

Analyse pour l'économie 2

Lorenzo Brandolese

# 1 Dérivées partielles et différentiabilité des fonctions de plusieurs variables

## 1.1 Dérivées partielles

Soit  $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  et  $x_0 \in \overset{\circ}{D}$ . Une *direction dans*  $\mathbb{R}^n$  est un vecteur  $v \in \mathbb{R}^n$  tel que  $\|v\| = 1$ . La droite  $d$  passant par  $x_0$  et de direction  $v$  est l'ensemble des point  $d = \{x_0 + tv \in \mathbb{R}^n : t \in \mathbb{R}\}$ . Si l'on restreint la fonction  $f$  à cette droite on obtient une fonction d'une seule variable (la variable  $t \in \mathbb{R}$ ) :  $\phi(t) = f(x_0 + tv)$ . Le calcul de  $\phi'(t)$  nous renseigne sur le taux d'accroissement de  $f$  le long de cette droite. Par définition, la *dérivée de  $f$  le long de la direction  $v$  en  $x_0$*  est  $\phi'(0)$  (à condition que cette dérivée existe). Autrement dit :

$$\frac{\partial f}{\partial v}(x_0) := \lim_{t \rightarrow 0} \frac{f(x_0 + tv) - f(x_0)}{t}.$$

Si la limite ci-dessus existe on dira que  $f$  est dérivable en  $x_0$  le long de la direction  $v$ .

**Exemple 1.1.** Considérons dans  $\mathbb{R}^3$  la fonction  $f(x, y, z) = xyz$ . Soit  $v$  la direction donnée par le vecteur de norme unitaire  $v = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)$ . Au point  $(1, 2, 3) \in \mathbb{R}^3$ , on a

$$\frac{\partial f}{\partial v}(1, 2, 3) = \lim_{t \rightarrow 0} \frac{(1 + \frac{t}{\sqrt{2}})(2 + \frac{t}{\sqrt{2}})3 - 6}{t} = \frac{6}{\sqrt{2}}.$$

Dans  $\mathbb{R}^n$ , on note généralement par  $x_1, \dots, x_n$  les  $n$  variables et  $e_1 = (1, 0, \dots, 0), \dots, e_n = (0, \dots, 0, 1)$  les directions correspondant aux  $n$ -axes. Les dérivées directionnelles le long de ces vecteurs s'appellent *dérivées partielles* et on utilise les notations

$$\frac{\partial f}{\partial x_i}(x_0) = f_{x_i}(x_0) = D_i(f)(x_0) = \frac{\partial f}{\partial e_i}(x_0), \quad i = 1, \dots, n.$$

Dans  $\mathbb{R}^2$  et  $\mathbb{R}^3$  on préfère souvent appeler  $x, y$  et respectivement  $x, y, z$  les variables. On utilise alors, par exemple, la notation

$$\frac{\partial f}{\partial x}(x_0, y_0) = f_x(x_0, y_0) = D_x(f)(x_0, y_0)$$

pour la première dérivée partielle d'une fonction de deux variables au point  $(x_0, y_0) \in \mathbb{R}^2$ .

**Exemple 1.2.** Calculons les dérivées partielles  $f_x(x, y)$  et  $f_y(x, y)$  en tout point  $(x, y) \in \mathbb{R}^2$ , pour la fonction  $f(x, y) = x \sin(xy)$  :

$$f_x(x, y) = \sin(xy) + xy \cos(xy), \quad \text{et} \quad f_y(x, y) = x^2 \cos(xy).$$

**Remarque 1.3.** L'existence des dérivées partielles d'une fonction en un point ne garantit pas la continuité de la fonction en ce point. Considerons par exemple la fonction  $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  définie par

$$f(x, y) = \left( \frac{x^2 y}{x^4 + y^2} \right)^2, \quad (x, y) \neq (0, 0), \quad f(0, 0) = 0.$$

On peut vérifier que cette fonction possède les dérivées partielles en  $(0, 0)$ , et même des dérivées le long toutes les directions à l'origine. Cependant cette fonction n'est pas continue en  $(0, 0)$  (comme on le voit en étudiant la suite  $f(1/n, 1/n^2)$ ). Cela est en contraste avec les fonctions réelles d'une seule variable, pour lesquelles la dérivabilité implique la continuité.

Ce n'est donc pas l'existence des dérivées partielles qui implique la continuité d'une fonction, mais la notion plus restrictive de différentiabilité, définie ci-après.

## 1.2 Fonctions différentiables (cas des fonctions scalaires de $n$ variables)

**Le cas des fonctions d'une seule variable. Rappels.** Rappelons qu'une fonction  $f: \mathbb{R} \rightarrow \mathbb{R}$  est dérivable en un point  $x_0 \in \mathbb{R}$  si et seulement s'il existe un nombre réel, noté  $f'(x_0)$  (et appelé dérivée de  $f$  en  $x_0$ ) tel que

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0) - f'(x_0)h}{h} = 0 \quad (\text{ici } h \in \mathbb{R}).$$

Pour les fonctions  $f: \mathbb{R} \rightarrow \mathbb{R}$  dérivables en un point  $x_0$ , on peut alors donner un sens à la notion de *droite tangente* au graphe de la fonction  $f$  en  $x_0$  : il s'agit de la droite

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

Généralisons cette idée aux fonctions de plusieurs variables.

**Le cas des fonctions de plusieurs variables.** Rappelons qu'une *application linéaire*  $L: \mathbb{R}^n \rightarrow \mathbb{R}$  est une application de la forme

$$L(x) = a_1x_1 + \dots + a_nx_n, \quad \text{pour tout } x = (x_1, \dots, x_n) \in \mathbb{R}^n,$$

où  $a_1, \dots, a_n$  sont des nombres réels. Le graphe d'une telle application est une droite de  $\mathbb{R}^2$  (si  $n = 1$ ) ou un plan de  $\mathbb{R}^3$  (si  $n = 2$ ) passant par l'origine. Plus en général, le graphe de  $L$  est un *hyperplan* de  $\mathbb{R}^{n+1}$  passant par l'origine :  $\{x \in \mathbb{R}^{n+1} : x_{n+1} = L(x_1, \dots, x_n) : x \in \mathbb{R}^n\}$ .

Une *fonction affine* :  $\mathbb{R}^n \rightarrow \mathbb{R}$  est une fonction de la forme

$$x \mapsto \alpha + L(x), \quad x \in \mathbb{R}^n,$$

où  $L$  est une application linéaire. L'addition du terme  $\alpha \in \mathbb{R}$  ayant l'effet d'une translation, le graphe des applications affines sont des hyperplans de  $\mathbb{R}^{n+1}$  (ne passant pas par l'origine en général, mais plutôt par le point  $(0, \dots, 0, \alpha)$ ).

**Définition 1.1.** Soit  $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  et  $x_0 \in \overset{\circ}{D}$ . On dit que  $f$  est différentiable au point  $x_0$  s'il existe une application linéaire  $L$  (dépendante de  $x_0$ ) telle que

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0) - L(h)}{\|h\|} = 0 \quad (\text{ici } h \in \mathbb{R}^n).$$

Si  $f$  est différentiable en tout point  $x_0 \in D$  on dit que  $f$  est différentiable en  $D$ . L'application  $L$  s'appelle différentielle de  $f$  en  $x_0$ . Cette application linéaire  $L$  est notée  $df_{x_0}$ .

Il est utile de comparer les deux notions suivantes :

$$f \text{ continue en } x_0 \iff f(x_0 + h) = f(x_0) + o(1) \text{ pour } \|h\| \rightarrow 0.$$

$$f \text{ différentiable en } x_0 \iff f(x_0 + h) = f(x_0) + df_{x_0}(h) + o(\|h\|) \text{ pour } \|h\| \rightarrow 0.$$

Ici  $o(1)$  désigne une fonction (de la variable  $h \in \mathbb{R}^n$ ) de limite nulle pour  $\|h\| \rightarrow 0$ , et  $o(\|h\|)$  désigne une fonction  $\varepsilon(h)$  telle que  $\lim_{h \rightarrow 0} \frac{\varepsilon(h)}{\|h\|} = 0$ .

**Proposition 1.1.** Soit  $f$  une fonction différentiable en  $x_0$ . Alors  $f$  possède de dérivées le long toute direction et

$$\frac{\partial f}{\partial v}(x_0) = df_{x_0}(v) = \sum_{i=1}^n v_i \frac{\partial f}{\partial x_i}(x_0). \quad (1.1)$$

*Dém.* Soit  $v \in \mathbb{R}^n$ , une direction, On a, grâce à la linéarité de  $df_{x_0}$ ,  $df_{x_0}(tv) = t df_{x_0}(v)$ . Ainsi,

$$\frac{f(x_0 + tv) - f(x_0)}{t} = \frac{df_{x_0}(tv) + o(\|tv\|)}{t} = df_{x_0}(v) + o(1), \quad \text{pour } t \rightarrow 0.$$

Mais alors, en prenant  $t \rightarrow 0$  dans cette expression, on trouve  $\frac{\partial f}{\partial v}(x_0) = df_{x_0}(v)$ . En développant  $v$  avec la base canonique,  $v = \sum_{i=1}^n v_i e_i$ , la dernière égalité de (1.1) en découle.  $\square$

Si  $f$  est différentiable en  $x_0$  la différentielle de  $f$  en  $x_0$  est unique, comme on le voit grâce à l'équation (1.1).

Pour une fonction différentiable en  $x_0$ , le graphe de la fonction  $h \mapsto f(x_0) + df_{x_0}(h)$  étant un hyperplan, la notion de différentiabilité signifie précisément qu'au voisinage de  $(x_0, f(x_0)) \in \mathbb{R}^{n+1}$ , le graphe de la fonction  $f$  est convenablement approché par un hyperplan. Si l'on observe que

$$f \text{ différentiable en } x_0 \iff f(x) = f(x_0) + df_{x_0}(x - x_0) + o(\|x - x_0\|) \text{ pour } x \rightarrow x_0,$$

cela conduit à la définition suivante.

**Définition 1.2.** Pour une fonction différentiable en  $x^0 \in \overset{\circ}{D}$ , l'hyperplan tangent au graphe de la fonction  $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  est l'hyperplan d'équation

$$\left\{ x \in \mathbb{R}^{n+1} : x_{n+1} = f(x^0) + \sum_{i=1}^n (x_i - x_i^0) \frac{\partial f}{\partial x_i}(x_0) \right\}.$$

**Exemple 1.4.** Construisons, dans  $\mathbb{R}^3$ , le plan tangent au graphe de la fonction de deux variables  $f(x, y) = x^2 + \sin y + 10$  au point  $(1, 0)$ . On a  $f_x(1, 0) = 2$  et  $f_y(1, 0) = 1$ . De plus,  $f(1, 0) = 11$ . Le plan cherché est donc le plan d'équation  $z = 2(x - 1) + y + 11$ , ou encore  $z = 2x + y + 9$ .

**Exemple 1.5.** La différentielle  $dL_{x_0}$  d'une application linéaire  $L: \mathbb{R}^n \rightarrow \mathbb{R}$  est indépendante de  $x_0$  et coïncide avec l'application  $L$ . En particulier, la différentielle de l'application linéaire (de "projection sur la  $i$ -ème composante")  $x \mapsto x_i$ , que l'on note abusivement  $x_i$ , coïncide avec l'application elle même). Cela explique pourquoi on note  $dx_i$  l'application

$$dx_i: \mathbb{R}^n \rightarrow \mathbb{R}, \quad \text{où } \forall h \in \mathbb{R}^n, dx_i(h) = h_i.$$

Pour une fonction  $f$ , la différentielle de  $f$ , notée  $df$  est l'application  $x_0 \mapsto df_{x_0}$ . C'est pourquoi l'on peut écrire l'égalité d'applications linéaires

$$\forall x_0 \quad df(x_0) = df_{x_0} = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_0) dx_i,$$

(si on applique cette formule à  $v \in \mathbb{R}^n$  on trouve exactement la formule (1.1)) ou en abrégé

$$df = \sum_{i=1}^n \frac{\partial f}{\partial x_i} dx_i.$$

**Définition 1.3** (Gradient). Pour une fonction  $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ , admettant toutes les dérivées partielles en  $x_0 \in \overset{\circ}{D}$ , le gradient de  $f$  en  $x_0$ , noté  $\nabla f(x_0)$  est le vecteur de  $\mathbb{R}^n$  de composantes

$$\nabla f(x_0) = \left( \frac{\partial f}{\partial x_1}(x_0), \dots, \frac{\partial f}{\partial x_n}(x_0) \right)^T.$$

Le gradient est par convention un vecteur colonne, c'est pourquoi on prend ici la transposition du vecteur  $(\frac{\partial f}{\partial x_1}(x_0), \dots, \frac{\partial f}{\partial x_n}(x_0))$ .

Observons que la relation (1.1) s'écrit

$$df_{x_0}(v) = \frac{\partial f}{\partial v}(x_0) = \langle \nabla f(x_0), v \rangle. \quad (1.2)$$

Autrement dit, la dérivée le long d'une direction  $v$  s'obtient en prenant le produit scalaire entre le vecteur gradient et le vecteur  $v$ .

Pour une fonction  $f$  et un point  $x_0$  donné, calculons  $\frac{\partial f}{\partial v}(x_0)$  le long différentes directions  $v$  : d'après (1.2) et l'inégalité de Cauchy-Schwarz (et le fait que  $\|v\| = 1$ ),

$$-\|\nabla f(x_0)\| \leq \frac{\partial f}{\partial v}(x_0) \leq \|\nabla f(x_0)\|.$$

Les égalités sont possibles lorsque  $v$  est parallèle au vecteur  $\nabla f(x_0)$ . Plus précisément, si le vecteur gradient ne s'annule pas en  $x_0$ ,

$$\begin{aligned} v = \nabla f(x_0) / \|\nabla f(x_0)\| &\Rightarrow \frac{\partial f}{\partial v}(x_0) = \|\nabla f(x_0)\|, \\ v = -\nabla f(x_0) / \|\nabla f(x_0)\| &\Rightarrow \frac{\partial f}{\partial v}(x_0) = -\|\nabla f(x_0)\|. \end{aligned}$$

Mais  $\frac{\partial f}{\partial v}(x_0)$  exprime la pente du graphe de  $f$  le long de la direction  $v$ . Donc, au point  $x_0$  la pente de la surface qui représente la fonction  $f$  sera maximale (et positive) dans la direction du gradient, minimale (et négative) si on prend l'orientation opposée à  $\nabla f(x_0)$ .

Par la formule (1.2) on définit

$$\frac{\partial f}{\partial w}(x_0) = \langle \nabla f(x_0), w \rangle = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_0) w_i$$

pour tout vecteur de  $\mathbb{R}^n$ ,  $w \neq 0$  (non nécessairement de norme = 1).

