Chapitre 7

Calcul différentiel à plusieurs variables

7.1 L'espace $\mathbb{R}^n$

7.1.1 Structures sur $\mathbb{R}^n$

– L'ensemble $\mathbb{R}^n$ est l'ensemble des n-uplets $(x_1, x_2, \ldots, x_n)$ où $x_i \in \mathbb{R}$. On notera

$$\mathrm{x} = (x_1, x_2, \ldots, x_n).$$

– x est un **point** de $\mathbb{R}^n$.

– Le point $\mathrm{x} = (x_1,\ x_2,\ \ldots,\ x_n)$ sera aussi considéré comme un vecteur-colonne

$$\mathrm{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

– Les x_i sont les **composantes** de x.

Notation :
Dans $\mathbb{R}^2$ on notera (x, y) plutôt que (x_1, x_2)
Dans $\mathbb{R}^3$ on notera (x, y, z) au lieu de (x_1, x_2, x_3).

L'espace $\mathbb{R}^n$ a les propriétés suivantes :

• L'espace $\mathbb{R}^n$ est muni d'une **addition** :

$$(x_1,\ x_2,\ \ldots,\ x_n) + (y_1,\ y_2,\ \ldots,\ y_n) = (x_1 + y_1,\ x_2 + y_2,\ \ldots,\ x_n + y_n)$$

et d'une **multiplication par un scalaire** :

$$\lambda \cdot (x_1,\ \ldots,\ x_n) = (\lambda x_1,\ \ldots,\ \lambda x_n).$$

Ces 2 opérations font de $\mathbb{R}^n$ un espace vectoriel.
L'élément neutre pour l'addition est l'origine : $O(0, 0, \ldots, 0)$.

- L'espace $\mathbb{R}^n$ est muni d'un **produit scalaire** :

$$< \cdot\,, \cdot > \,:\, \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R} \qquad \text{défini par}$$

$$< \mathrm{x}, \mathrm{y} > = \sum_{i=1}^{n} x_i y_i = \mathrm{x}^T \cdot \mathrm{y} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

- L'espace $\mathbb{R}^n$ est muni d'une **norme** :

$$\| \cdot \| : \mathbb{R}^n \longrightarrow \mathbb{R} \qquad \text{définie par}$$

$$\|\mathrm{x}\| = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{< \mathrm{x}, \mathrm{x} >}.$$

- Il est muni d'une **métrique** (ou **distance**) :

$$d : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R} \qquad \text{définie par}$$

$$d(\mathrm{x}, \mathrm{y}) = \|\mathrm{x} - \mathrm{y}\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

qui satisfait les propriétés suivantes :

(i) $d(\mathrm{x}, \mathrm{y}) \geqslant 0$

(ii) $d(\mathrm{x}, \mathrm{y}) = 0 \quad \Leftrightarrow \quad \mathrm{x} = \mathrm{y}$

(iii) $d(\mathrm{x}, \mathrm{y}) = d(\mathrm{y}, \mathrm{x})$ (symétrie)

(iv) $d(\mathrm{x}, \mathrm{z}) \leqslant d(\mathrm{x}, \mathrm{y}) + d(\mathrm{y}, \mathrm{z})$ (inégalité du triangle)

pour tout x, y, $\mathrm{z} \in \mathbb{R}^n$.

- Les vecteurs

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad e_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

forment une **base orthonormale** de $\mathbb{R}^n$.

Remarque 7.1. Si $n = 1$, on a $\|x\| = |x|$ et $d(x, y) = |x - y|$ pour tout $x, y \in \mathbb{R}$.

Dans ce chapitre, on va considérer les fonctions

$$f : D \longrightarrow \mathbb{R}^m$$

avec $D \subset \mathbb{R}^n$. A un point $\mathrm{x} \in D$ on associe donc un vecteur $f(\mathrm{x}) \in \mathbb{R}^m$.

Exemples 7.2.

1. $f : \mathbb{R}^2 \longrightarrow \mathbb{R} : f(x, y) = x^2 + \sin y$

2. $f : \mathbb{R} \longrightarrow \mathbb{R}^3 : f(t) = \begin{pmatrix} t \\ \cos t \\ t \sin t \end{pmatrix}$

3. $f : \mathbb{R}^3 \longrightarrow \mathbb{R}^3 : f(x, y, z) = \begin{pmatrix} x + yz \\ x^2 + y^2 + e^z \\ \sin(yz) \end{pmatrix}$

7.2 Courbes dans $\mathbb{R}^n$

7.2.1 Définition

Définition : Une courbe (dans $\mathbb{R}^n$) est une fonction continue $\gamma : I \longrightarrow \mathbb{R}^n$ où I est un intervalle de R. On associe donc pour chaque $t \in I$ un vecteur

$$\gamma(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix}$$

où les $x_i : I \longrightarrow \mathbb{R}$ sont des fonctions réelles.

Proposition 7.3. *γ est continue si et seulement si les x_i sont continues pour tout i.*

Remarques :
- C'est une courbe paramétrisée. La variable t s'appelle le **paramètre** de γ.
- Si l'on peut écrire $x_n = f(x_1, x_2, \ldots, x_{n-1})$ pour tous les points de γ, alors γ est le graphe d'une fonction à $n-1$ variables.

Exemple :

$$\gamma(t) = \begin{pmatrix} t \\ e^t \end{pmatrix} = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}.$$

Alors $y = e^x$.
- Si l'on peut écrire $F(x_1, x_2, \ldots, x_n) = 0$ en éliminant le paramètre t, on a obtenu l'**équation cartésienne implicite** de γ. (ce n'est pas toujours possible). On perd cependant la notion du "temps".

Exemple :

$$\gamma(t) = \begin{pmatrix} R\cos t \\ R\sin t \end{pmatrix}$$

Alors $x^2 + y^2 = R^2$.

7.2.2 Dérivabilité

Définition 7.4 (Vecteur tangent)**.** Si les fonctions $x_i(t)$ sont dérivables pour tout i, on définit le **vecteur tangent** par

$$\dot\gamma(t) = \begin{pmatrix} \dot x_1(t) \\ \dot x_2(t) \\ \vdots \\ \dot x_n(t) \end{pmatrix}$$

où $\dot{x_i}(t) = \dfrac{dx_i}{dt}(t)$.

On note parfois $v(t) = \dot\gamma(t)$.

Interprétation physique : si t représente **le temps** et $\gamma(t)$ la position d'un mobile, alors $\dot\gamma(t)$ est le **vecteur-vitesse**. Sa norme

$$\|\dot\gamma(t)\| = \sqrt{\dot{x_1}^2(t) + \cdots + \dot{x}_n^2(t)}$$

est la **vitesse** du mobile.
En dérivant encore $\dot\gamma(t)$, on obtient le **vecteur d'accélération**

$$\ddot\gamma(t) = \begin{pmatrix} \ddot{x_1}(t) \\ \ddot{x_2}(t) \\ \vdots \\ \ddot{x_n}(t) \end{pmatrix}$$

Quelques exemples

1) Mouvement rectiligne uniforme (MRU). Soit $a, v \in \mathbb{R}^n$ deux vecteurs fixés. Alors la courbe

$$\gamma(t) = a + v \cdot t \qquad (*)$$

est une droite passant par le point a et de vecteur directeur v.

Vecteur tangent (vitesse) : $\dot\gamma(t) = v$ est constante.

- Si $n = 2$, équation cartésienne : $v_2 x - v_1 y = v_2 a_1 - v_1 a_2$ avec $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$.
- Si $n = 3$, l'élimination de t donne 2 équations cartésiennes, chacune étant l'équation d'un plan et l'intersection des 2 plans donne la droite.

2) Mouvement circulaire uniforme (MCU) : soit

$$\gamma(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} R\cos t \\ R\sin t \end{pmatrix}$$

- Equation cartésienne : $x^2 + y^2 = R^2\cos^2 t + R^2\sin^2 t = R^2$.
- Vitesse : $\dot\gamma(t) = \begin{pmatrix} -R\sin t \\ R\cos t \end{pmatrix}$ et donc

$$\|\dot\gamma(t)\| = \sqrt{R^2\sin^2 t + R^2\cos^2 t} = R.$$

La vitesse est constante.
- Accélération : $\ddot\gamma(t) = \begin{pmatrix} -R\cos t \\ -R\sin t \end{pmatrix} = -\gamma(t)$. L'accélération est dirigée vers le centre.

3) Hélice : soit

$$\gamma(t) = \begin{pmatrix} R\cos t \\ R\sin t \\ ct \end{pmatrix} \qquad \text{avec } R, c > 0.$$

C'est la superposition d'un mouvement circulaire uniforme (MCU) dans le plan Oxy et d'un mouvement rectiligne uniforme (MRU) le long de l'axe Oz.

4) Chute libre : $\gamma(t) = \begin{pmatrix} x_0 + v_1 t \\ y_0 + v_2 t \\ z_0 + v_3 t - \frac{1}{2} g t^2 \end{pmatrix}$

$$\dot\gamma(t) = \begin{pmatrix} v_1 \\ v_2 \\ v_3 - gt \end{pmatrix} \qquad \ddot\gamma(t) = \begin{pmatrix} 0 \\ 0 \\ -g \end{pmatrix}$$

Changement de paramétrisation

Soit $I = [a, b]$ et $\gamma : I \longrightarrow \mathbb{R}^n$ une courbe. Soit $J = [c, d]$ et

$$\begin{aligned} \varphi : J &\longrightarrow I \\ s &\mapsto t \end{aligned}$$

une fonction continue bijective telle que $\varphi(c) = a$ et $\varphi(d) = b$.
Alors la courbe

$$\begin{aligned} \gamma \circ \varphi \, : \, J &\longrightarrow \mathbb{R}^n \\ s &\mapsto \gamma\left(\varphi(s)\right) \end{aligned}$$

est la "même" courbe que γ mais avec une paramétrisation différente.
La fonction $t = \varphi(s)$ est le **changement de paramétrisation**.

Interprétation physique : si t est le temps, le changement de paramétrisation revient à parcourir la courbe γ avec une autre vitesse. Plus précisement

$$\frac{d}{ds}\gamma(\varphi(s)) = \frac{d}{ds}\begin{pmatrix} x_1(\varphi(s)) \\ \vdots \\ x_n(\varphi(s)) \end{pmatrix} = \begin{pmatrix} \dot x_1(\varphi(s)) \cdot \varphi'(s) \\ \vdots \\ \dot x_n(\varphi(s)) \cdot \varphi'(s) \end{pmatrix} = \dot\gamma(\varphi(s)) \cdot \varphi'(s).$$

Exemple : Reprenons la courbe

$$\gamma(t) = \begin{pmatrix} R\cos t \\ R\sin t \end{pmatrix} \qquad 0 \leqslant t \leqslant 2\pi$$

Si l'on pose $t = s^2 = \varphi(s)$, la courbe devient

$$\tilde{\gamma}(s) = \gamma \circ \varphi(s) = \begin{pmatrix} R\cos s^2 \\ R\sin s^2 \end{pmatrix} \qquad 0 \leqslant s \leqslant \sqrt{2\pi}$$

Alors $\tilde{v}(s) = \dfrac{d}{ds}\tilde{\gamma}(s) = \begin{pmatrix} -2sR\sin s^2 \\ 2sR\cos s^2 \end{pmatrix}$ et donc

$$v(s) = \|\tilde{v}(s)\| = \sqrt{4s^2R^2\sin^2 s^2 + 4s^2R^2\cos^2 s^2} = 2sR$$

La vitesse n'est plus constante.

7.2.3 Longueur d'une courbe

Soit $P \leftrightarrow \gamma(t)$ et $Q \leftrightarrow \gamma(t+\Delta t)$ deux points de la courbe proches l'un de l'autre (avec $\Delta t > 0$). On a

$$u = \overrightarrow{PQ} = \gamma(t+\Delta t) - \gamma(t) = \begin{pmatrix} x_1(t+\Delta t) - x_1(t) \\ \vdots \\ x_n(t+\Delta t) - x_n(t) \end{pmatrix}$$

Alors la longueur de la corde PQ est égale à

$$\begin{aligned} \Delta L = \|u\| &= \sqrt{[x_1(t+\Delta t) - x_1(t)]^2 + \cdots + [x_n(t+\Delta t) - x_n(t)]^2} \\ &= \sqrt{\left[\frac{x_1(t+\Delta t) - x_1(t)}{\Delta t}\right]^2 + \cdots + \left[\frac{x_n(t+\Delta t) - x_n(t)}{\Delta t}\right]^2} \cdot \Delta t \\ &\xrightarrow{\Delta t \to 0} \sqrt{\dot{x}_1(t)^2 + \cdots + \dot{x}_n(t)^2} \cdot dt \\ &= \|\dot{\gamma}(t)\|\, dt. \end{aligned}$$

La longueur de la courbe entre $t = a$ et $t = b$ vaut donc

$$\boxed{L = \int_a^b \|\dot{\gamma}(t)\|\, dt}$$

Théorème 7.5. *La longueur d'une courbe est indépendante de la paramétrisation*

DÉM : On a $t = \varphi(s)$ avec $\varphi(c) = a$ et $\varphi(d) = b$. On doit montrer que

$$\int_a^b \|\dot{\gamma}(t)\|\, dt = \int_c^d \left\| \frac{d}{ds}[\gamma \circ \varphi(s)] \right\| ds.$$

On suppose que $c < d$ ce qui implique, comme φ est une bijection, que $\varphi'(s) > 0$ pour tout $s \in J$. Alors

$$\begin{aligned}
\int_c^d \left\| \frac{d}{ds} [\gamma \circ \varphi(s)] \right\| ds &= \int_c^d \left\| \dot{\gamma}\big(\varphi(s)\big) \cdot \varphi'(s) \right\| ds \\
&= \int_c^d \|\dot{\gamma}(\varphi(s))\| \cdot |\varphi'(s)| \, ds \\
&= \int_c^d \|\dot{\gamma}(\varphi(s))\| \cdot \varphi'(s) \, ds \\
&= \int_c^d \sqrt{\dot{x}_1(\varphi(s))^2 + \dot{x}_2(\varphi(s))^2 + \cdots + \dot{x}_n(\varphi(s))^2} \cdot \varphi'(s) \, ds \\
&\text{on pose } t = \varphi(s) \text{ et donc } dt = \varphi'(s) ds \\
&= \int_{\varphi(c)}^{\varphi(d)} \sqrt{\dot{x}_1(t)^2 + \dot{x}_2(t)^2 + \cdots + \dot{x}_n(t)^2} \; dt \\
&= \int_a^b \|\dot{\gamma}(t)\| \, dt = L
\end{aligned}$$

□

7.2.4 Angle entre deux courbes

Soient γ_1 et γ_2 deux courbes et $P = \gamma_1(t_1) = \gamma_2(t_2)$ un point d'intersection. Alors l'angle entre γ_1 et γ_2 au point P est donnée par la formule

$$\boxed{\cos\theta = \frac{\langle \, \dot{\gamma}_1(t_1) \, , \, \dot{\gamma}_2(t_2) \, \rangle}{\|\dot{\gamma}_1(t_1)\| \cdot \|\dot{\gamma}_2(t_2)\|}}$$

7.3 Fonctions réelles à plusieurs variables

Une <u>fonction réelle de n variables réelles</u> est une application

$$f \, : \, D \longrightarrow \mathbb{R} \qquad \text{avec } D \subset \mathbb{R}^n.$$

On associe ainsi à tout point $\mathrm{x} = (x_1, \ldots, x_n) \in D$, un **nombre réel** $f(\mathrm{x}) \in \mathbb{R}$.

