 f}{\partial y^2}(a,b) + o(h^2 + k^2) \end{aligned}$$

qui s'écrit aussi :

$$\begin{aligned} f(a+h, b+k) &\underset{(0,0)}{=} f(a,b) + \begin{pmatrix} \frac{\partial f}{\partial x}(a,b) & \frac{\partial f}{\partial y}(a,b) \end{pmatrix} \times \begin{pmatrix} h \\ k \end{pmatrix} \\ &\quad + \frac{1}{2} \begin{pmatrix} h & k \end{pmatrix} \times \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(a,b) & \frac{\partial^2 f}{\partial x \partial y}(a,b) \\ \frac{\partial^2 f}{\partial y \partial x}(a,b) & \frac{\partial^2 f}{\partial y^2}(a,b) \end{pmatrix} \times \begin{pmatrix} h \\ k \end{pmatrix} + o(h^2 + k^2) \end{aligned}$$

Exercice 6. Donner le développement de Taylor d'ordre 2 au voisinage de (a, b) de :

$$f : (x, y) \mapsto \sin(xy)$$

Résolution.

La formule de Taylor au second ordre s'avère très utile pour déterminer les extremums locaux de f .

DÉFINITION 11. (Extremum local/global)

Soit $f : U \longrightarrow \mathbb{R}$; un point $a \in U$ est un :

- *extremum (global) de f si :*
 - Pour tout $x \in U$, $f(x) \geq f(a)$; dans ce cas a est un *minimum (global)*, ou
 - Pour tout $x \in U$, $f(x) \leq f(a)$; dans ce cas a est un *maximum (global)*.
- *extremum local de f si il existe une boule ouverte $B(a, r)$ centrée en a tel que :*
 - Pour tout $x \in U \cap B(a, r)$, $f(x) \geq f(a)$; dans ce cas a est un *minimum local*, ou
 - Pour tout $x \in U \cap B(a, r)$, $f(x) \leq f(a)$; dans ce cas a est un *maximum local*.

Remarque. Bien sur tout extremum global est aussi un extremum local. La réciproque est fausse ; par exemple $x \mapsto x + \cos(x)$ admet un extremum local en tout point $x \equiv \frac{\pi}{2}[\pi]$ et aucun extremum global.

THÉORÈME 31. (Un extremum est un point critique)

Si $f \in \mathcal{C}^1(U, \mathbb{R})$ admet un extremum en a , alors en a toutes les dérivées partielles d'ordre 1 s'annulent :

$$a \text{ extremum local} \implies \nabla f(a) = 0$$

Un tel point a où les dérivées partielles d'ordre 1 s'annulent s'appelle un point critique.

Remarque. Pour plus de lisibilité on note $0 = 0_{\mathbb{R}^n}$.

Démonstration. Puisque U est un ouvert, il existe $r > 0$ tel que $B(a, r) \subset U$. Soit $(e_1, \dots, e_n)$ la base canonique de $\mathbb{R}^n$; si a est un extremum local alors toutes les fonctions de la variable réelle :

$$\varphi_i : t \mapsto f(a + t \cdot e_i)$$

admettent un extremum local en 0. Or les applications φ_i sont définies sur l'intervalle ouvert $] -r, r[$ contenant 0 et dérivables en 0, de dérivée $\varphi'_i(0) = \partial_i f(a)$; or une application de la variable réelle définie sur un intervalle ouvert et admettant un extremum local en un point où elle est dérivable, a une dérivée qui s'annule en ce point. Ainsi :

$$\forall i \in \llbracket 1, n \rrbracket, \varphi'_i(0) = \frac{\partial f}{\partial x_i}(a) = 0$$

■

C'est une condition nécessaire non suffisante (par exemple 0 est un point critique de $x \mapsto x^3$ mais n'est pas un extremum local).

Une propriété très intéressante pour s'assurer de l'existence d'un minimum local est la suivante :

Exercice 7.

1) Soit : $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une application continue ; montrer que si :

$$\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty \quad (\text{on dit que } f \text{ est coercive})$$

alors f admet un minimum global.

2) Montrer que :

$$f(x, y) = \frac{e^{2(x^2+y^2)}}{1 + (x+y)^2}$$

admet un minimum global et le déterminer.

Résolution.

On peut raffiner le résultat à l'ordre 2 en une condition nécessaire plus fine, ainsi qu'en une condition suffisante.

Rappelons qu'une matrice symétrique réelle $M \in \mathcal{M}_n(\mathbb{R})$ est dite :

- positive si $\forall X \in \mathcal{M}_{n,1}(\mathbb{R}), X^T M X \geq 0$; c'est le cas précisément lorsque toutes les valeurs propres de M sont positives. On note $\mathcal{S}_n^+(\mathbb{R})$ l'ensemble des matrices symétriques réelles positives.
- définie positive si $\forall X \in \mathcal{M}_{n,1}(\mathbb{R}) \setminus \{O_{n,1}\}, X^T M X > 0$; c'est le cas précisément lorsque toutes les valeurs propres de M sont strictement positives. On note $\mathcal{S}_n^{++}(\mathbb{R})$ l'ensemble des matrices symétriques réelles définies positives.