**Théorème 1.2** (de la différentielle totale). Si  $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  est une application admettant toutes les dérivées partielles au voisinage d'un point  $x^0$  intérieur à  $D$  et que celles-ci sont continues en  $x_0$ , alors  $f$  est différentiable en  $x^0$ .

*Dém.* (dans le cas des fonctions de deux variables.) Pour  $x = (x_1, x_2)$  et  $x^0 = (x_1^0, x_2^0)$ , en appliquant le théorème des accroissements finis, on peut trouver un réel  $\xi_1$  compris entre  $x_1^0$  et  $x_1$  et un réel  $\xi_2$  compris entre  $x_2$  et  $x_2^0$  tels que

$$\begin{aligned} f(x) - f(x^0) &= f(x_1, x_2) - f(x_1^0, x_2) + f(x_1^0, x_2) - f(x_1^0, x_2^0) \\ &= D_1(f)(\xi_1, x_2)(x_1 - x_1^0) + D_2(f)(x_1^0, \xi_2)(x_2 - x_2^0). \end{aligned}$$

Mais alors

$$\begin{aligned} f(x) - f(x^0) - D_1(f)(x^0)(x_1 - x_1^0) - D_2(f)(x^0)(x_2 - x_2^0) \\ &= [D_1(f)(\xi_1, x_2) - D_1(f)(x^0)](x_1 - x_1^0) + [D_2(f)(x_1^0, \xi_2) - D_2(f)(x^0)](x_2 - x_2^0) \\ &= o(\|x - x^0\|) \quad \text{pour } x \rightarrow x_0, \end{aligned}$$

car  $D_1(f)(\xi_1, x_2) \rightarrow D_1(f)(x^0)$  et  $D_2(f)(x_1^0, \xi_2) \rightarrow D_2(f)(x^0)$  grâce à la continuité des dérivées partielles de  $f$ , ce qui implique que les deux termes dans les crochets sont de limite nulle.  $\square$

### 1.3 Dérivées successives

Notons, pour une fonction définie sur un ouvert de  $\mathbb{R}^2$ ,  $\frac{\partial^2 f}{\partial x \partial y} = (f_y)_x$  et  $\frac{\partial^2 f}{\partial y \partial x} = (f_x)_y$ .

**Théorème 1.3** (Théorème de Schwarz. Démonstration hors programme). *Soit  $U$  un ouvert de  $\mathbb{R}^2$  et  $f: U \rightarrow \mathbb{R}$  et  $(x_0, y_0) \in U$ . Si les dérivées mixtes  $\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)$  et  $\frac{\partial^2 f}{\partial y \partial x}(x_0, y_0)$  existent dans  $U$  et sont continues dans  $(x_0, y_0)$ , alors elles coïncident.*

Plus en général, si  $f$  est une fonction définie sur un ouvert  $U$  de  $\mathbb{R}^n$ , on dit qu'elle est de classe  $C^k$  dans  $U$  si toutes les dérivées itérées de la fonction  $f$  existent et sont continues dans  $U$ , jusqu'à l'ordre  $k$ . On peut généraliser le théorème de Schwarz de la manière suivante : pour une fonction de classe  $C^k$  toutes les dérivées itérées jusqu'à l'ordre  $k$  coïncident.

### 1.4 La différentielle des fonctions $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Matrice jacobienne

Désignons par  $(e_1, \dots, e_n)$  et  $(e_1, \dots, e_m)$  les bases canoniques de  $\mathbb{R}^n$  et  $\mathbb{R}^m$ . Rappelons qu'à toute application linéaire  $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$  on associe une (et une seule) matrice  $A$ , de taille,  $m \times n$ , par la relation

$$L(x) = Ax, \quad \text{où} \quad Ax = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n a_{1j}x_j \\ \vdots \\ \sum_{j=1}^n a_{mj}x_j \end{pmatrix}$$

Si  $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , où  $L = (L_1, \dots, L_m)$  est une application linéaire, la matrice  $m \times n$  qui représente  $L$  est la matrice où le vecteur  $L(e_j) \in \mathbb{R}^m$ , a pour composantes  $a_{ij} = L_i(e_j)$ , pour  $i = 1, \dots, m$ .

**Exercice 1.6.** Démontrer que, pour toute matrice  $A$  de taille  $m \times n$ , il existe une constante  $K \geq 0$  telle que  $\|Ax\| \leq K\|x\|$ . Conclure que, si  $L(x) = Ax$ , alors l'application linéaire  $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$  est  $K$ -lipschitzienne (et donc continue).\*

**Définition 1.4.** Soit  $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  et  $x_0 \in \overset{\circ}{D}$ . On dit que  $f$  est différentiable au point  $x_0$  s'il existe une application linéaire,  $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$  telle que

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0) - L(h)}{\|h\|} = 0 \quad (\text{ici } h \in \mathbb{R}^n).$$

L'application  $L$  s'appelle différentielle de  $f$  en  $x_0$ . Cette application linéaire  $L$  est notée  $df_{x_0}$ . Ainsi,  $df_{x_0}: \mathbb{R}^n \rightarrow \mathbb{R}^m$ .

On peut aussi écrire la limite précédente par

$$f(x + h) = f(x_0) + df_{x_0}(h) + o(\|h\|), \quad \text{pour } h \rightarrow 0. \quad (1.3)$$

Observons que  $f = (f_1, \dots, f_m)$  et  $L = (L_1, \dots, L_m)$ , les composantes de  $L$  étant des applications linéaires :  $\mathbb{R}^n \rightarrow \mathbb{R}$ . De plus, la limite précédente (qui est une limite dans  $\mathbb{R}^m$ ) est nulle si et seulement si elle est nulle composante par composante. Autrement dit,  $f$  est différentiable si et seulement si ses composantes  $f_j: \mathbb{R}^n \rightarrow \mathbb{R}$  sont des applications différentiables. On peut aussi expliciter les composantes de  $df_{x_0}: \mathbb{R}^n \rightarrow \mathbb{R}^m$  en écrivant

$$df_{x_0} = (d(f_1)_{x_0}, \dots, d(f_m)_{x_0}).$$

---

\*. Solution.  $|(Ax)_i| = |\sum_{j=1}^n a_{ij}x_j| \leq \sqrt{a_{i1}^2 + \dots + a_{in}^2} \|x\|$  par l'inégalité de Cauchy-Schwarz. Mais alors  $\|Ax\| = \sqrt{\sum_{i=1}^m (Ax)_i^2} \leq \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2} \|x\|$ . Il suffit alors de prendre  $K = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}$ . Maintenant, d'après la linéarité de  $L$ ,  $\|L(x) - L(y)\| = \|L(x - y)\| \leq K\|x - y\|$ . Cela montre que  $L$  est lipschitzienne.

Appliquons ceci à  $df_{x_0}$ . Il s'agit de calculer  $d(f_i)_{x_0}(e_i)$ . Mais  $d(f_i)_{x_0}(e_j) = \frac{\partial f_i}{\partial x_j}(x_0)$  d'après (1.1). En conclusion, la matrice représentative de la différentielle d'une application  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  (en un point  $x_0$  donné) est la *matrice jacobienne*, donnée par :

$$J_f(x_0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \cdots & \frac{\partial f_1}{\partial x_n}(x_0) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \cdots & \frac{\partial f_m}{\partial x_n}(x_0) \end{pmatrix}.$$

Ainsi, la formule (1.1) se généralise aux fonctions à valeurs vectorielles par

$$df_{x_0}(v) = J_f(x_0)v, \quad (1.4)$$

le vecteur à droite dans (1.4) étant le produit entre la matrice Jacobienne de  $f$  en  $x_0$  et le vecteur  $v$  (vu comme vecteur colonne).

Si  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  est différentiable en un point  $x_0 \in \mathbb{R}^n$ , observons que l'on a l'inégalité suivante qui découle de l'exercice 1.6 :

$$\exists K \geq 0 \text{ t.q. } \forall v \in \mathbb{R}^n, \quad \|df_{x_0}(v)\| \leq K\|v\|. \quad (1.5)$$

Le fait de choisir une norme non euclidienne a pour seule conséquence de modifier la valeur de la constante  $M$ .

**Théorème 1.4.** Soient  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  et  $g: \mathbb{R}^m \rightarrow \mathbb{R}^k$  deux fonctions telles que  $f$  est différentiable en  $x_0 \in \mathbb{R}^n$  et  $g$  est différentiable en  $y_0 = f(x_0) \in \mathbb{R}^m$ . Alors la fonction composée  $F = g \circ f: \mathbb{R}^n \rightarrow \mathbb{R}^k$  est différentiable en  $x_0$ . De plus, on a l'égalité (entre applications linéaires :  $\mathbb{R}^n \rightarrow \mathbb{R}^k$ ) :

$$dF_{x_0} = dg_{y_0} \circ df_{x_0} \quad (1.6)$$

et les matrices jacobiennes correspondantes vérifient la relation

$$J_F(x_0) = J_g(y_0) \cdot J_f(x_0). \quad (1.7)$$

Démonstration \* hors programme.

\*. On a

$$\begin{aligned} f(x_0 + h) &= f(x_0) + df_{x_0}(h) + o(\|h\|), & \text{pour } h \rightarrow 0, \\ g(y_0 + k) &= g(y_0) + dg_{y_0}(k) + o(\|k\|), & \text{pour } k \rightarrow 0. \end{aligned}$$

Et, pour  $h \rightarrow 0$ ,

$$\begin{aligned} F(x_0 + h) &= g(f(x_0 + h)) = g\left(\underbrace{f(x_0) + df_{x_0}(h) + o(\|h\|)}_{:=k(h)}\right) \\ &= g(y_0) + dg_{y_0}(k(h)) + o(\|k(h)\|) \\ &= g(y_0) + dg_{y_0}(df_{x_0}(h)) + dg_{y_0}(o(\|h\|)) + o(\|k(h)\|), & \text{(par linéarité de } dg_{y_0}) \\ &= g(y_0) + dg_{y_0}(df_{x_0}(h)) + o(\|h\|). \end{aligned} \quad (1.8)$$

Détaillons la dernière égalité : il s'agit de démontrer que  $dg_{y_0}(o(h)) + o(k(h)) = o(h)$  pour  $h \rightarrow 0$ , c'est à dire,

$$\lim_{h \rightarrow 0} \frac{dg_{y_0}(o(h)) + o(\|k(h)\|)}{\|h\|} = 0.$$

En effet, en appliquant l'inégalité (1.5) (avec  $g$  à la place de  $f$ ,  $y_0$  à la place de  $x_0$  et  $M'$  à la place de  $M$ ) on trouve, pour  $h \rightarrow 0$ ,

$$\frac{\|dg_{y_0}(o(h))\|}{\|h\|} \leq M' \frac{\|o(\|h\|)\|}{\|h\|} \rightarrow 0.$$

**Exemple 1.7.** Soit  $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  la fonction  $f(x, y) = (x^2, y^2)$  et  $g: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  la fonction  $g(u, v) = (e^{uv}, uv)$ . Écrivons d'abord la matrice jacobienne de la fonction composée  $g \circ f$  en appliquant le théorème 1.4. Retrouvons ensuite le même résultat en explicitant la fonction composée  $g \circ f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ .

On a

$$J_f(x, y) = \begin{pmatrix} 2x & 0 \\ 0 & 2y \end{pmatrix}, \quad J_g(u, v) = \begin{pmatrix} ve^{uv} & ue^{uv} \\ v & u \end{pmatrix}$$

Donc,

$$J_{g \circ f}(x, y) = J_g(x^2, y^2)J_f(x, y) = \begin{pmatrix} y^2 e^{x^2 y^2} & x^2 e^{x^2 y^2} \\ y^2 & x^2 \end{pmatrix} \begin{pmatrix} 2x & 0 \\ 0 & 2y \end{pmatrix} = \begin{pmatrix} 2xy^2 e^{x^2 y^2} & x^2 e^{x^2 y^2} \\ y^2 & 2x^2 y \end{pmatrix}.$$

La seconde méthode est plus rapide : on a  $(g \circ f)(x, y) = (e^{x^2 y^2}, x^2 y^2)$ , comme on le voit en imposant  $u = x^2$  et  $v = y^2$ . On retrouve alors facilement le résultat précédent en calculant les deux dérivées partielles de la dernière fonction vectorielle.

### Cas particuliers.

- i) L'équation (1.7) permet de retrouver la formule classique de la dérivée de la fonction composée pour les fonctions d'une seule variable.

$$\begin{array}{ccccc} \mathbb{R} & \xrightarrow{f} & \mathbb{R} & \xrightarrow{g} & \mathbb{R} \\ t & \mapsto & f(t) & \mapsto & g(f(t)). \end{array}$$

Dans ce cas, la composée est  $F: \mathbb{R} \rightarrow \mathbb{R}$  et  $F(t) = g(f(t))$ . De plus, pour une fonction  $: \mathbb{R} \rightarrow \mathbb{R}$  la matrice jacobienne ( $1 \times 1$ ) est simplement la dérivée de la fonction. Ainsi, l'égalité matricielle (1.7) n'est rien d'autre que l'égalité classique

$$F'(t) = g'(f(t))f'(t).$$

- ii) Considérons un cas un peu plus compliqué :

$$\begin{array}{ccccc} \mathbb{R} & \xrightarrow{f=(f_1, f_2)} & \mathbb{R}^2 & \xrightarrow{g} & \mathbb{R} \\ t & \mapsto & (f_1(t), f_2(t)) & \mapsto & g(f_1(t), f_2(t)). \end{array}$$

Dans ce cas, la composée est  $F: \mathbb{R} \rightarrow \mathbb{R}$ , où  $F(t) = g(f_1(t), f_2(t))$ . Observons qu'ici  $g$  est une fonction de deux variables. Appelons  $(x, y)$  les variables de  $g$ . L'égalité matricielle (1.7) équivaut à l'égalité

$$F'(t) = \frac{\partial g}{\partial x}(f_1(t), f_2(t))f_1'(t) + \frac{\partial g}{\partial y}(f_1(t), f_2(t))f_2'(t).$$

**Exemple 1.8.** Soit  $g = g(x, y)$  une fonction différentiable  $: \mathbb{R}^2 \rightarrow \mathbb{R}$ . Considérons la composition de  $g$  avec la fonction  $t \mapsto (\sin t, t^3)$  et calculons ensuite la dérivée :

$$\frac{d}{dt} [g(\sin t, t^3)] = \frac{\partial g}{\partial x}(\sin t, \ln t) \cos t + 3 \frac{\partial g}{\partial y}(\sin t, \ln t) t^2.$$

De plus,

$$\frac{o(\|k(h)\|)}{\|h\|} = \frac{o(\|k(h)\|)}{\|k(h)\|} \cdot \frac{df_{x_0}(h) + o(\|h\|)}{\|h\|} \rightarrow 0,$$

grâce à (1.5), parce que le premier facteur converge vers 0. Cela prouve la dernière égalité dans (1.8).