Exemple : $f : \mathbb{R}^3 \longrightarrow \mathbb{R}$ définie par $f(x, y, z) = x^2 + 2y^2 - 10z$.

7.3.1 Graphe

Si $n = 2$, on peut définir et dessiner le graphe G_f d'une fonction $f(x,y)$. C'est le sous-ensemble de $\mathbb{R}^3$

$$G_f = \{(x,y,z) \mid z = f(x,y)\}.$$

Remarque : le graphe d'une fonction à 2 variables est de dimension 3. De ce fait, si $n \geqslant 3$, le graphe d'une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ peut toujours être défini mais il est de dimension $n+1$ et ne peut donc pas être représenté ni même imaginé.

7.3.2 Ensembles et courbes de niveau

Soit $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction et c un nombre réel. Alors l'ensemble

$$N_f(c) = \{\mathrm{x} \in \mathbb{R}^n \mid f(\mathrm{x}) = c\}$$

est appelé **ensemble de niveau c de f**.

- Dans le cas d'une fonction à 2 variables, $N_f(c)$ est appelé une **courbe de niveau**.

Interprétation : si $f(x,y)$ est l'altitude du point (x,y) sur une carte, une courbe de niveau $N_f(c)$ représente l'ensemble des points qui ont une altitude donnée c.

- Si $n = 3$, on parle de **surface de niveau**.

Exemples 7.6.

1. Soit $f(x_1, x_2, x_2, x_4) = x_1^2 + x_2^2 + x_3^2 + x_4^2$ Alors pour $c > 0$, l'ensemble de niveau

$$N(c) = \left\{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1^2 + x_2^2 + x_3^2 + x_4^2 = c\right\} = \left\{\mathrm{x} \in \mathbb{R}^4 \mid \|\mathrm{x}\| = \sqrt{c}\right\}$$

   est une (hyper)-sphère de rayon $\sqrt{c}$.

2. Graphe et courbes de niveau de la fonction $z = f(x,y) = xy$.

   Courbes de niveau $c \neq 0$: $f(x,y) = c \Longleftrightarrow xy = c \Longleftrightarrow y = \dfrac{c}{x}$
   Les courbes ne niveau $c \neq 0$ sont des hyperboles.
   La courbe de niveau 0 est donnée par $xy = 0 \iff x = 0$ ou $y = 0$. Donc

$$N_f(0) = \text{axe Ox} \,\cup\, \text{axe Oy}$$

   Le graphe de $f(x,y)$ est un hyperboloïde.

3. Graphe et courbes de niveau de $z = f(x,y) = x^2 + 4y^2$.

   Courbes de niveau : $N_f(c) = \{(x,y) \mid x^2 + 4y^2 = c\}$.
   Si $c < 0$, l'ensemble $N_f(c)$ est vide.
   Si $c = 0$, la courbe de niveau est un point : $(0,0)$
   Si $c > 0$ alors $N_f(c)$ est une ellipse d'équation standard

   $$\left(\frac{x}{\sqrt{c}}\right)^2 + \left(\frac{2y}{\sqrt{c}}\right)^2 = 1$$

   Le graphe $z = f(x,y) = x^2 + 4y^2$ est un paraboloïde.

4. Graphe et courbes de niveau de $f(x,y) = x + 2y + 6$.

   Le graphe $z = f(x,y) = x+2y+6$ est un plan dont l'équation normale est $x+2y-z+6 = 0$.
   Son vecteur normal est donc $n = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$.

   Les courbes de niveau c sont d' équation $x + 2y + 6 = c$. Ce sont des droites parallèles de vecteur directeur $v_c = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$.

5. Soit $f(x,y,z) = x^2 + y^2 + z^2$.
   Alors, pour $c > 0$, l'ensemble de niveau est la surface d'équation $x^2 + y^2 + z^2 = c$. C'est une sphère de rayon $\sqrt{c}$.
   Si $c = 0$, l'ensemble de niveau $N_f(0)$ est un point : $(0,0)$.
   Si $c < 0$, $N_f(c)$ est vide.

6. Les surfaces de niveau de la fonction

   $$f(x,y,z) = x - 3y + 4z$$

   sont des plans parallèles d'équation $x - 3y + 4z = c$. Leur vecteur normal est $\begin{pmatrix} 1 \\ -3 \\ 4 \end{pmatrix}$.

7.4 Dérivées partielles

Dans ce paragraphe D désigne un ouvert de $\mathbb{R}^n$.
Soit $f : D \longrightarrow \mathbb{R}$ une fonction à n variables et $a = (a_1, a_2, \ldots, a_n) \in D$ un point fixé. Considérons la fonction

$$g_i(\xi) = f(a_1, \ldots, a_{i-1}, \xi, a_{i+1}, \ldots, a_n).$$

C'est une fonction à une variable.
La dérivée de g est appelée la **dérivée partielle de f par rapport à la variable x_i**.

$$\begin{aligned}\frac{\partial f}{\partial x_i}(a) &:= \frac{dg_i}{d\xi}(a_i)\\ &= \lim_{h\to 0}\frac{g_i(a_i+h)-g_i(a_i)}{h}\\ &= \lim_{h\to 0}\frac{f(a_1,\ldots,a_{i-1},a_i+h,a_{i+1},\ldots,a_n)-f(a_1,\ldots,a_{i-1},a_i,a_{i+1},\ldots,a_n)}{h}\\ &= \lim_{h\to 0}\frac{f(a+he_i)-f(a)}{h}.\end{aligned}$$

On a les notations équivalentes :

$$\boxed{\frac{\partial f}{\partial x_i}(a) = D_i f(a) = f_{x_i}(a) = \left.\frac{\partial f}{\partial x_i}\right|_{\mathrm{x}=a}}$$

Une fonction $f : D \longrightarrow \mathbb{R}$ est dite **partiellement différentiable sur** D si les dérivées partielles $f_{x_i}(\mathrm{x})$ existent pour tout $\mathrm{x} \in D$ et pour tout $i = 1, 2, \ldots, n$.
Ceci définit alors une fonction $f_{x_i} : D \longrightarrow \mathbb{R}$ pour tout i.

Calcul effectif : la dérivée partielle par rapport à x_i se calcule donc en dérivant par rapport à x_i en considérant les autres variables comme des constantes.

Exemples 7.7.
1. Soit $f(x_1,x_2,x_3,x_4) = 2x_1x_2^2 + x_1\sin x_4 + e^{x_2^2}$. Alors

$$\begin{aligned} f_{x_1} &= 2x_2^2+\sin x_4 & f_{x_2} &= 4x_1x_2+2x_2e^{x_2^2}\\ f_{x_3} &= 0 & f_{x_4} &= x_1\cos x_4\end{aligned}$$

2.

$f(x,y)$	f_x	f_y
xy^2	y^2	$2xy$
$\sin(x+2y)$	$\cos(x+2y)$	$2\cos(x+2y)$
$e^{\frac{y}{x}}$	$-\frac{y}{x^2}e^{\frac{y}{x}}$	$\frac{1}{x}e^{\frac{y}{x}}$
x^{2y}	$2yx^{2y-1}$	$2\ln x\cdot x^{2y}$

3. Soit

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

Nous avons déjà vu que f n'est pas continue en $(0,0)$
Mais

$$f_x(0,0) = \lim_{h\to 0}\frac{f(h,0)-f(0,0)}{h} = \lim_{h\to 0}\frac{0}{h} = 0$$

et

$$f_y(0,0) = \lim_{h\to 0}\frac{f(0,h)-f(0,0)}{h} = \lim_{h\to 0}\frac{0}{h} = 0.$$

Ainsi la fonction $f(x,y)$ est dérivable en $(0,0)$ bien qu'elle ne soit pas continue.

Donc en dimension $n \geqslant 2$, dérivable $\not\Longrightarrow$ continue

Remarquons cependant que

$$f_x(x,y) = \frac{y(x^2+y^2)-(xy)(2x)}{(x^2+y^2)^4} = \frac{y^3-x^2y}{(x^2+y^2)^2} \qquad \text{pour tout } (x,y) \neq (0,0)$$

Sur la droite $x = 0$ ($y \neq 0$) on a donc

$$f_x(0,y) = \frac{y^3}{y^4} = \frac{1}{y} \xrightarrow{y \to 0} +\infty$$

et sur la droite $y = 0$ ($x \neq 0$), on a

$$f_x(x,0) = 0 \xrightarrow{x \to 0} 0$$

La fonction $f_x(x,y)$ n'est donc pas continue en $(0,0)$.

On peut montrer que si les $f_{x_i}(\mathrm{x})$ sont continues en a pour tout i, alors $f(\mathrm{x})$ est continue en a.

7.4.1 Gradient

Soit $D \subset \mathbb{R}^n$ et $f : D \longrightarrow \mathbb{R}$ une fonction partiellement différentiable et a un point de D. Le vecteur-ligne

$$\left(\begin{array}{cccc} f_{x_1}(a) & f_{x_2}(a) & \ldots & f_{x_n}(a) \end{array} \right)$$

est appelé **le gradient de f en a**. Il est noté $\nabla f(a)$.
Le gradient définit donc une fonction

$$\begin{aligned} \nabla f \,:\, D &\longrightarrow \mathbb{R}^n \\ \mathrm{x} &\mapsto \nabla f(\mathrm{x}) \end{aligned}$$

Le gradient en a est parfois aussi noté $\overrightarrow{\mathrm{grad}} f(a)$.

Exemples 7.8.

1. si $f(x,y) = x^2 + 2y^2 - 10$ alors

$$\nabla f(x,y) = \left(\begin{array}{cc} 2x & 4y \end{array} \right)$$

2. $f(x,y,z) = x^2y + 3xyz + \sin(xy)$ alors

$$\nabla f(x,y,z) = \left(\begin{array}{ccc} 2xy + 3yz + y\cos(xy) & x^2 + 3xz + x\cos(xy) & 3xy \end{array} \right)$$

7.4.2 Approximation du 1er ordre and plan tangent

Définition 7.9 (Fonction de classe C^1). Une fonction $f : D \longrightarrow \mathbb{R}$ définie sur un ouvert $D \subset \mathbb{R}^n$ est dite **continûment différentiable** si toutes ses dérivées partielles existent et sont continues sur D. On dit dans ce cas que f est de classe C^1 et on note $f \in C^1(D, \mathbb{R})$.

Théorème 7.10 (Théorème d'approximation du premier ordre). *Soit D un ouvert de $\mathbb{R}^n$ et $f \in C^1(D,\mathbb{R})$. Alors pour tout point $a \in D$, il existe une fonction réelle R_1 définie sur un voisinage V de a telle que*

$$f(\mathrm{x}) = f(a) + \underbrace{\nabla f(a) \cdot (\mathrm{x} - a)}_{=\langle \nabla f(a),\, \mathrm{x}-a\rangle} \quad + \quad R_1(\mathrm{x}) \qquad \forall \mathrm{x} \in V$$

avec

$$\lim_{\mathrm{x}\to a} \frac{R_1(\mathrm{x})}{\|\mathrm{x}-a\|} = 0.$$

La fonction R_1 est appelée le reste d'ordre 1. C'est un $o(\|\mathrm{x}-a\|)$.
Ici, x et a sont vus comme des vecteurs-colonne

$$\mathrm{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \qquad a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

Remarque 1 : en posant $u = \mathrm{x} - a \in \mathbb{R}^n$, la formule précédente s'écrit alors

$$f(a+u) = f(a) + \nabla f(a) \cdot u + r(u)$$

avec $\displaystyle\lim_{u\to 0} \frac{r(u)}{\|u\|} = 0$. Autrement dit avec $r(u) = o(\|u\|)$.

Remarque 2 :

- Si $n = 1$, on retrouve l'approximation de Taylor du premier ordre autour de $a = x_0$:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) \ + \ R(x)$$

avec $R(x) = o(x - x_0)$ et l'équation $y = f(x_0) + f'(x_0)(x - x_0)$ est celle de la tangente au graphe de f en x_0.

- Si $n = 2$, on trouve, autour du point $a = (x_0, y_0)$:

$$\begin{aligned} f(x,y) &= f(x_0,y_0) + \begin{pmatrix} f_x(x_0,y_0) & f_y(x_0,y_0) \end{pmatrix} \cdot \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} + R(x,y) \\ &\approx f(x_0,y_0) + f_x(x_0,y_0)\cdot(x - x_0) + f_y(x_0,y_0)\cdot(y - y_0). \end{aligned}$$

L'équation

$$\boxed{z = f(x_0,y_0) + f_x(x_0,y_0)\cdot(x - x_0) + f_y(x_0,y_0)\cdot(y - y_0)}$$

est l'équation du <u>plan tangent au graphe de f au point (x_0, y_0)</u>.