Notons encore qu'une matrice symétrique réelle $M \in \mathcal{M}_n(\mathbb{R})$:

- négative si $\forall X \in \mathcal{M}_{n,1}(\mathbb{R}), X^T M X \leq 0$; c'est le cas précisément lorsque toutes les valeurs propres de M sont négatives. On note $\mathcal{S}_n^-(\mathbb{R})$ l'ensemble des matrices symétriques réelles négatives.
- définie négative si $\forall X \in \mathcal{M}_{n,1}(\mathbb{R}) \setminus \{O_{n,1}\}, X^T M X < 0$; c'est le cas précisément lorsque toutes les valeurs propres de M sont strictement négatives. On note $\mathcal{S}_n^{--}(\mathbb{R})$ l'ensemble des matrices symétriques réelles définies négatives.

THÉORÈME 32. (Condition nécessaire/suffisante en un extremum local)

Soit $f \in \mathcal{C}^2(U, \mathbb{R})$ et $a \in U$.

- (Condition nécessaire pour que a soit un extremum local)
Si a est un minimum (resp. maximum) local alors $\nabla f(a) = 0$ et $H_f(a)$ est positive (resp. négative).
- (Condition suffisante pour que a soit un extremum local)
Si $\nabla f(a) = 0$ et $H_f(a)$ est définie positive (resp. définie négative) alors a est un minimum (resp. maximum) local.

Démonstration. Si a est un extremum local, alors a est un point critique $\nabla f(a) = 0$, et le développement de Taylor à l'ordre 2 devient :

$$f(a+h) - f(a) = \frac{1}{2} h^T H_f(a) h + o(\|h\|^2)$$

Par définition il existe $r_1 > 0$ tel que pour tout $h \in U \cap B(a, r_1)$, $f(a+h) - f(a)$ garde un signe constant ; ≥ 0 en un minimum, ≤ 0 en un maximum. Puisque U est un ouvert, a est un point intérieur et donc il existe $r_2 > 0$ tel que $B(a, r_2) \subset U$. Ainsi, en posant $r = \min(r_1, r_2) > 0$, $B(a, r) \subset U$ et pour tout $h \in B(a, r)$, $f(a+h) - f(a)$ est défini et de signe constant. Fixons une direction $d \in \mathbb{R}^n \setminus \{0_{\mathbb{R}^n}\}$ et appliquons le développement de Taylor pour $h = t \cdot d$ pour t un réel proche de 0 :

$$f(a+t \cdot d) - f(a) = t^2 \left(\frac{1}{2} d^T H_f(a) d + \|d\|^2 \times \varepsilon(t) \right)$$

avec $\varepsilon(t) \xrightarrow[t \rightarrow 0]{} 0$. Le fait que a soit un minimum (resp. maximum) local nous assure alors que le membre de gauche est ≥ 0 (resp. ≤ 0) pour tout t vérifiant $|t| < \frac{r}{\|d\|}$. Ainsi :

$$0 < |t| < \frac{r}{\|d\|} \implies \frac{1}{2} d^T H_f(a) d + \|d\|^2 \times \varepsilon(t) \text{ garde le même signe constant } (\geq 0 \text{ en un min, } \leq 0 \text{ en un max})$$

et donc par passage à la limite lorsque $t \rightarrow 0$:

$$d^T H_f(a) d \text{ est de même signe } (\geq 0 \text{ en un min, } \leq 0 \text{ en un max})$$

Puisque c'est vrai pour tout vecteur $d \in \mathbb{R}^n$ non nul, et trivialement aussi pour $d = 0_{\mathbb{R}^n}$, on en déduit que $H_f(a)$ est positive en a minimum local et négative en a maximum local.

Si $\nabla f(a) = 0$ et $H_f(a)$ est définie positive. L'application $d \mapsto d^T H_f(a) d$ est une forme bilinéaire entre $\mathbb{R}$ -ev de dimensions finies, elle est donc continue. D'après le théorème des bornes atteintes (cf. Chapitre "Espaces vectoriels normés"), sur le fermé borné $S(0_{\mathbb{R}^n}, 1)$ elle est bornée et atteint ses bornes. Ainsi il existe $m > 0$ tel que pour tout $d \in \mathbb{R}^n$, $\|d\| = 1 \implies d^T H_f(a) d \geq m > 0$. Soit $h \in \mathbb{R}^n \setminus \{0_{\mathbb{R}^n}\}$; notons $d_h = \frac{1}{\|h\|} \cdot h \in S(0_{\mathbb{R}^n}, 1)$; ainsi $h = \|h\| \cdot d_h$. Alors si h est suffisamment proche de $0_{\mathbb{R}^n}$:

$$\begin{aligned} f(a+h) - f(a) &= \frac{1}{2} h^T H_f(a) h + o(\|h\|^2) \\ &= \|h\|^2 \left(\underbrace{\frac{1}{2} d_h^T H_f(a) d_h}_{\geq m > 0} + \varepsilon(h) \right) \quad \text{avec } \varepsilon(h) \xrightarrow[h \rightarrow 0_{\mathbb{R}^n}]{} 0 \end{aligned}$$