Maintenant, (1.8) signifie précisément que  $F$  est différentiable en  $x_0$ , avec différentielle donnée par la formule (1.6). La formule (1.7), découle alors de la propriété bien connue des applications linéaires, que la matrice représentative d'une application linéaire  $L_1 \circ L_2$  (composé de deux applications linéaires  $L_1$  et  $L_2$  est donné par le produit entre la matrice de l'application  $L_1$  avec la matrice de l'application  $L_2$ .  $\square$

iii) Considérons un autre cas :

$$\begin{array}{ccc} \mathbb{R}^2 & \xrightarrow{f=(f_1, f_2)} & \mathbb{R}^2 & \xrightarrow{g} & \mathbb{R} \\ (x, y) & \mapsto & (f_1(x, y), f_2(x, y)) & \mapsto & g(f_1(x, y), f_2(x, y)). \end{array}$$

Dans ce cas, la composée est  $F: \mathbb{R}^2 \rightarrow \mathbb{R}$ , où  $F(x, y) = g(f_1(x, y), f_2(x, y))$ . La fonction  $g$  est une fonction de deux variables. Appelons  $(u, v)$  les variables de  $g$  pour les distinguer des variables  $(x, y)$  de la fonction  $f$ . L'égalité matricielle (1.7) équivaut au système

$$\begin{cases} \frac{\partial F}{\partial x}(x, y) = \frac{\partial g}{\partial u}(f_1(x, y), f_2(x, y)) \frac{\partial f_1}{\partial x}(x, y) + \frac{\partial g}{\partial v}(f_1(x, y), f_2(x, y)) \frac{\partial f_2}{\partial x}(x, y), \\ \frac{\partial F}{\partial y}(x, y) = \frac{\partial g}{\partial u}(f_1(x, y), f_2(x, y)) \frac{\partial f_1}{\partial y}(x, y) + \frac{\partial g}{\partial v}(f_1(x, y), f_2(x, y)) \frac{\partial f_2}{\partial y}(x, y), \end{cases}$$

**Exemple 1.9.** Soit  $g(u, v) = uv$  et  $f(x, y) = (e^x, \sin(x - y))$ . Il y a deux manières de calculer la dérivée partielle de  $F(x, y) = (g \circ f)(x, y)$  par rapport à  $x$ .

On peut calculer directement  $F(x, y) = e^x \sin(x - y)$  et l'on trouve  $\frac{\partial}{\partial x}[F(x, y)] = e^x \sin(x - y) + e^x \cos(x - y)$ . Mais on pouvait aussi observer que  $\frac{\partial g}{\partial u} = v$  et  $\frac{\partial g}{\partial v} = u$ , remplacer  $(u, v) = (e^x, \sin(x - y))$  et ensuite appliquer la première équation du système ci-dessus. On aboutit au même résultat.

## 1.5 L'inégalité des accroissements finis

L'inégalité des accroissements finis pour les fonctions d'une seule variable est bien connue : si  $F: [a, b] \rightarrow \mathbb{R}$  est dérivable, alors  $|F(b) - F(a)| \leq (\sup_{t \in [a, b]} |F'(t)|) |b - a|$ . Il existe un analogue pour les fonctions de plusieurs variables :

**Proposition 1.5.** Soit  $U$  un ouvert de  $\mathbb{R}^n$ ,  $f: U \rightarrow \mathbb{R}$  une fonction de classe  $C^1$  et  $x, x' \in U$ , tels que le segment de  $x$  à  $x'$ , qui par définition est l'ensemble

$$[x, x'] = \{(1 - t)x + tx' : 0 \leq t \leq 1\}$$

est contenu dans  $U$ . Alors

$$|f(x) - f(x')| \leq \left( \sup_{z \in [x, x']} \|\nabla f(z)\| \right) \|x - x'\|.$$

*Dém.* Posons  $F(t) = f((1 - t)x + tx')$ . Alors  $F(0) = f(x)$  et  $F(1) = f(x')$ . De plus  $F$  est de classe  $C^1$  dans l'intervalle  $[0, 1]$  et  $F'(t) = \langle \nabla f((1 - t)x + tx'), x' - x \rangle$ . Par l'inégalité des accroissements finis, appliquée à  $F$ , ensuite par l'inégalité de Cauchy-Schwarz,

$$|f(x) - f(x')| = |F(1) - F(0)| \leq \left( \sup_{t \in [0, 1]} |F'(t)| \right) \leq \sup_{z \in [x, x']} \|\nabla f(z)\| \|x - x'\|.$$

□

## 2 Formule de Taylor et développements limités

Soit  $f$  une fonction de classe  $C^2$  dans un ouvert  $U$  de  $\mathbb{R}^n$  et soit  $x_0 \in U$ , et  $\delta > 0$  tel que  $B(x_0, \delta) \subset U$ . Si  $v \in \mathbb{R}^n$  est une direction et  $0 \leq t \leq \delta$ , on pose :

$$F(t) = f(x_0 + tv).$$



La formule de Taylor centrée en 0 pour la fonction  $F$  d'ordre 2 s'écrit

$$F(t) = F(0) + F'(0)t + \frac{F''(0)}{2!}t^2 + o(t^2), \quad \text{pour } t \rightarrow 0. \quad (2.1)$$

D'autre part,

$$F'(t) = \frac{\partial f}{\partial v}(x_0 + tv) = \sum_{i=1}^n D_i(f)(x_0 + tv) v_i,$$

$$F''(t) = \sum_{i=1}^n \frac{\partial D_i(f)}{\partial v}(x_0 + tv) v_i = \sum_{i,j=1}^n D_i D_j(f)(x_0 + tv) v_i v_j.$$

Cela donne,

$$f(x_0 + tv) = f(x_0) + \sum_{i=1}^n D_i(f)(x_0)tv_i + \sum_{i,j=1}^n \frac{D_i D_j(f)}{2!}(x_0)t^2 v_i v_j + o(t^2) \quad \text{pour } t \rightarrow 0.$$

En posant  $h = tv$  (et donc  $|t| = \|h\|$ ) on obtient, pour  $\|h\| \rightarrow 0$ ,

$$f(x_0 + h) = f(x_0) + \sum_{i=1}^n D_i(f)(x_0)h_i + \frac{1}{2} \sum_{i,j=1}^n D_i D_j(f)(x_0)h_i h_j + o(\|h\|^2).$$

C'est la formule de Taylor d'ordre 2 pour les fonctions de plusieurs variables. Le terme à droite est le *développement limité d'ordre 2* de  $f$  centré en  $x_0$ . En observer la structure typique :

constante + partie linéaire en  $h$  + partie quadratique en  $h$  + reste.

Plus en général, si  $f$  est de classe  $C^k$  au voisinage de  $x_0$ , on pourra écrire la formule de Taylor d'ordre  $k$  pour  $F$  au voisinage de 0 :

$$F(t) = F(0) + F'(0)t + \frac{F''(0)}{2!}t^2 + \dots + \frac{F^{(k)}(0)}{k!}t^k + o(t^k), \quad \text{pour } t \rightarrow 0.$$

D'autre part, en généralisant les formules pour  $F'(t)$  et  $F''(t)$  on trouve

$$F^{(k)}(t) = \sum_{i_1, \dots, i_k=1}^n D_{i_1} D_{i_2} \dots D_{i_k}(f)(x_0 + tv) v_{i_1} v_{i_2} \dots v_{i_k}.$$

On obtient alors, pour  $x \rightarrow x_0$ ,

$$f(x_0 + h) = f(x_0) + \sum_{i=1}^n D_i f(x_0)h_i + \frac{1}{2} \sum_{i,j=1}^n D_i D_j f(x_0)h_i h_j + \dots$$

$$+ \frac{1}{k!} \sum_{i_1, \dots, i_k=1}^n D_{i_1} D_{i_2} \dots D_{i_k} f(x_0) h_{i_1} h_{i_2} \dots h_{i_k} + o(\|h\|^k). \quad (2.2)$$

Le terme à droite est le développement limité d'ordre  $k$  pour  $f$  centré en  $x_0$ .

### 3 Extrema de fonctions de plusieurs variables

#### 3.1 Points stationnaires

**Définition 3.1** (extrema locaux). Soit  $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  et  $x_0 \in D$ . On dit que  $x_0$  est un point de maximum relatif (ou local) pour  $f$  s'il existe un voisinage  $V$  de  $x_0$  tel que  $f(x) \leq f(x_0)$  pour tout  $x \in V \cap D$ . Dans ce cas on dit que la valeur  $f(x_0)$  est un maximum relatif (ou local) de la fonction  $f$ .

On dit que  $x_0$  est un point de minimum relatif (ou local) pour  $f$  s'il existe un voisinage  $V$  de  $x_0$  tel que  $f(x_0) \leq f(x)$  pour tout  $x \in V \cap D$ . Dans ce cas on dit que la valeur  $f(x_0)$  est un minimum relatif (ou local) de la fonction  $f$ .

Si on a la condition plus forte  $f(x) \leq f(x_0)$  pour tout  $x \in D$  on dit que  $x_0$  est un (point de) maximum absolu, ou global, pour  $f$  dans  $D$ . De même, si  $f(x_0) \leq f(x)$  pour tout  $x \in D$  on dit que  $x_0$  est un point de minimum absolu, ou global, pour  $f$  dans  $D$ .

Les points de minimum s'appellent aussi *minimiseurs* et les points de maximum *maximiseurs*.

#### Exemple 3.1.

(i) Étudions les extrema locaux et globaux de la fonction  $f: \mathbb{R} \rightarrow \mathbb{R}$ , où  $f(x) = (x^2 - 1)^2$  sur  $\mathbb{R}$ .

(ii) Étudions les extrema locaux et globaux de même fonction  $f$  sur l'intervalle  $I = [\frac{1}{2}, 2]$ .

Les réponses suivantes s'obtiennent en traçant le graphe de la fonction  $f$  à l'aide d'un tableau de variations.

(i)  $f$  possède 3 extrema locaux en 0, 1 et  $-1$ , plus précisément 0 est un point de maximum locale (mais ce n'est pas un point de maximum globale : le maximum globale de  $f$  n'existe pas). De plus  $\pm 1$  sont deux points de minimum globaux sur  $\mathbb{R}$  et  $\min_{\mathbb{R}} f = 0$ .

(ii) Sur l'intervalle  $I$ , 1 est le seul point minimum global pour  $f$ . On a  $\min_I f = 0$ . De plus, 2 est le seul point de maximum globale pour  $f$  dans  $I$  et  $\max_I f = 9$ . D'autre part  $\frac{1}{2}$  est un point de maximum local (non global) de  $f$  dans  $I$ .

Observons que dans l'exemple précédent, dans le cas (i), tous les extrema locaux ont été trouvés parmi les points où  $f'$  s'annule : c'est dû au fait que la fonction a été étudiée sur un intervalle ouvert. Dans le cas (ii) ce n'est plus le cas.

Dans cette section on se focalise sur la recherche de point d'extrema *libres*, c'est-à-dire les points de minimum ou maximum à l'intérieur de l'ensemble de définition de la fonction. Cela revient à supposer que la fonction est définie sur un ouvert.

**Proposition 3.1** (CNPO ou CN1). Soit  $f: U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ , où  $U$  est un ouvert de  $\mathbb{R}^n$  et  $x_0 \in U$  un extremum local (c'est à dire un maximum local ou un minimum local) pour  $f$ . Si  $f$  est dérivable le long d'une direction  $v$ , alors  $\frac{\partial f}{\partial v}(x_0) = 0$ . En particulier, si  $f$  est différentiable en  $x_0$ , alors toutes les dérivées partielles de  $f$  s'annulent en  $x_0$  et  $\nabla f(x_0) = 0$ .

*Dém.* La fonction d'une seule variable  $F(t) = f(x_0 + tv)$  est dérivable en  $t = 0$ , où elle possède un extremum local. Par le critère de Fermat,  $F'(0) = 0$ . Mais alors  $0 = F'(0) = \frac{\partial f}{\partial v}(x_0)$ .  $\square$

Un point  $x_0$  où  $f$  est différentiable et  $\nabla f(x_0) = 0$  s'appelle un *point stationnaire* ou *point critique* pour  $f$ .

Les points de minimum et maximum locaux d'une fonction différentiable  $f: D \rightarrow \mathbb{R}$  sont alors à chercher parmi les points stationnaires, qui sont les solutions du système

$$\begin{cases} \frac{\partial f}{\partial x_1}(x) = 0 \\ \dots \\ \frac{\partial f}{\partial x_n}(x) = 0, \end{cases} \quad x \in \overset{\circ}{D}, \quad (\text{CN1})$$

ou éventuellement parmi les points de  $D \setminus \overset{\circ}{D}$ . Ces points sont situés sur la frontière de l'ensemble  $D$ .

Le système ci-dessus s'appelle en abrégé (CNPO) ("conditions nécessaires du premier ordre") ou (CN1), puisque ce sont les dérivées d'ordre 1 qui interviennent. Vérifier ces conditions n'est pas une condition suffisante pour garantir que les solutions trouvées soient bien des minimiseurs ou des maximiseurs : par exemple, pour la fonction  $f: \mathbb{R} \rightarrow \mathbb{R}$  telle que  $f(x) = x^3$  ce système se réduit à la seule équation  $3x^2 = 0$ . La solution  $x = 0$  ne correspond pas à un extremum pour  $f$ .

Observer que (CN1) est un système (non-linéaire!) de  $n$  équations et  $n$  inconnues. Il n'y a malheureusement pas de théorie générale pour trouver les solutions.

**Exemple 3.2.** Trouver les extrema locaux et globaux de la fonction  $f(x, y) = \sqrt{1 - x^2 - y^2}$  sur son ensemble de définition.

Réponse : L'ensemble de définition est le disque  $D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ . Ce n'est pas un ouvert. Cherchons les points stationnaires intérieurs à  $D$ , c'est à dire dans  $\overset{\circ}{D} = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$ . L'unique solution du système (CN1) est le point  $(0, 0)$ . Observons que  $f(0, 0) = 1$ . Mais on voit immédiatement que  $f(x, y) \leq 1$  pour tout  $(x, y) \in D$ , donc l'origine est un point de maximum global pour  $f$ . Il n'y a pas de point de minimum (local ou global) à l'intérieur de  $D$  (sinon on aurait trouvé d'autres points stationnaires) ; cherchons alors le minimum de  $f$  sur la frontière, qui est le cercle  $\partial D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ . Mais on voit que la fonction  $f$  s'annule sur  $\partial D$  ; d'autre part  $f \geq 0$  dans  $D$ . La conclusion est que tous les points de  $\partial D$  sont des points de minimum global pour  $f$ .

**Exemple 3.3.** On cherche à construire un caisson de  $20\text{m}^3$ . Le matériau pour le fond coûte 3 euros/ $\text{m}^2$ , pour le couvercle 2 euros/ $\text{m}^2$  et pour les côtés 1 euro/ $\text{m}^2$ . Quel est le caisson le moins cher ? Et le plus cher ?