Exemple : soit

$$f(x,y) = 2 + x^2 y - y^3.$$

Alors

$$\nabla f = \begin{pmatrix} 2xy & x^2 - 3y^2 \end{pmatrix}.$$

Soit $a = (x_0, y_0) = (1, 2)$. Ceci donne $f(1,2) = -4$ et $\nabla f(a) = \begin{pmatrix} 4 & -11 \end{pmatrix}$
Approximation du 1er ordre autour de a :

$$\begin{aligned} f(x,y) &= f(x_0,y_0) + \nabla f(x_0,y_0) \cdot \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} + R(x) \\ &\approx -4 + \begin{pmatrix} 4 & -11 \end{pmatrix} \cdot \begin{pmatrix} x-1 \\ y-2 \end{pmatrix} \\ &= -4 + 4(x-1) - 11(y-2). \end{aligned}$$

L'équation $z = -4 + 4(x-1) - 11(y-2)$ est l'équation du plan tangent au graphe de f passant par $P(1,\ 2,\ -4)$.
De manière générale, l'équation du plan tangent au graphe de $f(x,y)$ au point (x_0, y_0) est

$$z = f(x_0,y_0) + \nabla f(x_0,y_0) \cdot \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} = f(x_0,y_0) + f_x(x_0,y_0)\cdot(x-x_0) + f_y(x_0,y_0)\cdot(y-y_0).$$

On peut réécrire l'équation comme suit :

$$f_x(x_0,y_0)\cdot x + f_y(x_0,y_0)\cdot y - z + d = 0$$

ce qui montre que le vecteur normal au plan tangent est

$$\boxed{n = \begin{pmatrix} f_x(x_0,y_0) \\ f_y(x_0,y_0) \\ -1 \end{pmatrix}}$$

On peut généraliser ce résultat à n variables.
L'équation de **l'hyperplan tangent** au graphe de $f(x_1, x_2, \ldots, x_n)$ au point $a = (a_1, a_2, \ldots, a_n)$ est alors donnée par

$$\boxed{x_{n+1} = f(a_1, \ldots, a_n) + \nabla f(a) \cdot (\mathrm{x} - a)}$$

où $\mathrm{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

7.4.3 Règles de dérivation

• **Addition :** soit $f, g \in C^1(D, \mathbb{R})$. Alors

$$\nabla(f+g) = \nabla f + \nabla g.$$

• **Produit :** soient $f, g \in C^1(D, \mathbb{R})$ avec $D \subset \mathbb{R}^n$. On peut considérer la fonction fg définie par $(fg)(\mathrm{x}) = f(\mathrm{x})g(\mathrm{x})$. Alors

$$\nabla(fg)(a) = \nabla(f)(a)\cdot g(a) + f(a)\cdot\nabla g(a).$$

En abrégé :

$$\boxed{\nabla(fg) = f\nabla g + g\nabla f}.$$

- **Composition :** soit $I \subset \mathbb{R}$ et

$$\begin{aligned} \gamma : I &\longrightarrow \mathbb{R}^n \\ t &\mapsto \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} \end{aligned}$$

une courbe (différentiable) dans $\mathbb{R}^n$.
Soit $D \subset \mathbb{R}^n$ et $f : D \longrightarrow \mathbb{R}$ une fonction à n variables. Si l'on suppose que $\operatorname{Im}(\gamma) \subset D$, on peut définir la fonction composée

$$\Phi(t) = f(\gamma(t)) = f(x_1(t),\ x_2(t),\ \ldots,\ x_n(t))$$

qui est une fonction à une variable. Que vaut alors $\frac{d}{dt}\Phi(t)$?
Réponse :

$$\begin{aligned} \frac{d}{dt}\Phi(t) &= \frac{\partial f}{\partial x_1}(x_1(t),\ldots,x_n(t)) \cdot \frac{dx_1}{dt} + \frac{\partial f}{\partial x_2}(x_1(t),\ldots,x_n(t)) \cdot \frac{dx_2}{dt} + \ldots \\ &\qquad \cdots + \frac{\partial f}{\partial x_n}(x_1(t),\ldots,x_n(t)) \cdot \frac{dx_n}{dt} \\ &= \sum_{i=1}^{n} f_{x_i}(\gamma(t)) \cdot \dot{x}_i(t) \\ &= \nabla f(\gamma(t)) \ \cdot \ \dot{\gamma}(t) \\ &= \langle \nabla f(\gamma(t)) \ , \ \dot{\gamma}(t) \rangle. \end{aligned}$$

Ainsi

$$\boxed{\frac{d}{dt}\left(f \circ \gamma\right)(t) = \nabla f\left(\gamma(t)\right) \cdot \dot{\gamma}(t)} \tag{7.1}$$

Exemple (vérification) : $f(x,y) = x^2 + y^2$ et

$$\gamma(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}.$$

Alors

$$\Phi(t) = (f \circ \gamma)(t) = f(\cos t,\ \sin t) = \cos^2 t + \sin^2 t = 1$$

et donc $\Phi'(t) = 0$.
Avec la formule :

$$\begin{cases} f_x = 2x \\ f_y = 2y \end{cases} \qquad \begin{cases} \dot{x}(t) = -\sin t \\ \dot{y}(t) = \cos t \end{cases}$$

Ceci donne

$$\nabla f = (2x \quad 2y) = (2\cos t \quad 2\sin t) \qquad \text{et} \qquad \dot{\gamma}(t) = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix}$$

Alors

$$\Phi'(t) = \nabla f \cdot \dot{\gamma}(t) = -2\cos t \sin t + 2\sin t \cos t = 0.$$

DÉMONSTRATION DE (7.1) : Sans perte de généralité, on démontre la formule en $t = 0$. Posons $a = \gamma(0) \in \mathbb{R}^n$. Par l'approximation du 1er ordre autour de a, on a

$$f(\mathrm{x}) = f(a) + \langle \nabla f(a)\,,\, \mathrm{x} - a\rangle + R(\mathrm{x})$$

avec $\lim\limits_{\mathrm{x}\to a} \dfrac{R(\mathrm{x})}{\|\mathrm{x} - a\|} = 0$.
Ceci donne, en notant $\mathrm{x} = \gamma(h) \in \mathbb{R}^n$:

$$\begin{aligned}
\frac{\Phi(h) - \Phi(0)}{h} &= \frac{f\big(\gamma(h)\big) - f\big(\gamma(0)\big)}{h} = \frac{f(\mathrm{x}) - f(a)}{h} \\
&= \frac{\langle \nabla f(a)\,,\, \mathrm{x} - a\rangle}{h} + \frac{R(\mathrm{x})}{h} \\
&= \left\langle \nabla f(a)\,,\, \frac{\gamma(h) - \gamma(0)}{h} \right\rangle + \frac{R(\mathrm{x})}{\|\mathrm{x} - a\|} \cdot \underbrace{\frac{\|\mathrm{x} - a\|}{h}}_{= \frac{\|\gamma(h) - \gamma(0)\|}{h}} \\
&\xrightarrow{h\to 0} \quad \langle \nabla f(a)\,,\, \dot\gamma(0)\rangle + 0 \cdot \|\dot\gamma(0)\| \\
&= \langle \nabla f(a)\,,\, \dot\gamma(0)\rangle
\end{aligned}$$

ce qui montre que $\Phi'(0) = \langle \nabla f(a)\,,\, \dot\gamma(0)\rangle$. □

7.4.4 Dérivée directionnelle

Définition 7.11. Soit $f \in C^1(D, \mathbb{R})$ avec D un ouvert de R^n, a un point de D et $v \in \mathbb{R}^n$ un vecteur non nul. La quantité

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

s'appelle la dérivée directionnelle de f en a le long de v.

Théorème 7.12. *On a*

$$D_v f(a) = \nabla f(a) \cdot v = \langle \nabla f(a)\,,\, v\rangle.$$

DÉMONSTRATION : Approximation du 1er ordre :

$$f(a + tv) = f(a) + \nabla f(a) \cdot (tv) + R(a + tv)$$

et donc

$$\begin{aligned}
D_v f(a) = {} & \lim_{t\to 0} \frac{f(a + tv) - f(a)}{t} \\
= {} & \lim_{t\to 0} \frac{\nabla f(a) \cdot (tv) + R(a + tv)}{t} \\
= {} & \lim_{t\to 0} \frac{\nabla f(a) \cdot (tv)}{t} + \lim_{t\to 0} \frac{R(a + tv)}{t} \\
= {} & \lim_{t\to 0} \nabla f(a) \cdot v + 0 \\
= {} & \nabla f(a) \cdot v.
\end{aligned}$$

□

Remarque 1 : si $v = e_i$ alors

$$D_{e_i}f(a) = \nabla f(a) \cdot e_i = \begin{pmatrix} f_{x_1}(a) & f_{x_2}(a) & \dots & f_{x_n}(a) \end{pmatrix} \cdot \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = f_{x_i}(a).$$

Remarque 2 : si v est de norme 1 alors

$$D_v f(a) = \langle \nabla f(a)\,,\, v \rangle = \|\nabla f(a)\| \cdot \cos(\alpha)$$

où α est l'angle entre v et $\nabla f(a)$.

Ainsi

$$\begin{aligned} D_v f(a) \text{ est maximal} \quad &\Longleftrightarrow \quad \cos\alpha = 1 \quad \Longleftrightarrow \quad \alpha = 0 \\ &\Longleftrightarrow \quad v \text{ et } \nabla f(a) \text{ sont de même sens et de même direction.} \end{aligned}$$

En conclusion,

le gradient donne la direction dans laquelle la dérivée de f est la plus grande.

De plus

la pente de f dans cette direction vaut $\|\nabla f(a)\|$.

Résumé :

$$\boxed{D_v f(a) = \langle \nabla f(a)\,,\, v \rangle = \text{ dérivée directionnelle de } f \text{ le long de } v}$$

$$\boxed{\text{pente de } f \text{ dans la direction } v = \frac{D_v f(a)}{\|v\|} = \left\langle \nabla f(a)\,,\, \frac{v}{\|v\|} \right\rangle}$$

7.4.5 Gradient et courbes de niveau

Reprenons la règle de composition.
Si $f : D \longrightarrow \mathbb{R}$ est une fonction à n variables et $\gamma(t)$ une courbe dans D, alors

$$\frac{d}{dt}f(\gamma(t)) = \nabla f(\gamma(t)) \cdot \dot{\gamma}(t) = D_{\dot{\gamma}}f(\gamma(t)) \qquad (*)$$

On peut donc interpréter la dérivée $\frac{d}{dt}f(\gamma(t))$ comme la pente de la fonction f en suivant la courbe $\gamma(t)$.

Exemple : Soit $f(x, y) = x^2 + y^2$. Alors

$$\nabla f = \begin{pmatrix} 2x & 2y \end{pmatrix}.$$

Si l'on suit la courbe $\gamma(t) = \begin{pmatrix} t\cos t \\ t\sin t \end{pmatrix}$ alors $\dot{\gamma}(t) = \begin{pmatrix} \cos t - t\sin t \\ \sin t + t\cos t \end{pmatrix}$ et

$$\begin{aligned} \frac{d}{dt}f(\gamma(t)) &= D_{\dot{\gamma}}f(\gamma(t)) = \langle \nabla f(\gamma(t)) \,,\, \dot{\gamma}(t) \rangle \\ &= (2t\cos t \quad 2t\sin t) \cdot \begin{pmatrix} \cos t - t\sin t \\ \sin t + t\cos t \end{pmatrix} \\ &= 2t\cos^2 t - 2t^2\cos t\sin t + 2t\sin^2 t + 2t^2\sin t\cos t = 2t \end{aligned}$$

Courbes de niveau

Si $\gamma(t)$ est une courbe contenue dans l'ensemble de niveau $N_f(c)$ alors $f(\gamma(t)) = c$ par définition de $N_f(c)$ et on doit avoir

$$\frac{d}{dt}f(\gamma(t)) = 0.$$

Ceci donne par $(*)$ ci-dessus :

$$\langle \nabla f(\gamma(t)) \,,\, \dot{\gamma}(t) \rangle = 0.$$

Le gradient est donc orthogonal à $\dot{\gamma}(t)$.

Comme ceci est vrai pour toute courbe $\gamma(t)$ contenue dans $N_f(c)$, on en déduit que

le gradient est toujours orthogonal aux ensembles de niveau.

En particulier,

le gradient d'une fonction à 2 variables est toujours orthogonal aux courbes de niveau.

Exemple : $f(x,y) = x^2 + 2y^2$. Alors $\nabla f = (2x \quad 4y)$.
Les courbes de niveau c^2 sont d'équation

$$x^2 + 2y^2 = c^2.$$

Ce sont donc des ellipses de centre $(0,0)$ et de demi-axe c et $\dfrac{c}{\sqrt{2}}$.

On vérifie que le vecteur $\nabla f = (2x \quad 4y)$ est bien perpendiculaire aux ellipses ci-dessus.

Illustration : le point $P(1,2)$ est sur l'ellipse (γ) : $x^2 + 2y^2 = 9$.
En dérivant, on obtient $2x + 4y \cdot y' = 0$ ce qui donne $y' = -\dfrac{x}{2y}$. Au point $P(1,2)$, on a donc $y'\big|_P = -\dfrac{1}{4}$. Un vecteur tangent est donc

$$v = \begin{pmatrix} -4 \\ 1 \end{pmatrix}.$$

Le gradient de f en P vaut $(\,2 \quad 8\,)$. Il est bien perpendiculaire au vecteur v.

7.4.6 Dérivées partielles d'ordres supérieurs

Soit $f(\mathrm{x}) = f(x_1, x_2, \ldots, x_n)$ une fonction à n variables définie sur un ouvert $D \subset \mathbb{R}^n$. On définit les dérivées d'ordre 2 comme suit

Définition 7.13.

$$f_{x_i x_j} = \frac{\partial^2 f}{\partial x_j \partial x_j} = \frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}$$

Si $i = j$, on note $\dfrac{\partial^2 f}{\partial x_i^2}$ au lieu de $\dfrac{\partial^2 f}{\partial x_i \partial x_i}$.

Exemple :

$$f(x,y) = x^2 y + y e^x \,.$$

Alors

$$f_x = 2xy + ye^x \qquad \text{et} \qquad f_y = x^2 + e^x \,.$$

Ceci donne

$$\begin{aligned} f_{xx} &= \frac{\partial}{\partial x}(2xy + ye^x) = 2y + ye^x \\ f_{yy} &= \frac{\partial}{\partial y}(x^2 + e^x) = 0 \\ f_{xy} &= \frac{\partial}{\partial x}(x^2 + e^x) = 2x + e^x \\ f_{yx} &= \frac{\partial}{\partial y}(2xy + ye^x) = 2x + e^x. \end{aligned}$$

On constate dans cet exemple que $f_{xy} = f_{yx}$.