Puisque $\varepsilon(h) \xrightarrow[h \rightarrow 0_{\mathbb{R}^n}]{} 0$, par définition :

$$\exists r > 0, \forall h \in \mathbb{R}^n, \|h\| \leq r \implies |\varepsilon(h)| \leq \frac{m}{4}$$

et donc pour ce $r > 0$:

$$a+h \in U \text{ et } \|h\| \leq r \implies f(a+h) - f(a) \geq \frac{m\|h\|^2}{4} > 0$$

Ainsi il existe bien une boule centrée en a , $B(a, r)$, tel que $\forall x \in U \cap B(a, r)$:

$$f(x) \geq f(a)$$

(il suffit de poser $x = a+h$). Donc a est bien un minimum local.

Le cas où $\nabla f(a) = 0$ et $H_f(a)$ est définie négative, est similaire : en changeant : $\|d\| = 1 \implies d^T H_f(a) d \leq m < 0$ on obtient :

$$a+h \in U \text{ et } \|h\| \leq r \implies f(a+h) - f(a) \leq \frac{m\|h\|^2}{4} < 0$$

et a est un maximum local. ■

Méthode. Il s'agit en un point critique a d'une fonction f de classe $\mathcal{C}^2$:

- d'une condition nécessaire pour que a soit un extremum : $H_f(a) \in \mathcal{S}_n^+(\mathbb{R}) \cup \mathcal{S}_n^-(\mathbb{R})$, ainsi si $H_f(a) \notin \mathcal{S}_n^+(\mathbb{R}) \cup \mathcal{S}_n^-(\mathbb{R})$ alors a n'est pas un extremum local,
- d'une condition suffisante pour que a soit un extremum : $H_f(a) \in \mathcal{S}_n^{++}(\mathbb{R}) \cup \mathcal{S}_n^{--}(\mathbb{R})$, ainsi si $H_f(a) \in \mathcal{S}_n^{++}(\mathbb{R}) \cup \mathcal{S}_n^{--}(\mathbb{R})$ alors a est un extremum local.

Mais ce n'est pas une condition nécessaire et suffisante : si en a point critique, $H_f(a) \in \mathcal{S}_n^+(\mathbb{R}) \cup \mathcal{S}_n^-(\mathbb{R})$, on ne peut rien conclure ; c'est le cas d'indétermination.

Dans le cas particulier d'une fonction de classe $\mathcal{C}^2$ de $U \subset \mathbb{R}^2 \longrightarrow \mathbb{R}$, le théorème devient :

COROLLAIRE 33. (Cas particulier des fonctions de deux variables réelles)

Soit $f : U \subset \mathbb{R}^2 \longrightarrow \mathbb{R}$ une fonction de classe $\mathcal{C}^2$ et soit a un point critique de f (i.e. $\nabla f(a) = 0$) ; Notons la matrice Hessienne de f en a :

$$H = H_f(a) = \begin{pmatrix} r & s \\ s & t \end{pmatrix} \in \mathcal{S}_2(\mathbb{R}) \subset \mathcal{M}_2(\mathbb{R})$$

Alors :

- Si a est un extremum local, alors $\det(H) = rt - s^2 \geq 0$, et :
 - si a est un minimum local, alors $\text{tr}(H) = r + t \geq 0$,
 - si a est un maximum local, alors $\text{tr}(H) = r + t \leq 0$.
- Si $\det(H) = rt - s^2 > 0$, alors a est un extremum local, et :
 - si $\text{tr}(H) = r + t > 0$, alors a est un minimum local,
 - si $\text{tr}(H) = r + t < 0$, alors a est un maximum local.

Remarque. Les notations r, s, t sont appelées les notations de Monge.

Démonstration. Puisque la matrice Hessienne H est (d'après le théorème de Schwarz) symétrique réelle, elle est d'après le théorème spectral diagonalisable et donc :

- (resp. définie) positive ssi ses valeurs propres sont (resp. strictement) positives,
- (resp. définie) négative ssi ses valeurs propres sont (resp. strictement) négatives.

Mais le déterminant de H est le produit des valeurs propres, et sa trace est la somme des valeurs propres. Puisque $H \in \mathcal{M}_2(\mathbb{R})$, les valeurs propres sont de même signe ssi $\det(H) \geq 0$ et de plus non nulles ssi $\det(H) > 0$, et leur signe est alors celui de $\text{tr}(H)$. Le résultat découle alors du théorème précédent. ■

Exercice 8.

1) Déterminer tous les extremums locaux de :

$$f(x, y) = x^2 + xy + y^2 + \frac{x^3}{4}$$

2) Même question avec :

$$g(x, y) = x^4 + y^4 - 2(x - y)^2$$

Résolution.