Réponse :

*Modélisation :* Notons  $x$ (=longueur),  $y$ (=largeur),  $z$ (=hauteur) les mesures du caisson en mètres. Le coût de construction est  $C(x, y, z) = 3xy + 2xy + 2(xz + yz) = 5xy + 2xz + 2yz$ . Il s'agit de résoudre les problèmes de minimisation et maximisation pour  $C$  "avec contrainte" :  $\min\{C(x, y, z) : xyz = 20\}$  et  $\max\{C(x, y, z) : xyz = 20\}$ .

*Élimination de la contrainte et recherche des points stationnaires :* L'ensemble des points vérifiant  $xyz = 20$  n'est pas un ouvert, c'est pourquoi la proposition (CN1) ne s'applique pas directement à la fonction  $C$ . On impose  $z = 20/(xy)$  et on étudie la fonction

$$f(x, y) = C(x, y, \frac{20}{xy}) = 5xy + 40(\frac{1}{y} + \frac{1}{x}), \quad x > 0, y > 0.$$

Nous pouvons appliquer la proposition (CN1) à la fonction  $f$ , qui est bien définie dans un ouvert. Le système  $\nabla f(x, y) = 0$  possède une seule solution pour  $x > 0$  et  $y > 0$ . Elle est donnée par  $x = y = 2$  (et donc  $z = 5$ ).

*Synthèse :* On construit donc un caisson de mesures  $2 \times 2 \times 5$ . Ce choix correspond-t-il au caisson de coût minimum ou maximum ? Le problème de minimisation est-il *bien posé* ? Et celui de maximisation ? [Voir l'exemple ci-après].

Dans la pratique, les fonction dont on étudie les problèmes d'optimisation sont souvent continues. Le théorème de Weierstrass est alors un outil essentiel pour garantir que des solutions existent. Mais ce théorème ne s'applique pas toujours (comme par exemple dans l'exemple précédent). Le théorème suivant est une variante utile.

**Théorème 3.2** (Weierstrass - variante). *Soit  $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction continue. Si  $\bar{x} \in D$  et si l'ensemble de sous-niveau  $K = \{x \in D: f(x) \leq f(\bar{x})\}$  est compact, alors le problème de minimisation  $\min_{x \in D} f(x)$  possède une solution. Autrement dit, il existe  $x^* \in D$  tel que  $f(x^*) = \min_{x \in D} f(x)$ .*

*Dém.* En effet, il existe  $x^* \in K$  tel que  $f(x^*) = \min_{x \in K} f(x)$ , par le théorème de Weierstrass. Mais  $f(x^*) \leq f(\bar{x})$  puisque  $\bar{x} \in K$ . Si  $x \in K$  on a  $f(x) \geq f(x^*)$  par définition de  $x^*$ . Si  $x \in D$  et  $x \notin K$  on a  $f(x) > f(\bar{x}) \geq f(x^*)$ . En conclusion,  $f(x^*) = \min_{x \in D} f(x)$ .  $\square$

Le cas typique d'application est celui d'une fonction  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  telle que  $\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty$ . Une telle fonction a tous les ensembles de sous-niveau bornés (pourquoi?). Si de plus  $f$  est continue, ses ensembles de sous-niveau sont fermés (pourquoi?) et donc compacts. La fonction possède alors un minimum absolu.

Bien entendu on peut établir un théorème analogue pour l'existence d'un maximum absolu : il s'agit cette fois-ci de supposer que l'ensemble de "sur-niveau"  $\{x: f(x) \geq f(\bar{x})\}$  est borné. Le cas typique d'application est celui d'une fonction continue telle que  $\lim_{\|x\| \rightarrow +\infty} f(x) = -\infty$ .

**Exemple 3.4** (Retour à l'exemple 7.7). Appliquons cette variante du théorème de Weierstrass avec  $(\bar{x}, \bar{y}) = (1, 1)$  (ce choix est arbitraire). Observons que l'ensemble de sous-niveau  $K = \{(x, y): f(x, y) \leq f(1, 1)\} = \{5xy + 40(\frac{1}{y} + \frac{1}{x}) \leq 85\}$  est compact (en effet, il est manifestement fermé et il est borné, puisque  $x \geq 40/85$ ,  $y \geq 40/85$  et  $5xy \leq 85 \Rightarrow x \leq 17 \cdot 85/40$  et  $y \leq 17 \cdot 85/40$ ). Mais alors le problème de minimisation de l'exemple 7.7 est bien posé, c'est-à-dire que le caisson de coût minimum existe : c'est bien le caisson de mesures  $2 \times 2 \times 5$  trouvé avant. Le problème de maximisation est mal posé : le caisson de coût maximum n'existe pas.

### 3.2 Fonctions quadratiques et matrice hessienne

**Un peu d'algèbre linéaire.** Commençons par des rappels d'algèbre linéaire. Un système linéaire homogène de  $n$  équations et  $n$  inconnues se représente par l'équation matricielle

$$Ax = 0, \quad x = (x_1, \dots, x_n)^T \in \mathbb{R}^n.$$

où  $A$  est une matrice carrée  $n \times n$ . Ce système possède toujours comme solutions au moins le vecteur nul  $x = 0$ . Il est bien connu que la condition nécessaire et suffisante pour qu'il y ait d'autres solutions  $x \neq 0$  est que  $\det A = 0$ .

**Définition 3.2.** *Un nombre réel (ou complexe)  $\lambda$  est dit valeur propre d'une matrice carrée  $A$  s'il existe  $w \in \mathbb{R}^n \setminus \{0\}$  tel que  $Aw = \lambda w$ . Le vecteur  $w$  s'appelle alors vecteur propre pour  $A$ .*

Observons que si  $\lambda$  est une valeur propre alors  $(A - \lambda I)w = 0$ , avec  $w \in \mathbb{R}^n \setminus \{0\}$ . Mais cela est possible si et seulement si

$$\det(A - \lambda I) = 0. \tag{3.1}$$

Chercher les valeurs propres revient à résoudre l'équation d'inconnue  $\lambda$  (3.1). En général, (3.1) est une équation polynomiale en  $\lambda$ , de degré  $n$ . Le terme à gauche dans (3.1) s'appelle le *polynôme caractéristique* de  $A$ . Cette équation possède alors  $n$  solutions, comptées avec leur multiplicité. Ces solutions sont éventuellement complexes.

**Exemple 3.5.** La matrice  $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$  possède 2 valeurs propres  $\lambda_1$  et  $\lambda_2$ . Celles-ci sont les deux solutions d'une équation polynomiale de degré 2 :

$$0 = \det \left[ \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] = \det \begin{pmatrix} 1 - \lambda & 2 \\ 3 & 4 - \lambda \end{pmatrix} = \lambda^2 - 5\lambda - 1.$$

Donc  $\lambda_{1,2} = \frac{1}{2}(5 \pm \sqrt{29})$ .

**Théorème 3.3** (démonstration hors programme. Voir le cours d'algèbre.). *Si  $A$  est une matrice carrée symétrique de taille  $n \times n$ , c'est-à-dire  $a_{ij} = a_{ji}$  pour tout  $i, j = 1, \dots, n$ , alors les valeurs propres de  $A$  sont toutes réelles.*

Soit  $A = (a_{ij})$  une matrice *symétrique*  $n \times n$ . Le produit entre la matrice  $A$  et le vecteur  $h = (h_1, \dots, h_n)^T$  est donné par le vecteur de composantes  $(Ah)_i = \sum_{j=1}^n a_{ij}h_j$ . Une *forme quadratique* est une application :  $\mathbb{R}^n \rightarrow \mathbb{R}$  de la forme

$$h \mapsto \langle Ah, h \rangle = \sum_{i,j=1}^n a_{ij}h_ih_j.$$

**Exemple 3.6.** Les formes quadratiques de  $\mathbb{R}^2$  sont toutes et seules les fonctions de la forme

$$(x, y) \mapsto \left\langle \begin{pmatrix} r & s \\ s & t \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} \right\rangle = rx^2 + 2sxy + ty^2.$$

Les formes quadratiques de  $\mathbb{R}^3$  sont toutes et seules les fonctions de la forme

$$(x, y, z) \mapsto \left\langle \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right\rangle = ax^2 + 2bxy + 2cxz + dy^2 + 2ezy + fz^2.$$

**Le cas des fonctions de plusieurs variables.** Les formes quadratiques apparaissent naturellement dans la formule de Taylor d'ordre 2. En effet, soit  $f$  une fonction de classe  $C^2$  au voisinage d'un point  $x_0$ . Considérons la *matrice hessienne de  $f$  en  $x_0$* , c'est à dire la matrice symétrique de taille  $n \times n$  des dérivées partielles secondes

$$H = H_f(x_0) = \left( \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) \right)_{1 \leq i, j \leq n}.$$

Pour les fonctions  $f: \mathbb{R} \rightarrow \mathbb{R}$ , la matrice hessienne est réduite au seul nombre  $f''(x_0)$ .

La formule de Taylor pour  $f$  d'ordre 2 s'écrit alors

$$f(x_0 + h) = f(x_0) + \langle \nabla f(x_0), h \rangle + \frac{1}{2} \langle H_f(x_0)h, h \rangle + o(\|h\|^2).$$

Si de plus  $x_0$  est un point stationnaire, on a  $\nabla f(x_0) = 0$  et donc la formule précédente devient

$$f(x_0 + h) = f(x_0) + \frac{1}{2} \langle H_f(x_0)h, h \rangle + o(\|h\|^2), \quad (3.2)$$

qui est la généralisation naturelle de (3.4).

Pour déterminer la nature du point stationnaire  $x_0$  on est amené à diviser terme-à-terme par  $\|h\|^2$  et prendre  $\|h\| \rightarrow 0$ . Cela conduit à étudier la fonction  $h \mapsto \frac{\langle H_f(x_0)h, h \rangle}{\|h\|^2}$ .

Plus en général, étudions alors les fonctions de type

$$F(h) = \frac{\langle Ah, h \rangle}{\|h\|^2}, \quad h \neq 0. \quad (3.3)$$

avec  $A = (a_{ij})$  matrice symétrique de taille  $n \times n$ .

**Lemme 3.4.** Soit  $F(h) = \frac{\langle Ah, h \rangle}{\|h\|^2}$  avec  $A = (a_{ij})$  matrice symétrique de taille  $n \times n$ . Les points stationnaires de  $F$  sont des vecteurs propres de la matrice  $A$ . De plus, si  $v$  est un vecteur propre de  $A$ , alors

$$Av = F(v)v,$$

c'est à dire que la valeur propre associée à  $v$  est  $F(v)$ .

*Dém.* Observons que  $\frac{\partial}{\partial h_i}(\langle Ah, h \rangle) = 2 \sum_{j=1}^n a_{ij}h_j$  et que  $\frac{\partial}{\partial h_i}\|h\|^2 = 2h_i$ . Il en découle que

$$\frac{\partial F}{\partial h_i}(h) = 2 \frac{\sum_{j=1}^n a_{ij}h_j - F(h)h_i}{\|h\|^2} = 2 \frac{[Ah - F(h)h]_i}{\|h\|^2}.$$

Soit  $v$  un point stationnaire de  $F$ . Alors, pour tout  $i = 1, \dots, n$ , on a  $\frac{\partial F}{\partial h_i}(v) = 0$ . Mais alors les numérateurs de l'expression précédente s'annulent et on trouve

$$Av = F(v)v.$$

□

**Lemme 3.5.** Soit  $F(h) = \frac{\langle Ah, h \rangle}{\|h\|^2}$  avec  $A = (a_{ij})$  matrice symétrique de taille  $n \times n$ . Soit  $\lambda_1$  et  $\lambda_n$  la plus petite et la plus grande valeurs propre de la matrice  $A$ . Alors

$$\min_{h \neq 0} F(h) = \lambda_1, \quad \text{et} \quad \max_{h \neq 0} F(h) = \lambda_n.$$

*Dém.* Observons que  $F$  est une fonction telle que, pour tout  $\lambda > 0$ ,  $F(\lambda h) = F(h)$ . en particulier, en prenant  $\lambda = \frac{1}{\|h\|}$  on voit que  $F(h) = F(\frac{h}{\|h\|})$ . Mais alors  $\max_{h \neq 0} F(h) = \max\{F(\frac{h}{\|h\|}) : h \neq 0\} = \max\{F(h) : \|h\| = 1\}$ . De même,  $\min_{h \neq 0} F(h) = \min\{F(h) : \|h\| = 1\}$ . L'ensemble  $\{h : \|h\| = 1\}$  est la sphère unité qui est compact dans  $\mathbb{R}^n$ . La fonction  $F$  étant continue sur  $\{h : \|h\| = 1\}$ , le théorème de Weierstrass implique que  $F$  possède un maximum et un minimum absolu sur cet ensemble. Par la discussion précédente,  $F$  possède un minimum et un maximum absolu sur  $\mathbb{R}^n \setminus \{0\}$ .

Soit donc  $h^*$  un point de minimum de  $F$  dans  $\mathbb{R}^n \setminus \{0\}$ . Alors  $h^*$  est un point stationnaire et par le lemme précédent  $F(h^*)$  est une valeur propre de la matrice  $A$ . De plus,  $Ah^* = F(h^*)h^*$ .

Mais alors  $\lambda_1 \leq F(h^*) = \min_{h \neq 0} F(h)$ . Mais au valeur propre  $\lambda_1$  correspondent un vecteur propre  $v_1 \in \mathbb{R}^n \setminus \{0\}$ . et l'on a  $Av_1 = \lambda_1 v_1$ , donc  $F(v_1) = \lambda_1$ . Donc  $\min_{h \neq 0} F(h) \leq \lambda_1$ .

Si  $\hat{h}$  est un point de maximum pour  $F$ , on procède de la même manière et on trouve  $F(\hat{h}) = \lambda_n$ . □

### 3.3 Problèmes d'optimisations : conditions d'ordre 2

**Le cas de fonctions d'une seule variable. Rappels.** Soit  $f : I \rightarrow \mathbb{R}$  une fonction de classe  $C^2$  sur l'intervalle ouvert  $I = ]a, b[$  et  $x_0$  un point stationnaire pour  $f$  dans  $]a, b[$ , c'est-à-dire  $f'(x_0) = 0$ . Alors, par la formule de Taylor d'ordre 2,

$$f(x_0 + h) = f(x_0) + \frac{1}{2}f''(x_0)h^2 + o(h^2), \quad \text{pour } h \rightarrow 0. \quad (3.4)$$

Si on divise terme-à-terme par  $h^2$  et on prend  $h \rightarrow 0$  on en déduit :

$$\frac{1}{h^2}(f(x_0 + h) - f(x_0)) = \frac{1}{2}f''(x_0) + o(1), \quad \text{pour } h \rightarrow 0.$$

- Pour que  $x_0$  soit un point de minimum local il est nécessaire que  $f''(x_0) \geq 0$ . (En effet, si  $x_0$  est un point de minimum local, alors pour  $h$  assez petit le terme à gauche est  $\geq 0$ ).