ATTENTION : en général, $f_{x_ix_j} \neq f_{x_jx_i}$. Mais on a le théorème suivant :

Théorème 7.14. *Si $f_{x_ix_j}$ et $f_{x_jx_i}$ sont continues dans un voisinage du point a, alors $f_{x_ix_j}(a) = f_{x_jx_i}(a)$.*

"DÉMONSTRATION" : La démonstration est faite dans le cas $n = 2$.
Soit $f(x, y)$ et $a = (x_0, y_0)$. Soient $h, k > 0$ tels que le rectangle

$$\{(x,y) \,|\, x_0 - h < x < x_0 + h\,,\ y_0 - k < y < y_0 + k\} \subset D.$$

Définissons les fonctions

$$g_1(x) = f(x, y_0 + k) - f(x, y_0) \qquad \text{et} \qquad g_2(y) = f(x_0 + h, y) - f(x_0, y)$$

Alors

$$g_1'(x) = f_x(x, y_0 + k) - f_x(x, y_0) \qquad \text{et} \qquad g_2'(y) = f_y(x_0 + h, y) - f_y(x_0, y).$$

Posons encore

$$F(h,k) = f(x_0 + h, y_0 + k) - f(x_0 + h, y_0) - f(x_0, y_0 + k) + f(x_0, y_0).$$

Alors

$$F(h,k) = g_1(x_0 + h) - g_1(x_0) \tag{7.2}$$

et

$$F(h,k) = g_2(y_0 + k) - g_2(y_0). \tag{7.3}$$

Le théorème de la moyenne appliqué à (7.2) puis à f_x donne

$$\begin{aligned} F(h,\, k) &= g_1(x_0 + h) - g_1(x_0) \\ &= g_1'(x_0 + \theta_1 h) \cdot h \qquad \theta_1 \in]0;1[\\ &= [f_x(x_0 + \theta_1 h,\, y_0 + k) - f_x(x_0 + \theta_1 h,\, y_0)] \cdot h \\ &= f_{yx}(x_0 + \theta_1 h,\, y_0 + \theta_2 k) \cdot k \cdot h \qquad \theta_2 \in]0;1[. \end{aligned}$$

Ce même théorème de la moyenne appliqué à (7.3) et f_y donne

$$\begin{aligned} F(h,\, k) &= g_2(y_0 + k) - g_2(y_0) \\ &= g_2'(y_0 + \theta_3 k) \cdot k \qquad \theta_3 \in]0;1[\\ &= [f_y(x_0 + h,\, y_0 + \theta_3 k) - f_y(x_0,\, y_0 + \theta_3 k)] \cdot k \\ &= f_{xy}(x_0 + \theta_4 h,\, y_0 + \theta_3 k) \cdot k \cdot h \qquad \theta_4 \in]0;1[. \end{aligned}$$

Donc $f_{xy}(x_0 + \theta_4 h, y_0 + \theta_3 k) \cdot kh = f_{yx}(x_0 + \theta_1 h, y_0 + \theta_2 k) \cdot kh$.

Comme f_{xy} et f_{yx} sont continues, en passant à la limite $h, k \to 0$, on obtient

$$f_{xy}(x_0,\ y_0) = f_{yx}(x_0,\ y_0)\,.$$

$\square$

Définition 7.15. On dit que $f(x_1, x_2, \ldots, x_n)$ est de classe C^2 sur $D \subset \mathbb{R}^n$ si toutes ces dérivées d'ordre 2 sont continues en tout point de D. On note alors

$$f \in C^2(D, \mathbb{R})$$

On généralise sans peine les définitions précédentes pour définir les dérivées partielles d'ordre $k \geqslant 2$.

Exemple : si $f(x, y)$ est une fonction à 2 variables, il y a, à priori, 8 dérivées partielles d'ordre 3 qui sont

$$f_{xxx} \quad f_{yxx} \quad f_{xyx} \quad f_{yyx} \quad f_{xxy} \quad f_{yxy} \quad f_{xyy} \quad f_{yyy}$$

Mais si ces 8 dérivées partielles sont continues, alors

$$f_{yxx} = f_{xyx} = f_{xxy} \qquad \text{et} \qquad f_{yyx} = f_{xyy} = f_{yxy}$$

Définition 7.16 (Fonction de classe C^k)**.** On dit qu'une fonction $f(\mathrm{x})$ est de classe C^k sur $D \subset \mathbb{R}^n$ si toutes ses dérivées partielles d'ordre k (existent et) sont continues. On note alors

$$f \in C^k(D, \mathbb{R})$$

Théorème 7.17. *Si $f \in C^k(D, \mathbb{R})$, alors ses dérivées partielles d'ordre k ne dépendent pas de l'ordre de dérivation mais uniquement du nombre de fois qu'une variable intervient.*

Définition 7.18 (Matrice hessienne)**.** Soit $f(\mathrm{x})$ une fonction à n variables dont les dérivées partielles du 2ème ordre existent. La matrice

$$H_f(\mathrm{x}) = \left(f_{x_i x_j}(\mathrm{x})\right)_{ij} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathrm{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathrm{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathrm{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathrm{x}) & \frac{\partial^2 f}{\partial x_2^2}(\mathrm{x}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\mathrm{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathrm{x}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\mathrm{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\mathrm{x}) \end{pmatrix}$$

s'appelle la **matrice hessienne de f au point** x.

Remarque : Si $f \in C^2(D, \mathbb{R})$ alors la matrice H_f est symétrique : $H_f^T = H_f$.

Exemple : si $n = 2$ et $f = f(x, y)$ alors

$$H_f = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}.$$

Exemple : soit $f(x, y) = 2x^2y + xe^y + 10$. Alors

$$f_{xx} = 4y \qquad f_{xy} = 4x + e^y = f_{yx} \qquad f_{yy} = xe^y$$

et donc

$$H_f(x, y) = \begin{pmatrix} 4y & 4x + e^y \\ 4x + e^y & xe^y \end{pmatrix}$$

qui est symétrique. Au point $(2, 0)$, par exemple, on a $H_f(2, 0) = \begin{pmatrix} 0 & 9 \\ 9 & 2 \end{pmatrix}$.

Théorème 7.19 (Approximation de Taylor d'ordre 2). *Soit $D \subset \mathbb{R}^n$ et $f \in C^2(D, \mathbb{R})$. Soit $a = (a_1, \dots, a_n) \in D$. Alors*

$$f(\mathrm{x}) = f(a) + \nabla f(a) \cdot (\mathrm{x} - a) + \frac{1}{2}(\mathrm{x} - a)^T \cdot H_f(a) \cdot (\mathrm{x} - a) + R_2(\mathrm{x}) \qquad \forall \mathrm{x} \in D$$

avec $\lim\limits_{\mathrm{x} \to a} \dfrac{R_2(\mathrm{x})}{\|\mathrm{x} - a\|^2} = 0$.
La fonction R_2 est le **reste d'ordre 2**. *C'est un $o\left(\|\mathrm{x} - a\|^2\right)$.*

Cas particulier : Si $f = f(x, y)$ et $a = (x_0, y_0)$, la formule devient

$$\begin{aligned} f(x,y) &= f(x_0, y_0) + \begin{pmatrix} f_x(x_0, y_0) & f_y(x_0, y_0) \end{pmatrix} \cdot \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} \\ &\quad + \frac{1}{2} \begin{pmatrix} x - x_0 & y - y_0 \end{pmatrix} \begin{pmatrix} f_{xx}(x_0, y_0) & f_{xy}(x_0, y_0) \\ f_{yx}(x_0, y_0) & f_{yy}(x_0, y_0) \end{pmatrix} \cdot \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} + R_2(x, y) \end{aligned}$$

$$\begin{aligned} &= f(x_0, y_0) + f_x(x_0, y_0) \cdot (x - x_0) + f_y(x_0, y_0) \cdot (y - y_0) + \frac{1}{2} \Big[f_{xx}(x_0, y_0) \cdot (x - x_0)^2 \\ &\quad + 2 f_{xy}(x_0, y_0) \cdot (x - x_0) \cdot (y - y_0) + f_{yy}(x_0, y_0) \cdot (y - y_0)^2 \Big] + R_2(x, y). \end{aligned}$$

Si $a = O(0, 0)$ (approximation à l'origine), ceci devient :

$$f(x, y) = f(O) + f_x(O) \cdot x + f(O) \cdot y + \frac{1}{2} f_{xx}(O) \cdot x^2 + f_{xy}(O) \cdot xy + \frac{1}{2} f_{yy}(O) \cdot y^2 + o(\|(x, y)\|^2)$$

Exemple 7.20. Reprenons l'exemple précédent : $f(x, y) = 2x^2y + xe^y + 10$ autour du point $P(3, 0)$. On a déjà calculer que

$$H_f(x, y) = \begin{pmatrix} 4y & 4x + e^y \\ 4x + e^y & xe^y \end{pmatrix}$$

et donc

$$H_f(3,0) = \begin{pmatrix} 0 & 13 \\ 13 & 3 \end{pmatrix}$$

De plus $\nabla f(x,y) = (4xy + e^y \;\; 2x^2 + xe^y)$ ce qui donne $\nabla f(3,0) = (1 \;\; 21)$. On obtient donc

$$\begin{aligned} f(x,y) &\approx f(3,0) + (1 \;\; 21) \cdot \begin{pmatrix} x-3 \\ y \end{pmatrix} + \frac{1}{2}(x-3 \;\; y) \cdot \begin{pmatrix} 0 & 13 \\ 13 & 3 \end{pmatrix} \cdot \begin{pmatrix} x-3 \\ y \end{pmatrix} \\ &= 13 + (x-3) + 21y + \frac{1}{2}[26(x-3)y + 3y^2] \\ &= 10 + x - 18y + 13xy + \frac{3}{2}y^2 \end{aligned}$$

7.5 Etude de fonction

7.5.1 Extremum

Définition 7.21. Soit $a \in \mathbb{R}^n$ et $r > 0$. La boule $B(a,r)$ est le sous-ensemble des points de $\mathbb{R}^n$ qui sont à une distance strictement inférieure à r du point a :

$$B(a,r) = \{\mathrm{x} \in \mathbb{R}^n \mid \|\mathrm{x} - a\| < r\}$$

Définition 7.22. Un sous-ensemble $D \subset \mathbb{R}^n$ est dit borné s'il existe une boule $B(a,r)$ telle que $D \subset B(a,r)$.
Un sous-ensemble D fermé et borné est dit **compact**.

Définition 7.23. Soit $f \in C^0(D,\mathbb{R})$.

1. Un point $a \in D$ est un maximum (resp. minimum) absolu si $f(\mathrm{x}) \leqslant f(a)$ (resp. $f(\mathrm{x}) \geqslant f(a)$) pour tout $\mathrm{x} \in D$.
2. On dit que $f(\mathrm{x})$ admet un **maximum local** en $a \in D$ s'il existe une boule $B(a,r)$ telle que $f(\mathrm{x}) \leqslant f(a)$ pour tout $\mathrm{x} \in B(a,r)$. Ce maximum est **strict** si l'inégalité est stricte.
3. f admet un **minimum local** en $a \in D$ s'il existe une boule $B(a,r)$ telle que $f(\mathrm{x}) \geqslant f(a)$ pour tout $\mathrm{x} \in B(a,r)$. De même ce minimum est strict si l'inégalité est stricte.
4. On appelle **extremum** (local ou absolu) un point qui est un minimum ou un maximum (local ou absolu).

Théorème 7.24 (Condition nécessaire)**.** *Soit D un ouvert et $f : D \longrightarrow \mathbb{R}$ une fonction continue. Si $a \in D$ est un extremum local et si f est dérivable en a, alors*

$$\nabla f(a) = \vec{0} = \begin{pmatrix} 0 & 0 & \cdots & 0 \end{pmatrix}$$

autrement dit $\dfrac{\partial f}{\partial x_i}(a) = 0 \quad \forall i = 1, \ldots, n.$

Démonstration : La preuve est identique à celle faite en dimension 1. Supposons que a soit un maximum. Alors

$$\frac{\partial f}{\partial x_i}(a) = \lim_{h \to 0} \frac{1}{h} \cdot \underbrace{(f(a + he_i) - f(a))}_{\leqslant 0} \quad \begin{cases} \leqslant 0 & \text{si } h > 0 \\ \geqslant 0 & \text{si } h < 0 \end{cases}$$

Donc $\dfrac{\partial f}{\partial x_i}(a) = 0.$ □

Définition 7.25 (Point stationnaire)**.** Un point $a \in D$ tel que $\nabla f(a) = 0$ est appelé un **point stationnaire**

Pour les extremums absolus sur un fermé D, il faut tenir compte des frontières. Plus précisément :

Théorème 7.26. *Soit D un sous-ensemble compact de $\mathbb{R}^n$. Alors toute fonction continue $f : D \longrightarrow \mathbb{R}$ atteint son maximum et son minimum sur D ; c'est-à-dire qu'il existe $\mathrm{x}_M \in D$ et $\mathrm{x}_m \in D$ avec*

$$f(\mathrm{x}) \leqslant f(\mathrm{x}_M) \qquad et \qquad f(\mathrm{x}) \geqslant f(\mathrm{x}_m) \qquad \forall \mathrm{x} \in D.$$

De plus si $a \in D$ est un extremum absolu, alors
1) soit a est un point stationnaire (i.e. $\nabla f(a) = \vec{0}$) ;
2) soit $\nabla f(a)$ n'existe pas ;
3) soit a appartient à la frontière de D : $a \in \partial D$;

Démonstration : C'est un corollaire du théorème précédent. □

Contre-exemples :
- $f(x, y) = \dfrac{1}{x^2 + y^2}$ sur $D = \mathbb{R}^2 - \{(0, 0)\}$ n'atteint pas son maximum car D n'est pas fermé.
- $f(x, y) = x^2 + y^2$ sur $D = \mathbb{R}^2$ n'atteint pas son maximum car D n'est pas borné.

Application : Pour trouver les extremums de f sur D, il faut donc, en plus des points stationnaires, restreindre la fonction f à la frontière de D et calculer les extremums de la fonction ainsi obtenue.

Exemple 7.27. Cherchons le minimum et le maximum de

$$f(x, y) = 3xy - x^3 - y^3 + 2$$

sur le domaine D limité par les droites $x = 2$, $y = 2$ et $y = -2x + 2$.

(I) $f_x = 3y - 3x^2 \quad f_y = 3x - 3y^2$.
Les points stationnaires sont les solutions du système

$$\begin{cases} 3y - 3x^2 = 0 \\ 3x - 3y^2 = 0 \end{cases}$$

qui sont $S_1(0, 0)$ et $S_2(1, 1)$. Seul S_2 est dans D.

(II) Bords de D :

- Droite $x = 2$: Alors

$$h(y) = f(2, y) = 6y - 8 - y^3 + 2$$

et donc $h'(y) = 6 - 3y^2$. Les zéros de $h'(y)$ sont en $y = \pm\sqrt{2}$. Ceci donne 2 points supplémentaires : $P_1(2, \sqrt{2})$ et $P_2(2, -\sqrt{2})$.