- Pour que  $x_0$  soit un point de minimum local il est suffisant que  $f''(x_0) > 0$ . (En effet, si  $f''(x_0) > 0$  alors pour  $h$  assez petit  $f''(x_0) + o(1) > 0$  et donc le terme à gauche est positif).
- Pour que  $x_0$  soit un point de maximum local il est nécessaire que  $f''(x_0) \leq 0$ .
- Pour que  $x_0$  soit un point de maximum local il est suffisant que  $f''(x_0) < 0$ .

Observer que lorsque  $f''(x_0) = 0$  on ne peut rien conclure en général : on a parfois un minimum, ou un maximum, ou un point d'inflexion (considérer par exemple les cas  $f(x) = x^3$ , avec  $x_0 = 0$ ). Généralisons ces considérations aux fonctions de plusieurs variables.

Le théorème suivant établit des conditions nécessaires (CN2) et des conditions suffisantes (CS2), faisant intervenir les dérivées d'ordre 2, pour qu'un point soit un minimiseur ou un maximiseur local.

**Théorème 3.6** (CN2-CS2). *Si  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  est une fonction de classe  $C^2$  au voisinage de  $x_0$  et si  $x_0$  est un point stationnaire, alors :*

- Pour que  $x_0$  soit un point de minimum local il est nécessaire que tous les valeurs propres de la matrices hessiennes de  $f$  soient  $\geq 0$ . (On dit que  $H_f(x_0)$  est semi-définie positive).
- Pour que  $x_0$  soit un point de minimum local il est suffisant que tous les valeurs propres de la matrices Hessiennes de  $f$  soient  $> 0$ . (On dit que  $H_f(x_0)$  est définie positive).
- Pour que  $x_0$  soit un point de maximum local il est nécessaire que tous les valeurs propres de la matrices Hessiennes de  $f$  soient  $\leq 0$ . (On dit que  $H_f(x_0)$  est semi-définie négative).
- Pour que  $x_0$  soit un point de maximum local il est suffisant que tous les valeurs propres de la matrices Hessiennes de  $f$  soient  $< 0$ . (On dit que  $H_f(x_0)$  est définie négative).

*Dém.* Si  $x_0$  est un point de minimum local, alors  $\exists \delta > 0$  tel que pour tout  $\|h\| < \delta$  on a (voir (3.2))

$$\frac{1}{2} \langle H_f(x_0)h, h \rangle + o(\|h\|^2) = f(x_0 + h) - f(x_0) \geq 0.$$

Mais alors, avec les notations du lemme précédent, appliqué à  $A = H_f(x_0)$ , on trouve, pour  $\|h\| < \delta$ ,

$$F(h) + o(1) \geq 0.$$

Soit  $\lambda_1 = \min_{h \neq 0} F(h) = F(v_1)$ . Comme  $F(v_1) = F(\alpha v_1)$  pour tout  $\alpha > 0$ , en prenant  $\alpha \rightarrow 0$  on déduit de l'inégalité précédente

$$\lambda_1 = \lim_{\alpha \rightarrow 0^+} F(\alpha v_1) \geq 0.$$

Mais alors toutes les valeurs propres de  $H_f(x_0)$  sont  $\geq 0$ . Cela prouve la première conclusion.

Supposons maintenant que toutes les valeurs propres soient strictement positives, et donc  $\lambda_1 = \min_{h \neq 0} F(h) > 0$ . En utilisant que

$$\frac{1}{\|h\|^2} [f(x_0 + h) - f(x_0)] = F(h) + o(1) \geq \lambda_1 + o(1), \quad \text{pour } h \rightarrow 0.$$

on voit que pour  $\|h\|$  assez petit cette expression est (strictement) positive. Donc  $x_0$  est bien un point de minimum local (strict) pour  $f$ .

Les deux conclusions restantes se démontrent de la même manière. □

**Remarque 3.7** (Points de selle). Il arrive parfois qu'un point critique ne soit ni de minimum, ni de maximum local. C'est le cas, par exemple quand la plus petite et la plus grande valeur propre de la matrice Hessienne vérifient  $\lambda_1 < 0 < \lambda_n$ . On dit alors que  $x_0$  est un *point de selle* (ou de *col*, ou de *min-max*). Par exemple, pour les fonctions de 2 variables, si  $(x_0, y_0)$  est un point de selle, alors la fonction  $x \mapsto f(x, y_0)$  aura un minimum local en  $x_0$  et la fonction  $y \mapsto f(x_0, y)$  un maximum local en  $y_0$  (ou l'inverse).

**Règles pratiques pour appliquer le théorème 3.6** Pour appliquer le théorème 3.6 il n'est pas indispensable de calculer les valeurs propres de  $H_f(x_0)$ , puisque seul le signe des valeurs propres joue un rôle. Pour une fonctions de  $n$  variables, ces valeurs propres sont les solutions de l'équation (3.1), qui est une équation de la forme

$$\lambda^n + a_1\lambda^{n-1} + \dots + a_{n-1}\lambda + a_n = 0, \quad (3.5)$$

où le polynôme à gauche est le polynôme caractéristique de la matrice  $H_f(x_0)$ .

**Exercice 3.8.** Démontrer que les racines de (3.5) (dont on sait qu'elles sont toutes réelles) :

- (i) sont toutes strictement négatives si et seulement si tous les coefficients  $a_1 > 0, \dots, a_n > 0$ .
- (ii) sont toutes strictement positives si et seulement si les coefficients vérifient  $a_1 < 0, a_2 > 0, a_3 < 0$ , etc.

**Exemple 3.9.** La fonction  $f(x, y, z) = x^2 + 2y^2 + z^2 + xy - xz$  a comme seul point stationnaire l'origine. On calcule aisément les dérivées partielles secondes en 0. On trouve alors.

$$\det(H_f(0) - \lambda I) = \det \begin{pmatrix} 2 - \lambda & 1 & -1 \\ 1 & 4 - \lambda & 0 \\ -1 & 0 & 2 - \lambda \end{pmatrix} = -(\lambda^3 - 8\lambda^2 + 18\lambda - 10) = 0.$$

Compte tenu du résultat de l'exercice précédent, les valeurs propres de la matrice Hessienne sont toutes strictement positives. L'origine est alors un point de minimum local. Cet analyse ne donne aucun renseignement sur la nature globale de ce point de minimum.

**Exercice 3.10** (Une méthode pratique pour les fonction de 2 variables). Soit  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$  de classe  $C^2$ . Il est standard de noter  $r = \frac{\partial^2 f}{\partial x^2}$ ,  $s = \frac{\partial^2 f}{\partial x \partial y}$  et  $t = \frac{\partial^2 f}{\partial y^2}$ . Calculer  $\det \begin{pmatrix} r - \lambda & s \\ s & t - \lambda \end{pmatrix}$  et en déduire que, en un point  $(x_0, y_0)$  :

- Si  $rt - s^2 > 0$  alors  $(x_0, y_0)$  est un extremum local pour  $f(x, y)$  : un minimiseur si  $r > 0$  et un maximiseur si  $r < 0$ .
- Si  $rt - s^2 < 0$  alors  $(x_0, y_0)$  est un point de selle.

## 4 Intégrale de Riemann et intégrales multiples

### 4.1 Sommes de Darboux

**Définition 4.1.** Une subdivision  $\Delta$  d'un intervalle  $[a, b]$  est une suite finie de réels  $x_0, x_1, \dots, x_n$  tels que

$$a = x_0 < x_1 < \dots < x_{n-1} < x_n = b.$$

Si  $\Delta$  est une telle subdivision, on notera, pour  $k = 1, \dots, n$  :

$$I_k = [x_{k-1}, x_k], \quad \text{et} \quad |I_k| = x_k - x_{k-1}.$$

**Définition 4.2** (Sommes de Darboux inférieure et supérieure). Si  $f: [a, b] \rightarrow \mathbb{R}$  est une fonction bornée et  $\Delta$  une subdivision de  $[a, b]$ , on pose

$$\Sigma_{\Delta}(f) = \sum_{k=1}^n |I_k| \inf_{I_k} f, \quad \text{et} \quad \Sigma^{\Delta}(f) = \sum_{k=1}^n |I_k| \sup_{I_k} f$$



Dans la définition précédente  $f$  peut être continue ou discontinue sur  $[a, b]$ . Pour les fonctions positives, géométriquement chacune des sommes de Darboux s'interprètent comme l'aire d'un « pluri-rectangle ». Lorsqu'il y a de nombreux points de subdivisions,  $\Sigma_{\Delta}(f)$  et  $\Sigma^{\Delta}(f)$  approchent, respectivement par dessous et par dessus, l'aire du sous-graphe de la fonction, c'est-à-dire de l'ensemble

$$\{(x, y) \in \mathbb{R}^2 : a \leq x \leq b, 0 \leq y \leq f(x)\}.$$

Par exemple, si  $f: [0, 1] \rightarrow \mathbb{R}$  est la fonction  $f(x) = x^2$ , pour la subdivision  $\Delta = (0, \frac{1}{2}, \frac{2}{3}, 1)$ , on a  $|I_1| = \frac{1}{2}$ ,  $|I_2| = \frac{1}{6}$  et  $|I_3| = \frac{1}{3}$ . Alors

$$\Sigma_{\Delta}(f) = \frac{1}{2} \cdot 0 + \frac{1}{6} \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{4}{9} = \frac{31}{216} \quad \text{et} \quad \Sigma^{\Delta}(f) = \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{6} \cdot \frac{4}{9} + \frac{1}{3} \cdot 1 = \frac{115}{216}.$$

**Définition 4.3.** Si  $\Delta$  et  $\Delta'$  sont deux subdivisions d'un intervalle  $[a, b]$ , on dit que  $\Delta'$  est plus fine que  $\Delta$  si tous les points de la subdivision  $\Delta$  sont aussi des points de la subdivision  $\Delta'$ .

**Proposition 4.1.** (monotonie des subdivisions)

i) Si  $\Delta$  et  $\Delta'$  sont deux subdivisions de  $[a, b]$ , avec  $\Delta'$  plus fine que  $\Delta$ , et  $f: [a, b] \rightarrow \mathbb{R}$  est bornée, on a

$$\Sigma_{\Delta}(f) \leq \Sigma_{\Delta'}(f) \leq \Sigma^{\Delta'}(f) \leq \Sigma^{\Delta}(f).$$

ii) Si  $\Delta_1$  et  $\Delta_2$  sont deux subdivisions de  $[a, b]$ , alors

$$\Sigma_{\Delta_1}(f) \leq \Sigma^{\Delta_2}(f).$$

*Dém.* i) Il suffit de considérer le cas où  $\Delta'$  possède un seul point supplémentaire en plus que  $\Delta$ . (En effet, si  $\Delta'$  contient  $k \geq 2$  points supplémentaires, il suffit d'itérer  $k$  fois l'argument). Or, soit  $y$  ce point supplémentaire et soit  $k$  tel que

$$x_{k-1} < y < x_k.$$

La seule différence entre les sommes de Darboux inférieures  $\Sigma_{\Delta}(f)$  et  $\Sigma_{\Delta'}(f)$  est que le terme  $|I_k| \inf_{I_k} f$  de la première somme doit être remplacé par les deux termes

$$|y - x_{k-1}| \inf_{[x_{k-1}, y]} f + |y - x_{k-1}| \inf_{[y, x_k]} f$$

de la seconde. Les autres termes de  $\Sigma_{\Delta}(f)$  et  $\Sigma_{\Delta'}(f)$  sont les mêmes. Mais,

$$\begin{aligned} |I_k| \inf_{I_k} f &= (x_k - x_{k-1}) \inf_{I_k} f \\ &\leq (y - x_{k-1}) \inf_{I_k} f + (x_k - y) \inf_{I_k} f \\ &\leq (y - x_{k-1}) \inf_{[x_{k-1}, y]} f + (y - x_{k-1}) \inf_{[y, x_k]} f. \end{aligned}$$

Donc  $\Sigma_{\Delta}(f) \leq \Sigma_{\Delta'}(f)$ . L'inégalité  $\Sigma_{\Delta'}(f) \leq \Sigma^{\Delta'}(f)$  est évidente.

L'inégalité  $\Sigma^{\Delta'}(f) \leq \Sigma^{\Delta}(f)$  se prouve de manière semblable.

ii) Posons  $\Delta' = \Delta_1 \cup \Delta_2$ . Cette subdivision est plus fine que  $\Delta_1$  et  $\Delta_2$ . D'après le point précédent,

$$\Sigma_{\Delta_1}(f) \leq \Sigma_{\Delta'}(f) \leq \Sigma^{\Delta'}(f) \leq \Sigma^{\Delta_2}(f).$$

□

Pour une fonction bornée  $f: [a, b] \rightarrow \mathbb{R}$ , on introduit les nombres réels

$$I_*(f) \stackrel{\text{d\u00e9f}}{=} \sup_{\Delta} \Sigma_{\Delta}(f), \quad \text{et} \quad I^*(f) \stackrel{\text{d\u00e9f}}{=} \inf_{\Delta} \Sigma^{\Delta}(f), \quad (4.1)$$

où le supremum et l'infimum sont pris sur toutes les subdivisions possibles de  $[a, b]$ . Ces expressions sont bien définies (pourquoi?) et on a toujours

$$-\infty < I_*(f) \leq I^*(f) < +\infty.$$

**Exemple 4.1.** Considérons la fonction de Dirichlet  $\Phi: [a, b] \rightarrow \mathbb{R}$ , égale à 1 sur les rationnels et 0 sur les irrationnels de l'intervalle  $[a, b]$ . Si  $\Delta$  est une subdivision arbitraire, et  $I_k$  est un intervalle de cette subdivision, alors  $\inf_{I_k} \Phi = 0$  et  $\sup_{I_k} \Phi = 1$ . On voit alors que  $\Sigma_{\Delta}(\Phi) = 0$  et  $\Sigma^{\Delta}(\Phi) = b - a$ . Pour la fonction de Dirichlet sur l'intervalle  $[a, b]$ , on a alors  $I_*(\Phi) = 0$  et  $I^*(\Phi) = b - a$ .

## 4.2 Fonctions Riemann intégrables

**Définition 4.4.** Une fonction bornée  $f: [a, b] \rightarrow \mathbb{R}$  est Riemann-intégrable si  $I_*(f) = I^*(f)$ . Dans ce cas, on écrit

$$\int_a^b f = I_*(f) = I^*(f),$$

où  $\int_a^b f$  est « l'intégrale de Riemann de  $f$  sur  $[a, b]$  ».

Si  $f$  est à valeurs complexes, on dira qu'elle est Riemann-intégrable si ses parties réelles et imaginaires sont Riemann-intégrables. Dans ce cas, on pose  $\int_a^b f = \int_a^b (\operatorname{Re} f) + i \int_a^b (\operatorname{Im} f)$ .

Lorsqu'il s'agit de calculer explicitement l'intégrale d'une fonction  $x \mapsto f(x)$ , il est souvent utile d'utiliser la notation alternative  $\int_a^b f = \int_a^b f(x) dx$ .

Pour les fonctions Riemann-intégrables positives  $\int_a^b f$  exprime géométriquement l'aire du sous-graphe de  $f$  sur  $[a, b]$ , c'est-à-dire de l'ensemble  $\{(x, y) \in \mathbb{R}^2 \mid a \leq x \leq b, 0 \leq y \leq f(x)\}$ . Pour les fonctions réelles qui changent de signe,  $\int_a^b f$  exprime une « aire algébrique » (l'aire des portions négatives de la fonction est compté négativement).

Compatiblement avec cette interprétation, on pose, par définition,

$$\int_a^a f = 0 \quad \text{et} \quad \int_b^a f = - \int_a^b f.$$

La fonction de Dirichlet est un exemple de fonction non Riemann-intégrable. Il n'est pas possible de calculer l'aire du sous-graphe de cette fonction avec la théorie de Riemann, mais il serait possible de le faire avec d'autres théories d'intégration.

Commençons par rappeler une caractérisation des bornes sup et inf qu'on utilisera dans la proposition suivante :

- Si  $A \subset \mathbb{R}$  est majoré et  $S \in \mathbb{R}$ , on a

$$S = \sup A \iff \begin{cases} \forall a \in A, & a \leq S \\ \forall \epsilon > 0, \exists a \in A: & S - \epsilon < a \end{cases}$$

- Si  $A \subset \mathbb{R}$  est minoré  $I \in \mathbb{R}$ , on a

$$I = \inf A \iff \begin{cases} \forall a \in A, & I \leq a \\ \forall \epsilon > 0, \exists a \in A: & a < I + \epsilon. \end{cases}$$

**Proposition 4.2** (Une c.n.s. d'intégrabilité). *Une fonction bornée  $f: [a, b] \rightarrow \mathbb{R}$  est Riemann intégrable si et seulement si, pour tout  $\epsilon > 0$ , il existe une subdivision  $\Delta$  de l'intervalle  $[a, b]$  telle que*

$$0 \leq \Sigma^\Delta(f) - \Sigma_\Delta(f) \leq \epsilon.$$

Dans ce cas, on a

$$\Sigma^\Delta(f) - \epsilon \leq \int_a^b f \leq \Sigma_\Delta(f) + \epsilon$$

*Dém.* \* Il s'agit de démontrer l'équivalence

$$\forall \epsilon > 0, \exists \Delta \text{ tel que } 0 \leq \Sigma^\Delta(f) - \Sigma_\Delta(f) \leq \epsilon \iff I_*(f) = I^*(f)$$

Pour démontrer l'implication  $\Rightarrow$ , on procède ainsi : pour  $\epsilon > 0$ , on considère une subdivision  $\Delta$  telle que  $0 \leq \Sigma^\Delta(f) - \Sigma_\Delta(f) \leq \epsilon$ . Ensuite, par l'hypothèse,

$$I^*(f) \leq \Sigma^\Delta(f) \leq \Sigma_\Delta(f) + \epsilon \leq I_*(f) + \epsilon.$$

Mais  $\epsilon > 0$  étant arbitraire, on a  $I^*(f) \leq I_*(f)$ , et donc  $I^*(f) = I_*(f)$ .

Pour démontrer l'implication  $\Leftarrow$ , on pose  $I(f) = I_*(f) = I^*(f)$  et on procède ainsi : pour  $\epsilon > 0$ , on choisit deux subdivisions  $\Delta_1$  et  $\Delta_2$  telles que

$$\begin{aligned} I(f) - \epsilon/2 < \Sigma_{\Delta_1}(f) \leq I(f) & \quad (\text{propriété du sup}), \\ I(f) \leq \Sigma^{\Delta_2}(f) < I(f) + \epsilon/2 & \quad (\text{propriété de l'inf}). \end{aligned}$$

Considérons maintenant la subdivision plus fine  $\Delta = \Delta_1 \cup \Delta_2$ . Grâce à la propriété de monotonie des subdivisions :

$$\begin{cases} I(f) - \epsilon/2 < \Sigma_{\Delta_1}(f) \leq \Sigma_\Delta(f) \\ \Sigma^\Delta(f) \leq \Sigma^{\Delta_2}(f) < I(f) + \epsilon/2. \end{cases}$$

En prenant la différence on trouve  $\Sigma^\Delta(f) - \Sigma_\Delta(f) < \epsilon$ . L'encadrement de  $\int_a^b f = I(f)$  en découle.  $\square$

### 4.3 Classes de fonctions Riemann-intégrables

**Théorème 4.3.** *Si  $f \in C([a, b], \mathbb{R})$ , alors  $f$  est Riemann-intégrable.*

*Dém.* Par le théorème de Weierstrass,  $f$  est bornée sur  $[a, b]$ . Par le théorème de Heine,  $f$  est uniformément continue sur  $[a, b]$ .

Soit  $\epsilon > 0$ . La continuité uniforme de  $f$  garantit l'existence de  $\delta > 0$  tel que

$$|x - y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon/(b - a).$$

Choisissons une subdivision  $\Delta = x_0, x_1, \dots, x_n$  uniforme de  $[a, b]$ . Cela signifie que les points de la subdivision sont équidistants, et deux points consécutifs sont à une distance égale à  $\delta' = (b - a)/n$ . Nous choisissons  $n$  de manière que  $\delta' < \delta$ . Si  $I_k$  est un intervalle de cette subdivision, avec  $k = 1, \dots, (b - a)/\delta'$ , on a

$$|\sup_{I_k} f - \inf_{I_k} f| < \epsilon/(b - a).$$

Mais alors

$$\Sigma^\Delta(f) - \Sigma_\Delta(f) \leq \sum_{k=1}^n \delta' |\sup_{I_k} f - \inf_{I_k} f| \leq \sum_{k=1}^n \delta \epsilon/(b - a) = \epsilon.$$

La proposition précédente implique que  $f$  est Riemann-intégrable.  $\square$

---

\*. Non exigible à l'examen

Le résultat précédent se généralise aux fonctions bornées et continues par morceaux sur l'intervalle  $[a, b]$ . Par définition une fonction continue par morceaux est une fonction ayant un nombre fini de points de discontinuités. L'intégrale d'une telle fonction se calcule comme la somme des intégrales sur les sous-intervalles où la fonction est continue au moins à l'intérieur de ces intervalles.

**Théorème 4.4.** *Si  $f: [a, b] \rightarrow \mathbb{R}$  est bornée et continue par morceaux sur  $[a, b]$ , alors  $f$  est intégrable.*

*Dém.* Traitons d'abord le cas d'une fonction ayant un seul point de discontinuité  $c \in ]a, b[$ . Soit  $\epsilon > 0$  et  $M = \sup |f|$ . Soit  $\delta > 0$  tel que  $[c - \delta, c + \delta] \subset [a, b]$  et  $M\delta < \epsilon/3$ . La fonction  $f$  est continue, et donc intégrable, sur les intervalles  $[a, c - \delta]$  et  $[c + \delta, b]$ . On considère une subdivision  $\Delta$  de  $[a, c - \delta]$  et une subdivision  $\Delta'$  de  $[c + \delta, b]$  telles que  $\Sigma^\Delta(f) - \Sigma_\Delta(f) < \epsilon/3$  et  $\Sigma^{\Delta'}(f) - \Sigma_{\Delta'}(f) < \epsilon/3$ . On considère la subdivision  $\Delta''$  de  $[a, b]$  obtenue en prenant tous les points de  $\Delta$  et  $\Delta'$  ainsi que le point  $c$ . On a  $\Sigma^{\Delta''}(f) - \Sigma_{\Delta''}(f) < \epsilon/3 + M\delta + \epsilon/3 < \epsilon$ . Donc  $f$  est Riemann intégrable sur  $[a, b]$ .

Le cas où il y a plusieurs points de discontinuité, ou bien le cas où  $f$  est discontinue au point  $a$  ou au point  $b$  se traitent de manière semblable.  $\square$

#### 4.4 Propriétés de l'intégrales

Les propriétés basiques de l'intégrales de Riemann sont les suivantes. Les deux premières propriétés traduisent le fait que l'ensemble des fonctions Riemann-intégrables sur un intervalle  $[a, b]$  est un espace vectoriel, et la linéarité de l'intégrale.

**Théorème 4.5.** *Soient  $f: [a, b] \rightarrow \mathbb{R}$  et  $g: [a, b] \rightarrow \mathbb{R}$  deux fonctions bornées et continues par morceaux. Alors  $f + g$  est bornée et continue par morceaux. De plus,*

i)  $\int_a^b (f + g) = \int_a^b f + \int_a^b g.$

ii) Si  $\lambda \in \mathbb{R}$ ,  $\int_a^b \lambda f = \lambda \int_a^b f.$

iii) Si  $f \leq g$ , alors  $\int_a^b f \leq \int_a^b g.$

iv) (Règle de Chasles) Pour tout  $c \in ]a, b[$ ,

$$\int_a^b f = \int_a^c f + \int_c^b f.$$

*Dém.* \* Soit  $\epsilon > 0$ . Il existe une subdivision  $\Delta_1$  telle que

$$\Sigma^{\Delta_1}(f) - \epsilon \leq \int_a^b f \leq \Sigma_{\Delta_1}(f) + \epsilon$$

et une subdivision  $\Delta_2$  telle que

$$\Sigma^{\Delta_2}(g) - \epsilon \leq \int_a^b g \leq \Sigma_{\Delta_2}(g) + \epsilon$$

Considérons la subdivision  $\Delta = \Delta_1 \cup \Delta_2$ . Grâce à la propriété de monotonie des subdivisions, on a

$$\Sigma^\Delta(f) - \epsilon \leq \int_a^b f \leq \Sigma_\Delta(f) + \epsilon$$

et

$$\Sigma^\Delta(g) - \epsilon \leq \int_a^b g \leq \Sigma_\Delta(g) + \epsilon.$$

---

\*. Non exigible à l'examen

En sommant terme à terme ces inégalités

$$\Sigma^\Delta(f) + \Sigma^\Delta(g) - 2\epsilon \leq \int_a^b f + \int_a^b g \leq \Sigma_\Delta(f) + \Sigma_\Delta(g) + 2\epsilon$$

Mais,

$$\begin{aligned} \Sigma_\Delta(f) + \Sigma_\Delta(g) &= \sum_{k=1}^n |I_k| \left( \inf_{I_k}(f) + \inf_{I_k}(g) \right) \leq \Sigma_\Delta(f + g) \\ &\leq \sum_{k=1}^n |I_k| \inf_{I_k}(f + g) \\ &= \Sigma_\Delta(f + g) \end{aligned}$$

De même, par les propriétés du sup,

$$\Sigma^\Delta(f + g) \leq \Sigma^\Delta(f) + \Sigma^\Delta(g).$$

Ceci donne finalement,

$$\Sigma^\Delta(f + g) - 2\epsilon \leq \int_a^b f + \int_a^b g \leq \Sigma_\Delta(f + g) + 2\epsilon$$

et donc l'intégrabilité de  $f + g$ , ainsi que la formule cherchée.

Exercice : démontrer les autres assertions. □

**Corollaire 4.6.** *Si  $f$  est bornée continue par morceaux, alors  $|f|$  l'est aussi et*

$$\left| \int_a^b f \right| \leq \int_a^b |f|.$$

*Dém.* En effet, lorsque  $f$  est continue en un point  $x_0$ , alors  $|f|$  l'est aussi, donc  $|f|$  possède un nombre fini de points de discontinuité. Donc la proposition précédente s'applique avec  $\phi(x) = |x|$ .

De plus

$$-|f| \leq f \leq |f| \Rightarrow -\int_a^b |f| \leq \int_a^b f \leq \int_a^b |f|,$$

et ceci donne l'inégalité cherchée. □

## 4.5 Intégrales de Riemann et primitives.

**Définition 4.5** (primitive). *Une primitive d'une fonction  $f : [a, b] \rightarrow \mathbb{R}$  est une fonction  $F : [a, b] \rightarrow \mathbb{R}$  dérivable sur  $[a, b]$ , telle que  $F' = f$ .*

**Théorème 4.7.** *Soit  $f : [a, b] \rightarrow \mathbb{R}$  une fonction bornée et continue par morceaux. Posons, pour  $x \in [a, b]$ ,*

$$F(x) = \int_a^x f.$$

- *La fonction  $F : [a, b] \rightarrow \mathbb{R}$  est bien définie et lipschitzienne sur  $[a, b]$ .*
- *Si de plus  $f$  est continue sur  $[a, b]$  alors  $F$  est dérivable sur  $[a, b]$  et  $F' = f$ . En particulier, toute fonction continue sur un intervalle possède une primitive.*

*Dém.* La fonction  $f$  est bornée et continue par morceaux sur  $[a, x]$ , donc  $F(x)$  est bien définie.

Soit  $x, x' \in [a, b]$  et  $M = \sup_{[a, b]} |f|$ . Alors

$$\begin{aligned} |F(x') - F(x)| &= \left| \int_a^{x'} f - \int_a^x f \right| = \left| \int_x^{x'} f \right| \\ &\leq M|x' - x| \end{aligned}$$

et  $f$  est  $M$ -Lipschitzienne.