- Droite $y = 2$: Alors

$$g(x) = f(x, 2) = 6x - 8 - x^3 + 2$$

qui donne $g'(x) = 6 - 3x^2$ dont les zéros sont en $x = \pm\sqrt{2}$. Ceci donne un autre point dans D : $P_3(\sqrt{2}, 2)$

- Droite $y = -2x + 2$. La fonction f restreinte à cette droite donne

$$u(x) = f(x, -2x + 2) = 3x(-2x + 2) - x^3 - (-2x + 2)^3 + 2 = \cdots = 7x^3 - 30x^2 + 30x + 2$$

ce qui donne $u'(x) = 21x^2 - 60x + 30$ dont les zéros sont

$$x_{1,2} = \frac{30 \pm \sqrt{270}}{21} \qquad \Longrightarrow \qquad x_1 = 2.21... \qquad x_2 = 0.646.$$

Ceci donne un autre point dans D : $P_4(0.646\,,\ 0.707)$.

(III) Il faut encore calculer la fonction f dans les **coins de** D à savoir aux points $Q_1(2, 2)$, $Q_2(0, 2)$ et $Q_3(2, -2)$.
On obtient donc 8 points sur lesquels il faut étudier la fonction f :

$S_2 : f(1,\ 1) = 3$ $\qquad$ $P_1 : f(2,\ \sqrt{2}) = -0.343$
$P_2 : f(2,\ -\sqrt{2}) = -11.65$ $\qquad$ $P_3 : f(\sqrt{2},\ 2) = -0.343$
$P_4 : f(0.646,\ 0.707) = 2.748$ $\qquad$ $Q_1 : f(2,\ 2) = -2$
$Q_2 : f(0,\ 2) = -6$ $\qquad$ $Q_3 : f(2,\ -2) = -6.$

Le maximum est donc en $(1, 1)$ et le minimum en $(2, -\sqrt{2})$.

7.5.2 Classification des points stationnaires

On a vu que si $a \in D$ est un extremum local, que D est ouvert et que f est de classe C^1 alors $\nabla f(a) = 0$.
Mais la réciproque n'est pas vraie. On peut avoir $\nabla f(a) = 0$ et que a soit un point selle.

Exemple 7.28. $f(x, y) = xy + 4$.
Alors

$$\begin{cases} f_x = y \\ f_y = x \end{cases}$$

et donc $f_x(0, 0) = 0 = f_y(0, 0)$.
Mais le point $(0, 0)$ n'est pas un extremum local.
En effet, sur la droite $y = x$, on a

$$f(x, y) = f(x, x) = x^2 + 4 \geqslant 4 = f(0, 0)\ .$$

Restreinte à cette droite, la fonction f a donc un minimum local en $(0, 0)$.

Mais sur la droite $y = -x$, on a

$$f(x,y) = f(x,-x) = -x^2 + 4 \leqslant 4 .$$

Restreinte à cette droite, la fonction f a donc un maximum local en $(0,0)$.

Le point $(0,0)$ est un point selle.

Définition 7.29 (Point selle). Soit D un ouvert de $\mathbb{R}^n$, $f \in C^1(D,\mathbb{R})$ et $a \in D$ un point stationnaire (i.e. $\nabla f(a) = 0$). On dit que a est un **point selle de** f s'il existe deux vecteurs $u, v \in \mathbb{R}^n$ tels que

- la fonction $t \mapsto f(a+tu)$ admette un maximum local strict en $t = 0$ et
- la fonction $t \mapsto f(a+tv)$ admette un minimum local strict en $t = 0$.

Rappel : soit $f(x) \in C^2(I,\mathbb{R})$ une fonction à 1 variable et a un point stationnaire (i.e. $f'(a) = 0$). Alors

- si $f''(a) > 0$, le point a est un minimum strict ;
- si $f''(a) < 0$, le point a est un maximum strict ;
- si $f''(a) = 0$ on ne peut rien dire, il faut regarder $f^{(3)}(a)$ puis $f^{(4)}(a)$ etc ...

Nous allons établir un résultat analogue pour les fonctions à n variables faisant intervenir la matrice hessienne $H_f(a)$.

Matrices et formes quadratiques

Soit $A = (a_{ij}) \in M_n(\mathbb{R})$ une matrice carrée d'ordre n symétrique $(a_{ij} = a_{ji})$. L'application

$$q_A : \mathbb{R}^n \longrightarrow \mathbb{R}$$

définie par

$$q_A(\mathrm{x}) = \mathrm{x}^T A \mathrm{x} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \cdot \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i,j=1}^{n} a_{ij} x_i x_j$$

est appelée la **forme quadratique associée à** A.

Exemples 7.30.

1) La matrice

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 7 \end{pmatrix}$$

définit la forme quadratique

$$q(x,y) = \begin{pmatrix} x & y \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 2 & 7 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = x^2 + 2xy + 2yx + 7y^2 = x^2 + 4xy + 7y^2.$$

2) La matrice

$$A = \begin{pmatrix} 12 & 2 & -3 \\ 2 & -4 & 0 \\ -3 & 0 & 11 \end{pmatrix}$$

définit la forme quadratique

$$q(x,y,z) = 12x^2 + 2xy - 3xz + 2yx - 4y^2 - 3zx + 11z^2 = 12x^2 + 4xy - 6xz - 4y^2 + 11z^2.$$

3) La matrice identité I_n définit la forme quadratique

$$q(\mathrm{x}) = \mathrm{x}^T\mathrm{x} = \|\mathrm{x}\|^2 = x_1^2 + x_2^2 + \cdots + x_n^2.$$

4) La matrice diagonale

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

définit la forme quadratique

$$q(x_1, x_2, \ldots, x_n) = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2$$

En particulier $q(e_i) = q(0, \cdots, 0, 1, 0, \cdots, 0) = \lambda_i$.

Application : $A = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}$ donne la forme quadratique

$$q(x,y) = x^2 - 2y^2$$

donc $q(1,0) = 1$ et $q(0,1) = -2 < 0$

Propriétés : pour toute forme quadratique $q(\mathrm{x})$ et tout $\lambda \in \mathbb{R}$, on a

$$\boxed{q(\lambda \mathrm{x}) = \lambda^2 q(\mathrm{x})}$$

(d'où le nom de quadratique).

Définition 7.31. Une forme quadratique q est dite

(1) **définie positive** si $q(\mathrm{x}) > 0$ pour tout $\mathrm{x} \in \mathbb{R}^n$ non nul ;

(2) **définie négative** si $q(\mathrm{x}) < 0$ pour tout $\mathrm{x} \in \mathbb{R}^n$ non nul ;

(3) **indéfinie** s'il existe $\mathrm{x}, \mathrm{y} \in \mathbb{R}^n$ avec $q(\mathrm{x}) < 0$ et $q(\mathrm{y}) > 0$.

Une matrice symétrique A est dite définie positive (définie négative, etc..) si la forme quadratique associée q_A est définie positive (définie négative, etc...)

Liens avec les valeurs propres

Comme A est symétrique, elle est diagonalisable :

$$A = PDP^{-1}$$

avec $D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$.

Les λ_i sont les valeurs propres de A. De plus, P est une matrice orthogonale : $P^{-1} = P^T$.

Proposition 7.32. *Soit A, P et D comme ci-dessus. Alors A est définie positive (resp. définie négative, indéfinie) si et seulement si D est définie positive (resp. définie négative, indéfinie). En d'autres termes, la diagonalisation ne change pas la signature de la matrice.*

DÉMONSTRATION : Comme P est inversible (bijective), à tout $\mathrm{x} \in \mathbb{R}^n$, on peut associer un et un seul $\mathrm{y} \in \mathbb{R}^n$ en posant

$$\mathrm{y} = P^{-1}\mathrm{x} = P^T\mathrm{x} \quad \Longleftrightarrow \quad \mathrm{x} = P\mathrm{y}.$$

Alors

$$q_A(\mathrm{x}) = \mathrm{x}^T A\mathrm{x} = \mathrm{x}^T PDP^{-1}\mathrm{x} = \mathrm{x}^T PD\mathrm{y} = (P^T\mathrm{x})^T D\mathrm{y} = \mathrm{y}^T D\mathrm{y} = q_D(\mathrm{y})$$

Ainsi

$$q_A(\mathrm{x}) > 0 \quad \forall \mathrm{x} \in \mathbb{R}^n \qquad \Longleftrightarrow \qquad q_A(\mathrm{y}) > 0 \quad \forall \mathrm{y} \in \mathbb{R}^n .$$

Ceci montre que q_A est définie positive si et seulement si q_D l'est. La démonstration est la même dans le cas d'une matrice définie négative ou indéfinie. □

Remarque 7.33. L'exemple 4) ci-dessous montre de plus qu'une matrice diagonale D est définie positive si et seulement si $\lambda_i > 0$ pour tout $i = 1, 2, \ldots, n$.
De même D est définie négative si et seulement si $\lambda_i < 0$ pour tout i.
Et finalement D est indéfinie si et seulement s'il existe un $\lambda_i < 0$ et un $\lambda_j > 0$ (car alors $q(e_i) = \lambda_i < 0$ et $q(e_j) = \lambda_j > 0$).

Cette remarque et la proposition précédente implique le résultat suivant :

Théorème 7.34. *Une matrice A symétrique est*

(1) définie positive si et seulement si toutes ses valeurs propres sont > 0

(2) définie négative si et seulement si toutes ses valeurs propres sont < 0

(3) indéfinie si et seulement s'il existe $\lambda_i < 0$ et $\lambda_j > 0$.

Exemple : soit

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 7 \end{pmatrix}$$

Alors

$$\lambda_1\lambda_2 = \det(A) = 7 - 4 = 3 > 0$$
$$\lambda_1 + \lambda_2 = \mathrm{Tr}(A) = 8 > 0 .$$

Donc les 2 valeurs propres sont positives et la matrice A est définie positive.

Plus généralement,

Théorème 7.35. *Soit $A \in M_2(\mathbb{R})$ une matrice 2×2 symétrique.*

(1) Si $\det(A) > 0$ et $\mathrm{Tr}(A) > 0$ alors $\lambda_1 > 0$ et $\lambda_2 > 0$ et la matrice A est définie positive ;

(2) si $\det(A) > 0$ et $\mathrm{Tr}(A) < 0$ alors $\lambda_1 < 0$ et $\lambda_2 < 0$ et la matrice A est définie négative ;

(3) si $\det(A) < 0$ alors $\lambda_1 < 0$ et $\lambda_2 > 0$ et la matrice A est indéfinie.

Application aux extremum

Soit $f(\mathrm{x})$ une fonctions à n variables et a un point stationnaire (i.e. $\nabla f(a) = 0$). Alors l'approximation de Taylor d'ordre 2 autour de a devient :

$$\begin{aligned} f(a+u) &= f(a) + \nabla f(a) \cdot u + \frac{1}{2}\, u^T \cdot H_f(a) \cdot u + r_2(u) \\ &\approx f(a) + \frac{1}{2}\, u^T \cdot H_f(a) \cdot u \end{aligned}$$

Ainsi proche de a, la fonction f se comporte comme

$$f(a+u) \approx f(a) + \frac{1}{2}\, q(u)$$

où $q(u)$ est la forme quadratique associée à $H_f(a)$:

$$q(u) = u^T \cdot H_f(a) \cdot u\,.$$

• Ainsi, si $q(u)$ est définie positive, alors $q(u) > 0$ pour tout u et donc

$$f(a+u) = f(a) + \frac{1}{2}\, q(u) > f(a)\,.$$

Le point a est alors un minimum local strict.
• Si $q(u)$ est définie négative, alors $q(u) < 0$ pour tout u et donc

$$f(a+u) = f(a) + \frac{1}{2}\, q(u) < f(a)\,.$$

Le point a est alors un maximum local strict.
• Si $q(u)$ est indéfinie, alors il existe u_1 avec $q(tu_1) = t^2 q(u_1) > 0$ et il existe u_2 avec $q(tu_2) = t^2 q(u_2) < 0$. Ceci implique que

$$\begin{aligned} f(a+tu_1) &= f(a) + q(tu_1) > f(a) \qquad &\text{pour tout } t \neq 0 \text{ et} \\ f(a+tu_2) &= f(a) + q(tu_2) < f(a) \qquad &\text{pour tout } t \neq 0 \end{aligned}$$

ce qui montre que a est un point selle.

En résumé,

Théorème 7.36 (Classification des points stationnaires)**.** *Soit a un point stationnaire d'une fonction $f \in C^2(D, \mathbb{R})$ et $H_f(a)$ la matrice hessienne de f en a.*

1. Si $H_f(a)$ est définie positive, alors a est un minimum strict ;

2. si $H_f(a)$ est définie négative, alors a est un maximum strict ;

3. si $H_f(a)$ est indéfinie, alors a est un point selle.

Dans les autres cas, on ne peut rien dire.

Dans le cas d'une fonction $f(x, y)$ à 2 variables, ce résultat devient

Théorème 7.37. *Soit a un point stationnaire d'une fonction $f(x,y)$ de classe C^2 et $H_f(a)$ sa matrice hessienne (matrice 2×2).*

1. Si $\det(H_f(a)) > 0$ et $\mathrm{Tr}(H_f(a)) > 0$, alors a est un minimum strict;

2. si $\det(H_f(a)) > 0$ et $\mathrm{Tr}(H_f(a)) < 0$, alors a est un maximum strict;

3. si $\det(H_f(a)) < 0$, alors a est un point selle.

Si $\det(H_f(a)) = 0$, on ne peut rien dire.

Exemples 7.38.

1. Reprenons l'exemple vu précédemment : $f(x,y) = xy + 4$ et $P(0,0)$.
   On a vu

$$\begin{cases} f_x = y \\ f_y = x \end{cases} \quad \Longrightarrow \quad \begin{cases} f_{xx} = 0 = f_{yy} \\ f_{xy} = 1 = f_{yx} \end{cases} \quad \Longrightarrow \quad H_f = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = H_f(0,0)$$

   et donc $\det H_f(0,0) = -1 < 0$. Le point $(0,0)$ est un point selle. Conforme à ce que l'on avait déjà trouvé.

2. $f(x,y) = x^3 + 3xy^2 - 15x - 12y + 30$. Alors

$$\begin{cases} f_x = 3x^2 + 3y^2 - 15 \\ f_y = 6xy - 12 \end{cases}$$

   Les points stationnaires sont les solutions du système

$$\begin{cases} 3x^2 + 3y^2 - 15 = 0 \\ 6xy - 12 = 0 \end{cases} \Longleftrightarrow \begin{cases} x^2 + y^2 = 5 \\ x = \frac{2}{y} \end{cases} \Longleftrightarrow \begin{cases} \frac{4}{y^2} + y^2 = 5 \\ x = \frac{2}{y} \end{cases} \Longleftrightarrow \begin{cases} y^4 - 5y^2 + 4 = 0 \\ x = \frac{2}{y} \end{cases}$$

   ce qui donne $y^2 = 4$ ou $y^2 = 1$. Les points stationnaires sont donc

$$P_1(1,2) \qquad P_2(-1,-2) \qquad P_3(2,1) \qquad P_4(-2,-1)$$

   Calculons maintenant H_f :

$$\begin{cases} f_{xx} = 6x \\ f_{xy} = 6y = f_{yx} \\ f_{yy} = 6x \end{cases} \quad \Longrightarrow \quad H_f(x,y) = \begin{pmatrix} 6x & 6y \\ 6y & 6x \end{pmatrix}$$

   - Point P_1 :

$$H_f(1,2) = \begin{pmatrix} 6 & 12 \\ 12 & 6 \end{pmatrix} \quad \Longrightarrow \quad \det(H_f) = 36 - 144 < 0$$

   $\Longrightarrow P_1$ est un point selle.

   - Point P_2 :

$$H_f(-1,-2) = \begin{pmatrix} -6 & -12 \\ -12 & -6 \end{pmatrix} \quad \Longrightarrow \quad \det(H_f) = 36 - 144 < 0$$

   $\Longrightarrow P_2$ est un point selle.

   - Point P_3 :

$$H_f(2,1) = \begin{pmatrix} 12 & 6 \\ 6 & 12 \end{pmatrix} \quad \Longrightarrow \quad \det(H_f) = 144 - 36 > 0 \quad \text{et} \quad \mathrm{Tr}(H_f) = 24 > 0$$

$\Longrightarrow P_3$ est un minimum local.

- Point P_4 :

$$H_f(-2,-1) = \begin{pmatrix} -12 & -6 \\ -6 & -12 \end{pmatrix} \quad \Longrightarrow \quad \det(H_f) = 144-36 > 0 \quad \text{et} \quad \mathrm{Tr}(H_f) = -24 < 0$$

$\Longrightarrow P_4$ est un maximum local.

3. Soit
$$f(x,y,z) = \frac{1}{2}x^2y - \frac{3}{2}x^2 + xyz - xz + 2x - y^2 + \frac{3}{2}y - z^2$$
(i) Vérifier que le point $P(1,\,1,\,0)$ est un point stationnaire.
(ii) Etudier sa nature.

(i)

$$\begin{aligned} f_x &= xy - 3x + yz - z + 2 &&\Longrightarrow\quad f_x(1,\,1,\,0) = 0 \\ f_y &= \frac{1}{2}x^2 + xz - 2y + \frac{3}{2} &&\Longrightarrow\quad f_y(1,\,1,\,0) = 0 \\ f_z &= xy - x - 2z &&\Longrightarrow\quad f_z(1,\,1,\,0) = 0. \end{aligned}$$

On a bien $\nabla f(1,\,1,\,0) = \overrightarrow{0}$.

(ii)

$$\begin{array}{lll} f_{xx} = y-3 & f_{yy} = -2 & f_{zz} = -2 \\ f_{xy} = x+z & f_{xz} = y-1 & f_{yz} = x \end{array}$$

La matrice hessienne de f au point $P(1,\,1,\,0)$ vaut alors

$$H_f(1,\,1,\,0) = \begin{pmatrix} y-3 & x+z & y-1 \\ x+z & -2 & x \\ y-1 & x & -2 \end{pmatrix}\Bigg|_{P(1,1,0)} = \begin{pmatrix} -2 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -2 \end{pmatrix} =: A$$

Il faut calculer les valeur propres de A :

$$\begin{aligned} \det(\lambda I_3 - A) = 0 &\quad\Longleftrightarrow \\ \begin{vmatrix} \lambda+2 & -1 & 0 \\ -1 & \lambda+2 & -1 \\ 0 & 1 & \lambda+2 \end{vmatrix} = 0 &\quad\Longleftrightarrow \text{(Sarus)} \\ (\lambda+2)^3 - (\lambda+2) - (\lambda+2) = 0 &\quad\Longleftrightarrow \\ \lambda^3 + 6\lambda^2 + 10\lambda + 4 = 0 & \end{aligned}$$

$\lambda_1 = -2$ est une solution et par factorisation, on obtient

$$\lambda^3 + 6\lambda^2 + 10\lambda + 4 = (\lambda+2)(\lambda^2 + 4\lambda + 2) = 0$$

$\Longrightarrow \lambda_{2,3} = -2 \pm \sqrt{2}$.
Les 3 valeurs propres étant < 0, le point $P(1,\,1,\,0)$ est un maximum local.

7.6 Théorème des fonctions implicites

7.6.1 Exemple

Soit $F(x, y)$ une fonction à 2 variables. Considérons la courbe de niveau 0

$$\gamma = N_F(0) = \{(x, y) \mid F(x, y) = 0\}.$$

1) On peut essayer de résoudre l'équation $F(x, y) = 0$ en fonction de y pour établir $y = f(x)$. Si on peut le faire, alors la courbe γ est le graphe de la fonction $y = f(x)$ et on a

$$F(x, f(x)) = 0.$$

En dérivant cette dernière égalité par rapport à x, on trouve

$$F_x(x, f(x)) + F_y(x, f(x)) \cdot f'(x) = 0$$

ce qui donne

$$f'(x) = -\frac{F_x(x, f(x))}{F_y(x, f(x))}.$$

Exemple : soit

$$F(x, y) = x^2 + e^y + 2x = 0.$$

Alors

$$y = f(x) = \ln(-x^2 - 2x)$$

pour $x \in]-2; 0[$.
Calcul direct : $y' = f'(x) = \dfrac{-2x - 2}{-x^2 - 2x}$.
Calcul implicite : $F_x = 2x + 2$ et $F_y = e^y$. Alors

$$f'(x) = -\frac{F_x}{F_y} = -\frac{2x + 2}{e^y} = -\frac{2x + 2}{-x^2 - 2x}.$$

2) Dans le cas général, il n'est pas possible de calculer $y = f(x)$. Mais, si l'on suppose que localement - sur un ouvert contenant $P(x_0, y_0)$ - il existe $y = y(x)$ alors on a vu au chapitre 4 qu'on peut dériver la relation $F(x, y) = 0$ par rapport à x (dérivée droite) pour trouver $y'\big|_P$. Ceci donne

$$F_x(x_0, y_0) + F_y(x_0, y_0) \cdot y'(x_0, y_0) = 0 \quad \Longrightarrow \quad \boxed{y'\big|_P = -\frac{F_x(x_0, y_0)}{F_y(x_0, y_0)}}$$

La formule est la même que précédemment.

Exemple : soit

$$F(x, y) = x^2 + y^2 - 25 = 0$$

le cercle de centre $(0, 0)$ et de rayon 5. Alors

$$\begin{cases} F_x = 2x \\ F_y = 2y \end{cases}$$

On obtient ainsi

$$y'\big|_P = -\frac{2x_0}{2y_0} = -\frac{x_0}{y_0} \qquad \text{pour } P(x_0, y_0)$$

Sans calculer explicitement $y = y(x)$ on peut donc trouver la dérivée $y'(x_0)$ en tout point (x_0, y_0). L'équation de la tangente à γ au point $P(x_0, y_0)$ est alors

$$y = y'\big|_P \cdot (x - x_0) + y_0 = -\frac{F_x\big|_P}{F_y\big|_P} \cdot (x - x_0) + y_0$$

ce qui donne

$$\text{Equation de la tangente en } P : \quad \boxed{F_x\big|_P \cdot (x - x_0) + F_y\big|_P \cdot (y - y_0) = 0}$$

Exemple précédent : la tangente au cercle $x^2 + y^2 - 25 = 0$ au point $P(3, 4)$ est

$$F_x(3,4) \cdot (x - 3) + F_y(3,4) \cdot (y - 4) = 0$$

ce qui donne

$$6 \cdot (x - 3) + 8 \cdot (y - 4) = 0.$$

7.6.2 Théorème général

Soit

$$F(x_1, \dots, x_n) = 0$$

une équation à n variables.
Les résultats précédents se généralisent de la façon suivante.

Premièrement, le théorème suivant affirme que, localement, et sous certaines hypothèses, il est possible de résoudre l'équation en fonction d'une variable, par exemple $x_n = f(x_1, \dots, x_{n-1})$
Secondement, il affirme que les dérivées partielles de $f(x_1, \dots, x_{n-1})$ se calculent à partir des dérivées partielles F_{x_i} comme précédemment.

Théorème 7.39 (Théorème des fonctions implicites). *Soit*

- *$D \subset \mathbb{R}^n$ un ouvert,*
- *$F : D \longrightarrow \mathbb{R}$ une fonction de classe C^1 et*
- *$a = (a_1, \dots, a_n) \in D$ un point satisfaisant $F(a) = 0$.*
- *On suppose que $F_{x_n}(a) \neq 0$.*

Posons $\hat{a} = (a_1, \dots, a_{n-1}) \in \mathbb{R}^{n-1}$.
Alors il existe un ouvert $U \subset \mathbb{R}^{n-1}$ contenant $\hat{a}$ et une fonction à $n-1$ variables $f : U \longrightarrow \mathbb{R}$ satisfaisant :

1. *le graphe de f est contenu dans D :*

$$(x_1, \dots, x_{n-1}, f(x_1, \dots, x_{n-1})) \in D \qquad \text{pour tout } (x_1, \dots, x_{n-1}) \in U.$$

2. *$f(a_1, \dots, a_{n-1}) = a_n$*
3. *$F(x_1, \dots, x_{n-1}, f(x_1, \dots, x_{n-1})) = 0$ pour tout $(x_1, \dots, x_{n-1}) \in U$*
4. *la fonction f est de classe C^1 et*

$$f_{x_i} = -\frac{F_{x_i}}{F_{x_n}} \qquad \forall i \in \{1, \dots, n-1\}$$

Remarque 7.40. La condition $F_{x_n}(a) \neq 0$ est essentielle. Cependant si $F_{x_n}(a) = 0$ mais que $F_{x_k}(a) \neq 0$ pour un certain $k \neq n$, alors le résultat est vrai pour x_k c'est-à-dire qu'il existe une fonction f avec

$$x_k = f(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots x_n)$$

On explicite x_k au lieu de x_n.

En revanche, si $F_{x_i}(a) = 0$ pour tout $i = 1, \ldots, n$ (le point a est un point stationnaire), le théorème n'est plus vrai et il est impossible d'expliciter une variable x_i en fonction des autres.

7.6.3 Application : équation du plan tangent à une surface implicite

Considérons la surface Γ définie par l'équation

$$F(x, y, z) = 0$$

et un point $P(x_0, y_0, z_0) \in \Gamma$, i.e. $F(x_0, y_0, z_0) = 0$. Supposons de plus que $F_z\big|_P \neq 0$.
Alors, par le théorème des fonctions implicites, il existe une fonction

$$z = f(x, y)$$

décrivant la surface Γ au voisinage de P. En particulier $f(x_0, y_0) = z_0$.
L'approximation du 1er ordre de cette fonction f donne l'équation du plan tangent :

$$z = f(x_0, y_0) + \nabla f(x_0, y_0) \cdot \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} = f(P) + f_x(P) \cdot (x - x_0) + f_y(P) \cdot (y - y_0)$$

Or le théorème des fonctions implicites donne f_x et f_y :

$$f_x(P) = -\frac{F_x(P)}{F_z(P)} \qquad \text{et} \qquad f_y(P) = -\frac{F_y(P)}{F_z(P)}$$

L'équation du plan tangent devient :

$$z = f(P) - \frac{F_x(P)}{F_z(P)} \cdot (x - x_0) - \frac{F_y(P)}{F_z(P)} \cdot (y - y_0)$$

qui s'écrit encore

$$\boxed{F_x(P) \cdot (x - x_0) + F_y(P) \cdot (y - y_0) + F_z(P) \cdot (z - z_0) = 0.}$$

Ceci montre que le vecteur normal au plan tangent et donc à la surface Γ est

$$n = \begin{pmatrix} F_x(P) & F_y(P) & F_z(P) \end{pmatrix} = \nabla F(P)$$

Le gradient ∇F est donc bien orthogonal à la surface de niveau $N_F(0) = \Gamma$.

De manière générale si une hyper-surface Γ est donnée par l'équation

$$F(x_1, \ldots, x_n) = 0$$

alors l'équation de l'hyper-plan tangent à Γ passant par $a = (a_1, \dots, a_n) \in \Gamma$ est

$$\boxed{F_{x_1}(a) \cdot (x_1 - a_1) + F_{x_2}(a) \cdot (x_2 - a_2) + \cdots + F_{x_n}(a) \cdot (x_n - a_n) = 0}$$

qui s'écrit aussi

$$\nabla F(a) \cdot \begin{pmatrix} x_1 - a_1 \\ \vdots \\ x_n - a_n \end{pmatrix} = 0$$

Exemples 7.41.
1. Considérons l'ellipsoïde

$$\Gamma : F(x, y, z) = 3x^2 + 6y^2 + z^2 - 36 = 0.$$

Alors

$$\begin{cases} F_x = 6x \\ F_y = 12y \\ F_z = 2z \end{cases}$$

et l'équation du plan tangent au point $P(1, 2, 3) \in \Gamma$ est

$$\begin{aligned} F_x(1,2,3) \cdot (x-1) + F_y(1,2,3) \cdot (y-2) + F_z(1,2,3) \cdot (z-3) = 0 \quad &\iff \\ 6(x-1) + 24(y-2) + 6(z-3) = 0 \quad &\iff \\ 6x + 24y + 6z = 72. \end{aligned}$$

2. Considérons la surface

$$(\Gamma) : \quad xyz = 1$$

et le point $P(2, \frac{1}{6}, 3) \in \Gamma$. Alors, en notant $F(x, y, z) = xyz - 1$, on a

$$\begin{cases} F_x = yz \\ F_y = xz \\ F_z = xy \end{cases} \implies \begin{cases} F_x(P) = \frac{1}{2} \\ F_y(P) = 6 \\ F_z(P) = \frac{1}{3} \end{cases}$$

L'équation du plan tangent à Γ par P est alors

$$\frac{1}{2} \cdot (x-2) + 6 \cdot \left(y - \frac{1}{6}\right) + \frac{1}{3} \cdot (z-3) = 0.$$

7.7 Extremum sous contraintes : multiplicateur de Lagrange

7.7.1 Multiplicateur de Lagrange

On désire trouver le maximum (ou minimum) d'une fonction $f(x_1, \dots, x_n)$ sur un domaine D avec la contrainte $g(x_1, \dots, x_n) = 0$. On suppose que $\nabla g(\mathrm{x}) \neq 0$ sur D.

1ère méthode : si l'on peut expliciter

$$x_n = \tilde{g}(x_1, \dots, x_{n-1}),$$

alors

$$f(x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_{n-1}, \tilde{g}(x_1, \dots, x_{n-1})) = \Phi(x_1, \dots, x_{n-1})$$

et on cherche les extremums de Φ.