Soit  $\epsilon > 0$ . Si  $f$  est continue sur  $[a, b]$  et  $x_0 \in [a, b]$ , il existe  $\delta > 0$  tel que, si  $|x - x_0| < \delta$ ,  $|f(x) - f(x_0)| < \epsilon$ . Mais alors, si  $0 < h < \delta$

$$\begin{aligned} \left| \frac{F(x_0 + h) - F(x_0)}{h} - f(x_0) \right| &= \left| \frac{1}{h} \int_{x_0}^{x_0+h} (f - f(x_0)) \right| \\ &\leq \frac{1}{h} \int_{x_0}^{x_0+h} |f - f(x_0)| \\ &\leq \epsilon. \end{aligned}$$

En prenant  $h \rightarrow 0+$  on voit alors que  $F'(x_0+) = f(x_0)$ . On prouve de la même manière que  $F'(x_0-) = f(x_0)$ , donc  $F$  est dérivable et  $F' = f$ .  $\square$

**Théorème 4.8** (Théorème fondamental du calcul différentiel). *Si  $F: [a, b] \rightarrow \mathbb{R}$  est dérivable sur  $[a, b]$  et  $f = F'$  est Riemann-intégrable, alors*

$$\int_a^b f = F(b) - F(a).$$

*En particulier, si  $F$  est de classe  $C^1$  sur  $[a, b]$ , on a la formule  $\int_a^b F' = F(b) - F(a)$ .*

*Dém.* Soit  $\Delta = x_0, x_1, \dots, x_n$  une subdivision arbitraire de  $[a, b]$ , et  $I_k = [x_{k-1}, x_k]$ . Grâce au théorème des valeurs intermédiaires, appliqué à la fonction dérivable  $F$  sur l'intervalle  $I_k$ , il existe  $c_k \in I_k$  tel que

$$F(b) - F(a) = \sum_{k=1}^n [F(x_k) - F(x_{k-1})] = \sum_{k=1}^n |I_k| f(c_k).$$

Comme  $\inf_{I_k} f \leq f(c_k) \leq \sup_{I_k} f$ , on trouve

$$\Sigma_{\Delta}(f) \leq F(b) - F(a) \leq \Sigma^{\Delta}(f).$$

Mais  $f$  est Riemann-intégrable sur  $[a, b]$ . Si on passe au sup sur  $\Delta$  dans l'inégalité de gauche on trouve  $\int_a^b f \leq F(b) - F(a)$ . Si on passe à l'inf sur  $\Delta$  dans l'inégalité de droite on trouve  $F(b) - F(a) \leq \int_a^b f$ .  $\square$

## 4.6 Méthodes de calculs d'intégrales

Lorsqu'il s'agit de calculer explicitement l'intégrale d'une fonction  $x \mapsto f(x)$ , il est souvent pratique d'effectuer les calculs en utilisant la notation

$$\int_a^b f(x) dx \stackrel{\text{déf}}{=} \int_a^b f.$$

Le nom de la variable n'a pas d'importance et on peut remplacer la lettre  $x$  par une autre. On a, par exemple,  $\int_a^b f(x) dx = \int_a^b f(t) dt$ .

**Théorème 4.9** (Intégration par parties). *Si  $f$  et  $g$  sont deux fonctions de classe  $C^1$  sur  $[a, b]$ , alors*

$$\int_a^b fg' = f(b)g(b) - f(a)g(a) - \int_a^b f'g.$$

*Dém.* Exercice. □

**Théorème 4.10** (changement de variable). *Soit  $\phi: [a, b] \rightarrow [c, d]$  de classe  $C^1$  et  $f: [c, d] \rightarrow \mathbb{R}$  continue. Alors,*

$$\int_a^b f(\phi(x))\phi'(x) dx = \int_{\phi(a)}^{\phi(b)} f(t) dt.$$

*Dém.* En effet,  $f$  possède une primitive  $F$ . Alors  $x \mapsto F(\phi(x))$  est une primitive de  $x \mapsto f(\phi(x))\phi'(x)$ . On voit alors, grâce au théorème fondamental du calcul différentiel, que les deux termes de l'égalité sont égaux à  $F(\phi(b)) - F(\phi(a))$ . □

Dans la pratique, pour calculer  $\int_c^d f(t) dt$ , on peut chercher une fonction  $\phi$  comme ci-dessus et qui est en plus bijective, et considérer l'application inverse  $\phi^{-1}: [c, d] \rightarrow [a, b]$ . La formule de changement de variable revient alors à poser :

$$t = \phi(x), \quad dt = \phi'(x) dx, \quad \begin{cases} t = c \iff x = \phi^{-1}(c) \\ t = d \iff x = \phi^{-1}(d) \end{cases} \quad \text{et} \quad \int_c^d f(t) dt = \int_{\phi^{-1}(c)}^{\phi^{-1}(d)} f(x) dx.$$

## 4.7 Sommes de Riemann

**Définition 4.6.** *Une subdivision pointée  $\Delta^\bullet$  d'un intervalle  $[a, b]$  est une subdivision  $\Delta = x_0, x_1, \dots, x_n$  munie d'une suite finie de points  $\xi_1, \dots, \xi_n$ , telle que*

$$a = x_0 < x_1 < \dots < x_n = b, \quad \text{et} \quad \xi_k \in I_k \stackrel{\text{d\u00e9f}}{=} [x_{k-1}, x_k] \quad \text{pour } k = 1, \dots, n.$$

*Le pas de cette subdivision est*

$$\text{pas}(\Delta^\bullet) \stackrel{\text{d\u00e9f}}{=} \text{pas}(\Delta) = \max_{k=1, \dots, n} |I_k|.$$

**Définition 4.7.** *Si  $f: [a, b] \rightarrow \mathbb{R}$  est une fonction born\u00e9e, et  $\Delta^\bullet$  est une subdivision point\u00e9e de  $[a, b]$  comme ci-dessus, la somme de Riemann associ\u00e9e \u00e0  $f$  et \u00e0  $\Delta^\bullet$  est*

$$R(f, \Delta^\bullet) = \sum_{k=1}^n |I_k| f(\xi_k).$$

Comme les sommes de Darboux, une somme de Riemann s'interpr\u00eate comme l'aire d'un pluri-rectangle. Bien entendu, \u00e0 toute subdivision point\u00e9e  $\Delta^\bullet$  on peut associer la subdivision non-point\u00e9e  $\Delta$  correspondante : on a alors

$$\Sigma_\Delta(f) \leq R(f, \Delta^\bullet) \leq \Sigma^\Delta(f).$$

Le th\u00e9or\u00e8me suivant illustre l'\u00e9quivalence entre notre d\u00e9finition d'int\u00e9grabilit\u00e9 et la notion d'int\u00e9grabilit\u00e9 historiquement introduite par B. Riemann, qui repose sur l'existence de la limite des sommes de Riemann, lorsque le pas de la subdivision tend vers z\u00e9ro.

**D\u00e9finition 4.8.** *Soit  $f: [a, b] \rightarrow \mathbb{C}$  et soit  $\ell \in \mathbb{C}$ . L'\u00e9criture*

$$\lim_{\text{pas}(\Delta) \rightarrow 0} R(f, \Delta^\bullet) = \ell$$

*signifie que  $\forall \epsilon > 0, \exists \delta > 0$  tel que, pour toute subdivision point\u00e9e  $\Delta^\bullet$  de  $[a, b]$  :*

$$\text{pas}(\Delta) < \delta \implies |R(f, \Delta^\bullet) - \ell| < \epsilon. \quad (\star)$$

**Théorème 4.11.** Une fonction bornée  $f: [a, b] \rightarrow \mathbb{R}$  est Riemann intégrable si et seulement si la limite  $\lim_{\text{pas}(\Delta) \rightarrow 0} R(f, \Delta^\bullet)$  existe. Dans ce cas,

$$\lim_{\text{pas}(\Delta) \rightarrow 0} R(f, \Delta^\bullet) = \int_a^b f.$$

*Dém.* \* On peut se ramener aux fonctions à valeurs dans  $\mathbb{R}$  en prenant la partie réelle et imaginaire.

Supposons que  $f$  soit Riemann intégrable (au sens des sommes de Darboux). Soit  $\epsilon > 0$ . Il s'agit de trouver  $\delta > 0$  vérifiant (\*). Pour cela, posons  $M = \sup_{[a,b]} |f|$ . Par l'hypothèse, on sait qu'il existe une subdivision  $\nabla = x_0, \dots, x_n$  telle que

$$\Sigma^\nabla(f) - \epsilon \leq \int_a^b f \leq \Sigma^\nabla(f) + \epsilon.$$

Notons, comme d'habitude,  $|I_k| = x_k - x_{k-1}$ . Démontrons que, si on choisit

$$0 < \delta < \min\left(\frac{\epsilon}{2(n+1)M}, \min_{k=1, \dots, n} |I_k|\right),$$

alors la condition (\*) est satisfaite.

Soit  $\Delta^\bullet$  une subdivision pointée telle que  $\text{pas}(\Delta) < \delta$ . Le choix de  $\delta$  implique que tous les intervalles  $I_k$  de la subdivision de départ  $\nabla$  contiennent au moins un point de la nouvelle subdivision  $\Delta$ . Observons que (faire un dessin pour comprendre la deuxième inégalité<sup>†</sup>)

$$\begin{aligned} R(f, \Delta^\bullet) &\leq \Sigma^\Delta(f) \\ &\leq M\delta(n+1) + \Sigma^\nabla(f) \\ &\leq \int_a^b f + 2\epsilon. \end{aligned}$$

Un calcul semblable donne

$$\int_a^b f - 2\epsilon \leq R(f, \Delta^\bullet).$$

Ceci prouve la condition cherchée (\*).

Réciproquement, supposons que

$$\lim_{\text{pas}(\Delta) \rightarrow 0} R(f, \Delta^\bullet) = \ell.$$

Soit  $\epsilon > 0$ . Par l'hypothèse, nous pouvons trouver  $\delta > 0$  et une subdivision  $\Delta = x_0, x_1, \dots, x_n$ , de pas inférieure à  $\delta$ , telle que, pour n'importe quel « pointage »  $\xi_1, \dots, \xi_n$  de cette subdivision, on a

$$\left| \sum_{k=1}^n |I_k| f(\xi_k) - \ell \right| < \epsilon.$$

Prenons  $\xi_k$  tel que  $f(\xi_k) \approx \sup_{I_k} f$  (l'égalité  $f(\xi_k) = \sup_{I_k} f$  est possible si le sup de  $f$  est atteint sur  $I_k$ , mais ce n'est pas toujours le cas). Un choix qui convient est de prendre  $\xi_k$  tel que

$$\sup_{I_k} f - \epsilon/(b-a) \leq f(\xi_k)$$

---

\*. Non exigible à l'examen

†. Ici, chaque terme de  $\Sigma^\Delta(f)$  exprime l'aire d'un rectangle étroit, de base  $\leq \delta$ . On distingue alors les rectangles étroits dont la base contiennent les points  $x_0, x_1, \dots, x_n$  (dont chacun a une aire majoré par  $M\delta$ ), des rectangles étroits dont la base est contenue dans un intervalle de type  $I_k$ . Les premiers ont une somme des aires majoré par  $M\delta(n+1)$ . Les autres ont une somme des aires majoré par  $\Sigma^\nabla(f)$ .



(ce qui est toujours possible par la propriété du sup). Ainsi,

$$\Sigma^\Delta(f) = \sum_{k=1}^n |I_k| \sup_{I_k} f \leq \sum_{k=1}^n |I_k| f(\xi_k) + \epsilon \leq \ell + 2\epsilon.$$

Avec un raisonnement semblable (en choisissant d'autres points  $\xi_k$  tels que  $f(\xi_k) \approx \inf_{I_k} f$ ), on trouve

$$\ell - 2\epsilon \leq \Sigma_\Delta(f).$$

En conclusion, nous avons trouvé une subdivision telle que

$$\Sigma^\Delta(f) - 2\epsilon \leq \ell \leq \Sigma_\Delta(f) + 2\epsilon.$$

Ceci implique la Riemann-intégrabilité de  $f$  et que  $\int_a^b f = \ell$ . □

Le corollaire suivant s'obtient en reconnaissant l'expression des sommes de Riemann dans le cas d'une subdivision uniforme,  $x_k = a + (b - a)/n$ , avec un pointage aux points  $\xi_k = x_k$  ( $k = 1, \dots, n$ ).

**Corollaire 4.12.** *Si  $f: [a, b] \rightarrow \mathbb{R}$  est une fonction bornée Riemann-intégrable sur  $[a, b]$ , alors*

$$\frac{b-a}{n} \sum_{k=1}^n f(a + k(b-a)/n) \rightarrow \int_a^b f.$$

**Définition 4.9** (convergence simple). *Soit  $D \subset \mathbb{R}$  et  $f_n: D \rightarrow \mathbb{R}$ , ( $n = 0, 1, 2, \dots$ ) et soit  $f: D \rightarrow \mathbb{R}$ . On dit que la suite  $(f_n)_{n \in \mathbb{N}}$  converge simplement vers  $f$  dans  $D$  si :*

$$\forall \epsilon > 0 \quad \forall x \in D, \quad \exists n_0 \quad \text{tel que,} \quad \forall n \geq n_0, \quad |f_n(x) - f(x)| \leq \epsilon.$$

Autrement dit,

$$\forall x \in D \quad \lim_{n \rightarrow +\infty} f_n(x) = f(x).$$

Cette notion de convergence est bien moins contraignante que la notion de convergence uniforme. Par exemple, la suite de fonction  $f_n: [0, 1] \rightarrow \mathbb{R}$ , définie par  $f_n(x) = x^n$  converge simplement vers la fonction  $f(x) = 0$  si  $0 \leq x < 1$  et telle que  $f(1) = 1$ , mais pas uniformément.

## 4.8 Intégrale double de Riemann

Soit  $D \subset \mathbb{R}^2$  un ensemble borné et  $f: D \rightarrow \mathbb{R}$ .

# 5 Courbes paramétrées

## 5.1 Courbes paramétrées

Dans tout ce chapitre  $\|x\|$  désigne la norme euclidienne du vecteur  $x \in \mathbb{R}^n$ .

Une courbe dans  $\mathbb{R}^n$  est un ensemble  $\Gamma \subset \mathbb{R}^n$  tel qu'il existe une application continue  $\varphi: [a, b] \rightarrow \mathbb{R}^n$ , où  $[a, b]$  est un intervalle de  $\mathbb{R}$ , telle que

$$\Gamma = \{\varphi(t) : t \in [a, b]\}.$$

On dit aussi que  $\varphi$  est une *paramétrisation* de  $\Gamma$ , ou aussi que  $\Gamma$  est le *support* de la courbe paramétrée  $\varphi$ . Une paramétrisation permet de définir une *orientation* (c'est-à-dire un sens de parcours) sur  $\Gamma$  : le parcours allant du point  $\varphi(a)$  au point  $\varphi(b)$ .

**Interpretation cinématique.** Si  $x(t)$ ,  $y(t)$  et  $z(t)$  sont les coordonnées d'un point matériel mobile à l'instant  $t$ , la loi horaire du mouvement, c'est à dire l'application  $t \mapsto (x(t), y(t), z(t))$  définit une courbe paramétrée  $\varphi$  dans  $\mathbb{R}^3$ . Dans cet interprétation cinématique, le support  $\Gamma$  de la courbe est la *trajectoire* du point matériel. La *vitesse* instantanée du point est donnée par le vecteur  $\varphi'(t) = (x'(t), y'(t), z'(t))$ .

**Définition 5.1.** - On dit qu'une paramétrisation  $\varphi: [a, b] \rightarrow \mathbb{R}^n$  d'une courbe  $\Gamma$  est régulière si  $\varphi$  est de classe  $C^1$  (c'est à dire que ses composantes  $\varphi_1, \dots, \varphi_n$  sont toutes des fonctions de classe  $C^1$ ) et si, pour tout  $t \in [a, b]$ , le vecteur  $\varphi'(t) \in \mathbb{R}^n$  n'est pas le vecteur  $(0, \dots, 0)$ . S'il existe une partition finie  $a = a_0 \leq a_1 \leq a_2 \leq \dots \leq a_N = b$  de l'intervalle  $[a, b]$  telle que  $\varphi$  est régulière dans tout intervalle  $[a_i, a_{i+1}]$  on dit que  $\varphi$  est régulière par morceaux. Un point où  $\varphi'(t_0)$  s'annule est dit point singulier.