<u>2ème méthode</u> : si l'on peut trouver une paramétrisation de la courbe $g(x_1, \dots, x_n) = 0$ de la forme

$$(\gamma) : \quad \begin{cases} x_1 &= x_1(t) \\ & \vdots \\ x_n &= x_n(t) \end{cases}$$

alors $f(x_1, \dots, x_n) = f(x_1(t), \dots, x_n(t)) = \Phi(t)$ et on cherche les extremums de Φ.

<u>3ème méthode</u> : si aucune des 2 méthodes ne marche, on utilise un **multiplicateur de Lagrange**.

Cette méthode consiste à résoudre le système d'équations en $\mathrm{x} \in D$:

$$\boxed{\begin{cases} \nabla f(\mathrm{x}) + \lambda \nabla g(\mathrm{x}) = 0 \\ g(\mathrm{x}) = 0 \end{cases}}$$

C'est un système à $n+1$ inconnues $x_1, x_2, \dots, x_n, \lambda$ et à $n+1$ équations.

Cette méthode est basée sur le théorème suivant :

Théorème 7.42. *Soit $D \subset \mathbb{R}^n$ un ouvert et $f, g \in C^1(D, \mathbb{R})$ et $M = \{\mathrm{x} \in D \,|\, g(\mathrm{x}) = 0\}$. Soit $a = (a_1, \dots, a_n) \in M$ un extremum local de la fonction $f\big|_M$. On suppose de plus que $\nabla g(a) \neq 0$. Alors il existe un $\lambda \in \mathbb{R}$ tel que*

$$\nabla f(a) + \lambda \nabla g(a) = 0.$$

Autrement dit, $\nabla f(a)$ et $\nabla g(a)$ sont colinéaires.
λ est appelé un multiplicateur de Lagrange.

Démonstration : On peut supposer $\dfrac{\partial g}{\partial x_n}(a) \neq 0$. Alors par le théorème des fonctions implicites, il existe une fonction $\Phi(x_1, \dots, x_{n-1})$ dans un voisinage $U = B(\tilde{a}, r)$ de $\tilde{a} = (a_1, \dots, a_{n-1})$ telle que

$$\begin{cases} \Phi(a_1, \cdots, a_{n-1}) = a_n \\ g(x_1, \dots, x_{n-1}, \Phi(x_1, \dots, x_{n-1})) = 0 \qquad \forall \mathrm{x} \in U \end{cases}$$

avec

$$\frac{\partial \Phi}{\partial x_i}(\tilde{a}) = -\frac{\frac{\partial g}{\partial x_i}(a)}{\frac{\partial g}{\partial x_n}(a)} \qquad \text{pour tout } i = 1, \dots, n-1$$

Posons alors

$$F(x_1, \dots, x_{n-1}) = f(x_1, \dots, x_{n-1}, \Phi(x_1, \dots, x_{n-1})).$$

Alors, par hypothèse, la fonction F (qui est $f\big|_M$) admet un extremum en $\tilde{a}$. On a donc $\nabla F(\tilde{a}) = 0$ ce qui donne, avec la règle de composition :

$$\begin{aligned} 0 = \frac{\partial F}{\partial x_i}(\tilde{a}) &= \frac{\partial f}{\partial x_i}(a) + \frac{\partial f}{\partial x_n}(a) \cdot \frac{\partial \Phi}{\partial x_i}(\tilde{a}) \\ &= \frac{\partial f}{\partial x_i}(a) + \frac{\partial f}{\partial x_n}(a) \cdot \left(-\frac{\frac{\partial g}{\partial x_i}(a)}{\frac{\partial g}{\partial x_n}(a)} \right) \\ &= \frac{\partial f}{\partial x_i}(a) + \left(-\frac{\frac{\partial f}{\partial x_n}(a)}{\frac{\partial g}{\partial x_n}(a)} \right) \cdot \frac{\partial g}{\partial x_i}(a) \qquad \text{pour tout } i = 1, \dots, n-1 \end{aligned}$$

En posant $\lambda = -\dfrac{\frac{\partial f}{\partial x_n}(a)}{\frac{\partial g}{\partial x_n}(a)}$, on obtient

$$0 = \frac{\partial f}{\partial x_i}(a) + \lambda \cdot \frac{\partial g}{\partial x_i}(a) \qquad \text{pour tout } i = 1, \ldots, n-1, n$$

et donc $\nabla f(a) + \lambda \nabla g(a) = 0$.

□

ATTENTION : $\nabla f(a) + \lambda g(a) = 0$ est une condition nécessaire mais pas suffisante. Il faut donc vérifier que la solution cherchée est bien un extremum.

Exemples 7.43.
1) Trouver le maximum et le minimum de la fonction

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

sur le bord de l'ellipse $x^2 + 4y^2 = 1$.
Solution : la condition est $g(x, y) = x^2 + 4y^2 - 1 = 0$. Alors

$$\nabla f = \begin{pmatrix} 2x - 1 \\ 8y + 2 \end{pmatrix}^T \qquad \nabla g = \begin{pmatrix} 2x \\ 8y \end{pmatrix}^T$$

et le système à résoudre est alors

$$\begin{cases} 2x - 1 + \lambda 2x &= 0 \\ 8y + 2 + \lambda 8y &= 0 \\ x^2 + 4y^2 - 1 &= 0 \end{cases}$$

ou

$$\begin{cases} 2(\lambda + 1)x &= -1 \quad (I) \\ 4(1 + \lambda)y &= -1 \quad (II) \\ x^2 + 4y^2 &= 1 \end{cases}$$

En élevant au carré et en additionant les équations (I) et (II), on trouve

$$\begin{cases} 4(\lambda + 1)^2(x^2 + 4y^2) &= 2 \\ x^2 + 4y^2 &= 1 \end{cases}$$

ce qui donne

$$4(\lambda + 1)^2 = 2 \quad \Longrightarrow \quad \lambda + 1 = \pm\sqrt{1/2} \quad \Longrightarrow \quad \lambda = -1 \pm \sqrt{\frac{1}{2}}.$$

On trouve alors 2 points :

$$P\left(\frac{1}{\sqrt{2}}, -\frac{1}{2\sqrt{2}}\right) \qquad \text{et} \qquad Q\left(-\frac{1}{\sqrt{2}}, \frac{1}{2\sqrt{2}}\right).$$

Un calcul montre que P est le minimum et Q le maximum.

2) Construire une boite de forme parallélépipèdique sans couvercle, de volume donné V de telle sorte que la surface du fond et des parois soit minimale.

La fonction à minimiser est

$$A = f(x, y, z) = xy + 2xz + 2yz$$

et la contrainte est

$$g(x, y, z) = xyz - V = 0.$$

Le système à résoudre $\nabla f(x, y, z) + \lambda \nabla g(x, y, z) = 0$ devient

$$\begin{cases} y + 2z + \lambda yz &= 0 & (I) \\ x + 2z + \lambda xz &= 0 & (II) \\ 2x + 2y + \lambda xy &= 0 & (III) \\ xyz &= V & (IV) \end{cases}$$

$(I) \cdot x - (II) \cdot y$ donne $2xz - 2yz = 0$ et donc $x = y$.
Le système (partiel) devient :

$$\begin{cases} x + 2z + \lambda xz &= 0 & (II) \\ 4 + \lambda x &= 0 & (III) \\ x^2 z &= V & (IV) \end{cases}$$

$(II) - (III) \cdot z$ donne $x - 2z = 0$ donc $x = 2z$. Finalement (IV) devient $4z^3 = V$ ce qui donne

$$z = \sqrt[3]{\frac{V}{4}} \qquad x = y = 2z = \sqrt[3]{2V}.$$

7.7.2 Contraintes multiples

Pour trouver les extremums d'une fonction $f(\mathrm{x})$ avec m contraintes

$$\begin{cases} g_1(\mathrm{x}) = 0 \\ g_2(\mathrm{x}) = 0 \\ \vdots \\ g_m(\mathrm{x}) = 0 \end{cases}$$

on résoud le système

$$\begin{cases} \nabla f(\mathrm{x}) + \displaystyle\sum_{j=1}^{m} \lambda_j \nabla g_j(\mathrm{x}) &= 0 \\ g_1(\mathrm{x}) &= 0 \\ &\vdots \\ g_m(\mathrm{x}) &= 0 \end{cases}$$

C'est un système à $n + m$ inconnues : $x_1, x_2, \ldots, x_n, \lambda_1, \ldots, \lambda_m$ et $n + m$ équations.

Exemple : trouvons le maximum de la fonction

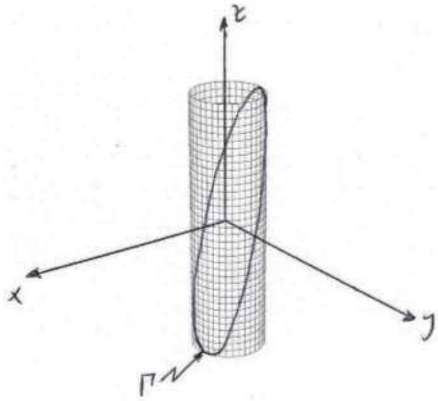
$$f(x, y, z) = x + y + z$$

avec les contraintes

$$\begin{cases} g_1(x, y, z) = x^2 + y^2 - 2 = 0 \\ g_2(x, y, z) = x + z - 1 \end{cases}$$

Géométriquement $g_1(x, y, z) = 0$ est l'équation d'un cylindre vertical et $g_2(x, y, z) = 0$ celle d'un plan. Leur intersection est une ellipse Γ. On cherche donc le maximum de f sur l'ellipse

Γ.



Comme

$$\nabla f(x,y,z) = (1 \quad 1 \quad 1) \qquad \nabla g_1(x,y,z) = (2x \quad 2y \quad 0) \qquad \nabla g_2(x,y,z) = (1 \quad 0 \quad 1),$$

il faut résoudre le système

$$\begin{cases} 1 + \lambda_1 2x + \lambda_2 &= 0 \quad (I) \\ 1 + \lambda_1 2y &= 0 \quad (II) \\ 1 + \lambda_2 &= 0 \quad (III) \\ x^2 + y^2 &= 2 \quad (IV) \\ x + z &= 1 \quad (V) \end{cases}$$

On voit que $\lambda_2 = -1$ par (III) et donc $x = 0$ ce qui donne $y = \pm\sqrt{2}$ et $z = 1$. On obtient 2 points

$$P(0, \sqrt{2}, 1) \qquad \text{et} \qquad Q(0, -\sqrt{2}, 1).$$

Comme $f(P) = 1 + \sqrt{2}$ et $f(Q) = 1 - \sqrt{2}$, on en déduit que P est le maximum cherché.

7.8 Fonctions de $\mathbb{R}^n$ dans $\mathbb{R}^m$

7.8.1 Gradient

Dans cette section, nous allons étudier les fonctions

$$f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

En physique, ces fonctions sont appelées **champs vectoriels.**
A tout point $\mathrm{x} = (x_1, \ldots, x_n) \in D$, on associe un vecteur

$$f(x_1, \ldots, x_n) = \begin{pmatrix} f_1(x_1, \ldots, x_n) \\ f_2(x_1, \ldots, x_n) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{pmatrix}$$

La fonction $f_k : D \longrightarrow \mathbb{R}$ est la **i-ème composante de** f.
L'image d'un point x sera vue comme un vecteur-colonne.

Définition-Proposition 7.44. Soit D un ouvert de $\mathbb{R}^n$ et $f : D \longrightarrow \mathbb{R}^m$ une fonction.
La fonction f est continue (resp. dérivable, resp. de classe C^1) si toutes ses composantes $f_k : D \longrightarrow \mathbb{R}$ sont continues (resp. dérivables, resp. de classe C^1).

Exemples 7.45.
1. Si $n = 1$, on retrouve les courbes déjà étudiées $\gamma : I \longrightarrow \mathbb{R}^m$.

2. Si $m = 1$, on retrouve les fonctions réelles à plusieurs variables $f : D \longrightarrow \mathbb{R}$.

3. La fonction $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ définie par

$$f(x,y) = \begin{pmatrix} x^2 + 2y \\ 2x + \sin y \end{pmatrix}$$

est continue et de classe C^1 sur tout $\mathbb{R}^2$.

4. Soit $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ la fonction définie par

$$f(x,y) = x^2 + 5xy - 4x.$$

Alors le gradient de f définit un champ vectoriel

$$\begin{aligned} \nabla f : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ (x,y) &\mapsto \begin{pmatrix} 2x + 5y - 4 \\ 5x \end{pmatrix} \end{aligned}$$

Définition 7.46. Soit $D \subset \mathbb{R}^n$ un ouvert, $f : D \longrightarrow \mathbb{R}^m$ une fonction dérivable et $\mathrm{x} \in D$. La matrice $m \times n$

$$\nabla f(\mathrm{x}) = \begin{pmatrix} \nabla f_1(\mathrm{x}) \\ \nabla f_2(\mathrm{x}) \\ \vdots \\ \nabla f_m(\mathrm{x}) \end{pmatrix} = \left(\frac{\partial f_i}{\partial x_j}(\mathrm{x}) \right)_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}}$$

est appelée le **gradient de f en** x ou également la **matrice jacobienne de f**.

Remarques :
1. si $f : D \longrightarrow \mathbb{R}$ $(m = 1)$, la matrice est un vecteur-ligne et on retrouve le gradient définit dans la section 7.4.1.
2. Dans le cas d'une courbe

$$\begin{aligned} \gamma : I &\longrightarrow \mathbb{R}^m \\ t &\mapsto \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_m(t) \end{pmatrix} \end{aligned}$$

on a alors $n = 1$ et le gradient de γ est un vecteur-colonne qui n'est rien d'autre que le vecteur tangent :

$$\nabla \gamma(t) = \dot{\gamma}(t).$$

Exemples 7.47.
1. Considérons la fonction de $\mathbb{R}^3$ dans $\mathbb{R}^2$:

$$f(x,y,z) = \begin{pmatrix} x^2y + 2yz + \sin x \\ xyz + 2x \end{pmatrix}$$

Le gradient de f en (x,y,z) est une matrice 2×3 :

$$\nabla f(x,y,z) = \begin{pmatrix} 2xy + \cos x & x^2 + 2z & 2y \\ yz + 2 & xz & xy \end{pmatrix}$$

2. Le gardient de la fonction identité $Id_{\mathbb{R}^n} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ est la matrice identité I_n.

Règle de composition

Soit $D \subset \mathbb{R}^n$, $U \subset \mathbb{R}^m$ deux ouverts, $f : D \longrightarrow \mathbb{R}^m$ et $g : U \longrightarrow \mathbb{R}^l$ deux fonctions de classe C^1. Supposons que $f(D) \subset U$. Alors la fonction composée

$$g \circ f : D \longrightarrow \mathbb{R}^l$$

est de classe C^1 et son gradient est donné par

$$\underbrace{\nabla(g \circ f)(a)}_{l \times n} = \underbrace{\nabla g(f(a))}_{l \times m} \cdot \underbrace{\nabla f(a)}_{m \times n} \qquad \text{pour tout } a \in D.$$

Cette formule s'écrit aussi

$$\boxed{\nabla(g \circ f) = [\nabla g \ \circ f] \cdot \nabla f.}$$

Remarques
1. Si $m = n = l = 1$ on retrouve la règle $(g \circ f)'(x) = g'(f(x)) \cdot f'(x)$

2. Si $n = 1$ (f est une courbe) et $l = 1$ (g est une fonction réelle), on retrouve la règle déjà vue

$$(g \circ f)'(t) = \nabla g(f(t)) \cdot \underbrace{\dot{f}(t)}_{=\nabla f(t)} .$$

3. Dans le cas $l = 1$, on peut réécrire la règle précédente comme suit :
soit $g(x_1, \ldots, x_m)$ une fonction de $\mathbb{R}^m$ dans $\mathbb{R}$. Posons

$$\begin{cases} x_1 = f_1(u_1, \ldots, u_n) \\ x_2 = f_2(u_1, \ldots, u_n) \\ \vdots \\ x_m = f_m(u_1, \ldots, u_n) \end{cases}$$

Si l'on note $\mathrm{x} = f(u)$, alors la formule précédente s'écrit

$$\Big(D_{u_1}(g \circ f) \ \ D_{u_2}(g \circ f) \cdots D_{u_n}(g \circ f)\Big)$$

$$= (D_{x_1}g \ \ D_{x_2}g \cdots D_{x_m}g) \cdot \begin{pmatrix} D_{u_1}f_1 & D_{u_2}f_1 & \cdots & D_{u_n}f_1 \\ D_{u_1}f_2 & D_{u_2}f_2 & \cdots & D_{u_n}f_2 \\ \vdots & \vdots & \ddots & \vdots \\ D_{u_1}f_m & D_{u_2}f_m & \cdots & D_{u_n}f_m \end{pmatrix}$$

ce qui donne pour tout $i = 1, 2, \ldots, n$:

$$\begin{aligned}\frac{\partial g \circ f}{\partial u_i}(u) &= \frac{\partial g}{\partial x_1}(\mathrm{x}) \cdot \frac{\partial x_1}{\partial u_i}(u) + \frac{\partial g}{\partial x_2}(\mathrm{x}) \cdot \frac{\partial x_2}{\partial u_i}(u) + \cdots + \frac{\partial g}{\partial x_m}(\mathrm{x}) \cdot \frac{\partial x_m}{\partial u_i}(u) \\ &= \sum_{k=1}^{m} \frac{\partial g}{\partial x_k}(f(u)) \cdot \frac{\partial x_k}{\partial u_i}(u)\end{aligned}$$

Notation physique

Si

$$L = L(x, y, z) \qquad \text{et} \qquad \left\{ \begin{array}{l} x = x(u, t) \\ y = y(u, t) \\ z = z(u, t) \end{array} \right.$$

on notera, par abus :

$$\frac{\partial L}{\partial u} = L_u = L_x \cdot x_u + L_y \cdot y_u + L_z \cdot z_u$$

et

$$L_t = L_x \cdot x_t + L_y \cdot y_t + L_z \cdot z_t$$

Fonction inverse et gradient

Soit

$$h : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

une fonction bijective et $h^{-1} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ sa fonction réciproque. On suppose que h et h^{-1} sont dérivables. Alors, comme $h \circ h^{-1} = \mathrm{id}_{\mathbb{R}^n}$, la règle de composition donne

$$\nabla h(h^{-1}(y)) \cdot \nabla h^{-1}(y) = I_n$$

ou encore

$$\nabla h^{-1}(y) = \left[\nabla h(h^{-1}(y))\right]^{-1}$$

Le gradient de h^{-1} est l'inverse du gradient de h.

Exemple : soit $h : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ définie par

$$h(\rho, \varphi) = \left(\begin{array}{c} \rho \cos \varphi \\ \rho \sin \varphi \end{array} \right)$$

C'est la transformation entre coordonnées cartésiennes et coordonnées polaires. Alors en se restreignant à un ouvert D avec $\rho > 0$ (par exemple $\varphi \in]0, \pi/2[$), la fonction h est bijective. Son inverse est donné par

$$h^{-1}(x, y) = \left(\begin{array}{c} \rho \\ \varphi \end{array} \right) = \left(\begin{array}{c} \sqrt{x^2 + y^2} \\ \arctan\left(\frac{y}{x}\right) \end{array} \right)$$

On a

$$\nabla h = \left(\begin{array}{cc} x_\rho & x_\varphi \\ y_\rho & y_\varphi \end{array} \right) = \left(\begin{array}{cc} \cos \varphi & -\rho \sin \varphi \\ \sin \varphi & \rho \cos \varphi \end{array} \right)$$

et

$$\nabla h^{-1} = \left(\begin{array}{cc} \rho_x & \rho_y \\ \varphi_x & \varphi_y \end{array} \right) = \left(\begin{array}{cc} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{array} \right)$$

Si on écrit ∇h^{-1} en fonction de ρ et φ, on obtient

$$\nabla h^{-1} = \begin{pmatrix} \frac{x}{\rho} & \frac{y}{\rho} \\ -\frac{y}{\rho^2} & \frac{x}{\rho^2} \end{pmatrix} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\frac{\sin\varphi}{\rho} & \frac{\cos\varphi}{\rho} \end{pmatrix}$$

et on vérifie que c'est bien la matrice inverse de ∇h.

7.8.2 Application : changement de coordonnées

Coordonnées polaires

Soient

$$\begin{cases} x = x(\rho,\varphi) = \rho\cos\varphi \\ y = y(\rho,\varphi) = \rho\sin\varphi \end{cases}$$

Considérons la fonction $h : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ définie par

$$h(\rho,\varphi) = (x(\rho,\varphi), y(\rho,\varphi)).$$

La fonction h décrit le changement entre coordonnées polaires et coordonnées cartésiennes. Son gradient est

$$\nabla h = \begin{pmatrix} x_\rho & x_\varphi \\ y_\rho & y_\varphi \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\rho\sin\varphi \\ \sin\varphi & \rho\cos\varphi \end{pmatrix}$$

Soit $f(x,y)$ une fonction réelle à 2 variables.
Posons

$$\tilde{f}(\rho,\varphi) = f(h(\rho,\varphi)) = f(x(\rho,\varphi), y(\rho,\varphi)).$$

C'est la fonction f exprimée en coordonnées polaires. Parfois, par abus de langage, on identifie f et $\tilde{f}$.
On définit également

$$\tilde{f}_x(\rho,\varphi) = f_x(x(\rho,\varphi), y(\rho,\varphi))$$

C'est la dérivée partielle de f selon x mais exprimée en coordonnées polaires. De même, on pose $\tilde{f}_y = f_y \circ h$ ce qui se résume par

$$\nabla_{x,y}\tilde{f} = (\tilde{f}_x \quad \tilde{f}_y) = \nabla f \ \circ h.$$

(c'est le gradient de f selon x et y mais exprimé en fonction de ρ et φ).
On notera $\nabla_{\rho,\varphi}\tilde{f}$ au lieu de $\nabla\tilde{f}$ pour plus de clarté.
Avec ces notations la règle de composition

$$\nabla\tilde{f} = [\nabla f \ \circ h] \cdot \nabla h = \nabla_{x,y}\tilde{f} \cdot \nabla h$$

devient

$$\boxed{\nabla_{\rho,\varphi}\tilde{f} = \nabla_{x,y}\tilde{f} \cdot \nabla h}$$

qui donne

$$(\tilde{f}_\rho \quad \tilde{f}_\varphi) = (\tilde{f}_x \quad \tilde{f}_y) \cdot \underbrace{\begin{pmatrix} \cos\varphi & -\rho\sin\varphi \\ \sin\varphi & \rho\cos\varphi \end{pmatrix}}_{=\nabla h}$$

La matrice ∇h est inversible si $\rho \neq 0$, car $\det(\nabla h) = \rho$. En multipliant à droite par ∇h^{-1}, on obtient

$$\begin{aligned}(\tilde{f}_x \;\; \tilde{f}_y) &= (\tilde{f}_\rho \;\; \tilde{f}_\varphi) \cdot \frac{1}{\rho} \cdot \begin{pmatrix} \rho\cos\varphi & \rho\sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \\ &= \begin{pmatrix} \cos(\varphi)\cdot\tilde{f}_\rho - \tilde{f}\frac{1}{\rho}\sin(\varphi)\cdot\tilde{f}_\varphi & \sin(\varphi)\cdot\tilde{f}_\rho + \frac{1}{\rho}\cos(\varphi)\cdot\tilde{f}_\varphi \end{pmatrix}\end{aligned}$$

ce qui donne

$$\boxed{\begin{aligned}\tilde{f}_x &= \cos\varphi\cdot\tilde{f}_\rho - \tfrac{1}{\rho}\sin\varphi\cdot\tilde{f}_\varphi \\ \tilde{f}_y &= \sin\varphi\cdot\tilde{f}_\rho + \tfrac{1}{\rho}\cos\varphi\cdot\tilde{f}_\varphi\end{aligned}} \tag{7.4}$$

Ces 2 formules permettent donc de calculer les dérivées partielles selon x et y d'une fonction donnée en coordonnées polaires.

Exemple 7.48. Considérons la fonction $\tilde{f}(\rho, \varphi) = 2\rho^2$ en coordonnées polaires. Calculons le gradient de $\tilde{f}$ selon x, y. On a
$\tilde{f}_\rho = 4\rho$ et $\tilde{f}_\varphi = 0$ ce qui donne, en appliquant la formule (7.4)

$$\begin{aligned}f_x = \tilde{f}_x &= \cos\varphi\cdot\tilde{f}_\rho - \frac{1}{\rho}\sin\varphi\cdot\tilde{f}_\varphi = \cos\varphi\cdot 4\rho - 0 = 4\rho\cos\varphi \\ f_y = \tilde{f}_y &= \sin\varphi\cdot\tilde{f}_\rho + \frac{1}{\rho}\cos\varphi\cdot\tilde{f}_\varphi = \sin\varphi\cdot 4\rho\end{aligned}$$

Le gradient est donc $4\rho\cdot(\cos\varphi \quad \sin\varphi)$.

Vérification par calcul direct :

$$f(x, y) = 2\rho^2 = 2(x^2 + y^2).$$

Alors $f_x = 4x = 4\rho\cos\varphi$ et $f_y = 4y = 4\rho\sin\varphi$.

Coordonnées cylindriques

Considérons le changement de coordonnées

$$\begin{cases} x = \rho\cos\varphi \\ y = \rho\sin\varphi \\ z = z \end{cases}$$

Ceci définit une fonction $h : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ en posant $h(\rho, \varphi, z) = (x, y, z)$. Sa matrice jacobienne est

$$\nabla h = \begin{pmatrix} x_\rho & x_\varphi & x_z \\ y_\rho & y_\varphi & y_z \\ z_\rho & z_\varphi & z_z \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\rho\sin\varphi & 0 \\ \sin\varphi & \rho\cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

On a $\det(\nabla h) = \rho$. Ainsi si $\rho \neq 0$ la matrice ∇h est inversible :

$$\nabla h^{-1} = \frac{1}{\rho}\begin{pmatrix} \rho\cos\varphi & \rho\sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & \rho \end{pmatrix}.$$

Soit $f(x,y,z)$ une fonction à 3 variables et $\tilde{f}(\rho,\varphi,z)$ la fonction correspondante en coordonnées cylindriques. En posant comme précédemment

$$\nabla_{x,y,z}\tilde{f} = \nabla f \circ h$$

et en notant $\nabla_{\rho,\varphi,z}\tilde{f}$ au lieu de $\nabla\tilde{f}$, on obtient

$$\nabla\tilde{f} = [\nabla f \circ h]\cdot\nabla h = \nabla_{x,y,z}\tilde{f}\cdot\nabla h.$$

ou encore

$$\boxed{\nabla_{x,y,z}\tilde{f} = \nabla_{\rho,\varphi,z}\tilde{f}\cdot\nabla h^{-1}}$$

ce qui donne

$$\boxed{\begin{array}{l} \tilde{f}_x = \tilde{f}_\rho\cos\varphi - \tilde{f}_\varphi\,\frac{1}{\rho}\,\sin\varphi \\ \tilde{f}_y = \tilde{f}_\rho\sin\varphi + \tilde{f}_\varphi\,\frac{1}{\rho}\,\cos\varphi \\ \tilde{f}_z = \tilde{f}_z \end{array}}$$

Coordonnées sphériques

Considérons le changement de coordonnées :

$$\left\{\begin{array}{l} x = r\,\cos\varphi\,\sin\theta \\ y = r\,\sin\varphi\,\sin\theta \\ z = r\cos\theta \end{array}\right. \qquad r>0,\ 0\leqslant\varphi\leqslant 2\pi$$

Ceci définit une fonction $h:\mathbb{R}^3 \longrightarrow \mathbb{R}^3$ en posant $h(r,\varphi,\theta)=(x,y,z)$. Sa matrice jacobienne est

$$\nabla h = \begin{pmatrix} x_r & x_\varphi & x_\theta \\ y_r & y_\varphi & y_\theta \\ z_r & z_\varphi & z_\theta \end{pmatrix} = \begin{pmatrix} \cos\varphi\,\sin\theta & -r\sin\varphi\,\sin\theta & r\,\cos\varphi\,\cos\theta \\ \sin\varphi\,\sin\theta & r\cos\varphi\,\sin\theta & r\,\sin\varphi\,\cos\theta \\ \cos\theta & 0 & -r\sin\theta \end{pmatrix}.$$

On a $\det(\nabla h) = r^2\sin\theta$. Ainsi si $r\neq 0$ et $\theta\neq\pm\pi$, la matrice ∇h est inversible :

$$\nabla h^{-1} = \frac{1}{r^2\sin\theta}\begin{pmatrix} r\cos\varphi & r\sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & r \end{pmatrix}$$

et on calcule, comme ci-dessus

$$\boxed{\nabla_{x,y,z}\tilde{f} = \nabla_{r,\varphi,\theta}\tilde{f}\cdot\nabla h^{-1}.}$$