- On dit qu'une courbe paramétrée  $\varphi: [a, b] \rightarrow \mathbb{R}^n$  est fermée si  $\varphi(a) = \varphi(b)$ .
- On dit qu'une courbe paramétrée  $\varphi: [a, b] \rightarrow \mathbb{R}^n$  est simple si  $\varphi(t) \neq \varphi(t')$ , pour tout  $t \neq t'$  et  $a < t < b$ ,  $a < t' < b$ . Cela signifie que la courbe n'a pas d'intersection avec elle même.

**Exemple 5.1.** - Si  $f: [a, b] \rightarrow \mathbb{R}$  est une fonction de classe  $C^1$ , on peut toujours lui associer une courbe paramétrée dans  $\mathbb{R}^2$  régulière, simple, non fermée, en prenant son graphe. Cela revient à poser  $\varphi(t) = (t, f(t))$ ,  $t \in [a, b]$ .

- La fonction  $\varphi(t) = (\cos(t), \sin(t))$ , où  $t \in [0, 2\pi]$  définit une paramétrisation régulière, simple, fermée. Le support de cette courbe paramétrée est le cercle de centre  $O$  et de rayon 1.
- La courbe paramétrée  $\varphi(t) = (t^2, t^3)$ ,  $t \in [-2, 2]$  est simple non régulière. En effet, observons que le vecteur dérivée  $\varphi'(t) = (2t, 3t^2)$  s'annule pour  $t = 0$ . Pour dessiner cette courbe, on part du système  $\begin{cases} x = t^2 \\ y = t^3 \end{cases}$  et on obtient, en éliminant  $t$  (il faut distinguer  $t \geq 0$  et  $t < 0$ )  $y = \pm x^{2/3}$ , avec  $0 \leq x \leq 4$ . On voit alors sans peine que le support  $\Gamma$  de cette courbe présente un cusp à l'origine.
- Soit  $a \in \mathbb{R}^n$  et  $v$  un vecteur de  $\mathbb{R}^n$ . L'équation paramétrique de la droite  $d$  passant par le point  $a = (a_1, \dots, a_n)$  et parallèle au vecteur  $v = (v_1, \dots, v_n)$  est l'application

$$\varphi(t) = a + tv \quad t \in \mathbb{R}$$

Si  $t_0 \in \mathbb{R}$ , la même droite  $d$  peut aussi être paramétrée par

$$\varphi(t) = a + (t - t_0)v \quad t \in \mathbb{R}.$$

- Les bords des polygones sont des exemples typiques de courbes qui admettent des paramétrisations régulières par morceaux, mais pas de paramétrisation régulières.

**Droite et vecteur tangent.** Soit  $\varphi$  une paramétrisation régulière, définie sur un intervalle  $[a, b]$ , et soit  $t_0$  et  $t_1$  deux points de cet intervalle. La droite passant par les points  $\varphi(t_0)$  et  $\varphi(t_1)$  s'identifie à la droite passant par  $\varphi(t_0)$  et parallèle au vecteur  $v = \frac{\varphi(t_1) - \varphi(t_0)}{t_1 - t_0}$ . Ainsi, cette droite est paramétrée par l'application

$$t \mapsto \varphi(t_0) + (t - t_0) \frac{\varphi(t_1) - \varphi(t_0)}{t_1 - t_0}, \quad t \in \mathbb{R}.$$

En passant à la limite pour  $t_1 \rightarrow t_0$  on trouve

$$\varphi(t_0) + \varphi'(t_0)(t - t_0), \quad t \in \mathbb{R}. \quad (5.1)$$

qui est l'équation paramétrique de la *droite tangente* au support de  $\varphi$  au point  $\varphi(t_0)$ . Observons que cette droite est parallèle au vecteur  $\varphi'(t_0)$ . Le vecteur

$$\vec{T}(t_0) = \frac{\varphi'(t_0)}{\|\varphi'(t_0)\|}$$

qui est de norme unitaire, est donc le *verseur tangent* à la courbe au point  $\varphi(t_0)$ .

## 5.2 Longueur d'une courbe

Pour définir la longueur d'une courbe paramétrée  $\varphi: [a, b] \rightarrow \mathbb{R}^n$ , considérons la famille  $\mathcal{P}$  de toutes les partitions finies possibles  $a = a_0 < a_1 < \dots < a_N = b$  de l'intervalle  $[a, b]$ . À partir d'une telle partition, on considère les points  $\varphi(a_i)$ ,  $i = 0, \dots, N$  appartenant au support de la courbe et les segments de  $\mathbb{R}^n$ ,  $[\varphi(a_i), \varphi(a_{i+1})] = \{(1-t)\varphi(a_i) + t\varphi(a_{i+1}) : t \in [0, 1]\}$ . Si  $P \in \mathcal{P}$  est une telle partition, la quantité

$$L(P) = \sum_{i=0}^{N-1} \|\varphi(a_i) - \varphi(a_{i+1})\|$$

n'est rien d'autre que la longueur de la ligne affine par morceaux qui relie les points  $\varphi(a_i)$ . Il est alors naturel de définir la longueur de la courbe comme

$$\text{Long}(\varphi) = \sup\{L(P) : P \in \mathcal{P}\}.$$

On peut démontrer que si  $\varphi$  est une courbe régulière par morceaux, alors la quantité ci-dessus est toujours finie.

**Théorème 5.1.** *Pour toute courbe paramétrée régulière par morceaux, on a*

$$\text{Long}(\varphi) = \int_a^b \|\varphi'(t)\| dt.$$

La démonstration de ce théorème est hors programme.

**Exemple 5.2** (Longueur du cercle). En paramétrant le cercle centré en  $O$  et de rayon  $R$  par  $\varphi(t) = (R \cos t, R \sin t)$ ,  $t \in [0, 2\pi]$  on voit que sa longueur est  $\int_0^{2\pi} \sqrt{(-R \sin t)^2 + (R \cos t)^2} dt = 2\pi R$ .

## 5.3 Paramétrisations équivalentes et abscisse curviligne

Une même courbe  $\Gamma$  peut être paramétrée de plusieurs manière différentes.

**Définition 5.2.** *On dit que deux paramétrisations régulières par morceaux  $\varphi: [a, b] \rightarrow \mathbb{R}^n$  et  $\psi: [c, d] \rightarrow \mathbb{R}^n$  d'une même courbe sont équivalentes s'il existe un difféomorphisme  $h: [a, b] \rightarrow [c, d]$  ( $h$  est donc une application bijective et  $h$  et  $h^{-1}$  sont de classe  $C^1$ ) telle que, pour tout  $t \in [a, b]$ , on a*

$$\varphi(t) = \psi(h(t)).$$

Dans ce cas on écrit  $\varphi \sim \psi$

La relation  $\sim$  est une relation d'équivalence. Observons que pour un difféomorphisme  $h: [a, b] \rightarrow [c, d]$  il y a deux possibilité : soit  $h$  est strictement croissante, soit  $h$  est strictement décroissante (c'est une conséquence de la bijectivité, de la continuité, et du théorème des valeurs intermédiaires). Dans le premier cas on dit que  $\varphi$  et  $\psi$  ont la même orientation.

**Exemple 5.3.** Un quart de cercle peut être paramétrée par  $t \mapsto (R \cos t, R \sin t)$ ,  $t \in [0, \pi/2]$ , mais aussi, en tant que graphe d'une fonction, par  $x \mapsto (x, \sqrt{R^2 - x^2})$ ,  $x \in [0, R]$ . On construit le difféomorphisme  $h: [0, \pi/2] \rightarrow [0, R]$  en posant  $x = h(t) = R \cos t$ . Les deux paramétrisations sont alors équivalentes. Observons que le sens de parcours n'est pas le même dans ce deux cas. En effet la fonction  $h$  est décroissante.

**Remarque 5.4.** Si  $\varphi: [a, b] \rightarrow \mathbb{R}^n$  et  $\psi: [c, d] \rightarrow \mathbb{R}^n$ , avec  $\varphi \sim \psi$  par un difféomorphisme  $h$ , et si  $s_0 = h(t_0)$ , alors le vecteur tangent au point  $\varphi(t_0) = \psi(s_0)$  est

$$\vec{T}(t_0) = \frac{\varphi'(t_0)}{\|\varphi'(t_0)\|} = \frac{\psi'(h(t_0))h'(t_0)}{\|\psi'(h(t_0))\| |h'(t_0)|} = \pm \frac{\psi'(s_0)}{\|\psi'(s_0)\|},$$

avec le signe  $+$  si  $\varphi$  et  $\psi$  ont la même orientation et le signe  $-$  dans le cas contraire. Ce calcul montre que la droite tangente à une courbe  $\Gamma$  ne dépend pas de la paramétrisation choisie.

**Définition 5.3.** Une paramétrisation normale, ou paramétrisation par abscisse curviligne d'une courbe  $\Gamma$ , est une paramétrisation  $\psi: [0, \text{Long}(\Gamma)] \rightarrow \mathbb{R}^n$  telle que  $\|\psi'(s)\| = 1$  pour tout  $0 \leq s \leq \text{Long}(\Gamma)$ .

**Proposition 5.2.** Toute courbe  $\Gamma$  régulière par morceaux admet une paramétrisation normale équivalente

*Dém.* À partir d'une paramétrisation régulière par morceaux  $\varphi: [a, b] \rightarrow \mathbb{R}^n$ , construisons une paramétrisation normale  $\psi \sim \varphi$ . Pour cela il suffit de définir  $h: [a, b] \rightarrow [0, \text{Long}(\varphi)]$ , avec

$$h(t) = \int_a^t \|\varphi'(\tau)\| d\tau.$$

Cette application  $h$  définit bien un difféomorphisme entre  $[a, b]$  et  $[0, \text{Long}(\varphi)]$ . Ensuite il suffit de poser  $\psi(s) = \varphi(h^{-1}(s))$  (ou  $\varphi(t) = \psi(h(t))$ , de manière équivalente). Observons que  $\psi'(s) = \varphi'(h^{-1}(s))(h^{-1})'(s)$ . Mais, pour tout  $s$ ,  $h(h^{-1}(s)) = s$  et en dérivant terme-à-terme cette relation on trouve  $(h^{-1})'(s) = 1/h'(h^{-1}(s))$  (c'est la formule usuelle de la dérivée de l'application inverse). Comme  $h'(t) = \|\varphi(t)\|$ , on trouve

$$\psi'(s) = \varphi'(h^{-1}(s)) \frac{1}{\|\varphi'(h^{-1}(s))\|}.$$

Mais alors  $\|\psi'(s)\| = 1$ .

□

L'abscisse curviligne (c'est-à-dire, le paramètre d'une paramétrisation normale) d'une courbe  $\Gamma$  est traditionnellement notée avec la lettre  $s$ . Observons que la différentielle  $ds = \|\varphi'(t)\| dt$  s'interprète comme un "élément infinitésimal de longueur de la courbe".

**Définition 5.4.** Soit  $f: \Gamma \rightarrow \mathbb{R}$  une fonction continue, définie sur le support  $\Gamma$  d'une courbe régulière par morceaux, paramétrée par  $\varphi: [a, b] \rightarrow \mathbb{R}$ . L'intégrale curviligne de  $f$  le long  $\Gamma$  est définie par

$$\int_{\Gamma} f ds = \int_a^b f(\varphi(t)) \|\varphi'(t)\| dt.$$

Observons que dans le cas particulier où  $\Gamma$  est un intervalle  $[a, b]$ , en le paramétrant par  $\varphi(t) = t$ , on trouve que  $\int_{\Gamma} f ds = \int_a^b f(t) dt$ . L'intégrale curviligne est donc une généralisation de l'intégrale usuelle.

Il est facile, à l'aide du théorème de changement de variables, de prouver que  $\int_{\Gamma} f ds$  est indépendant de la paramétrisation choisie pour  $\Gamma$ . En particulier (en prenant la fonction constante  $f = 1$ ) on voit que la longueur est indépendante du choix de la paramétrisation.

**Exemple 5.5.** Une paroi verticale est limitée par le haut par la courbe d'équation  $t \mapsto (t, t^2, 6t)$ ,  $t \in [0, 1]$ . Quelle est l'aire de la surface de cette paroi ? Réponse. Le profil au sol de la paroi est la courbe  $\Gamma$  de  $\mathbb{R}^2$  paramétrée par  $t \mapsto (t, t^2)$ ,  $t \in [0, 1]$ . Considérons ensuite la fonction  $f(x, y) = 6x$ . Il s'agit de calculer l'intégrale curviligne  $\int_{\Gamma} f \, ds = \int_0^1 \sqrt{1 + 4t^2} 6t \, dt = [(1 + 4t^2)^{3/2}]_{t=0}^{t=1} = 5^{3/2} - 1$ .

## 5.4 La courbure

Soit  $\psi: [0, \text{Long}(\Gamma)] \rightarrow \mathbb{R}^2$  la **paramétrisation normale** d'une courbe  $\Gamma$ . On note par  $s$  l'abscisse curviligne. On suppose que  $\varphi$  est de classe  $C^2$ .

On sait que  $\|\psi'(s)\| = 1$ , donc

$$\vec{T}(s) = \psi'(s),$$

est le vecteur tangent à la courbe au point  $\psi(s)$ . Soit  $\vec{N}(s)$  le vecteur normal à  $\vec{T}(s)$ , de manière que  $(\vec{T}(s), \vec{N}(s))$  soit une base orthonormée directe du plan\*.

**Proposition 5.3.** *Il existe une fonction continue  $s \mapsto \bar{\kappa}(s)$  telle que  $\psi''(s) = \bar{\kappa}(s)\vec{N}(s)$ . En particulier, si la courbe est paramétrée à l'aide de l'abscisse curviligne, alors la direction normale à la courbe au point  $\varphi(s)$  est donnée par le vecteur  $\varphi''(s)$ .*

*Dém.* En effet  $\psi'(s) \cdot \psi'(s) = \vec{T}(s) \cdot \vec{T}(s) = \|\vec{T}(s)\|^2 = 1$ . En dérivant par rapport à  $s$  on trouve  $2\psi''(s) \cdot \psi'(s) = 0$ . Mais alors  $\psi''(s)$  est orthogonal à  $\psi'(s)$ , et donc parallèle à  $\vec{N}(s)$ .  $\square$

**Définition 5.5.** La courbure algébrique  $\bar{\kappa}(s)$  de  $\psi$  au point  $\psi(s)$  est le nombre réel

$$\bar{\kappa}(s) = \psi''(s) \cdot \vec{N}(s) = \vec{T}'(s) \cdot \vec{N}(s).$$

La courbure de  $\varphi$  au point  $\varphi(s)$  est le nombre  $\geq 0$

$$\kappa(s) = |\bar{\kappa}(s)| = \|\psi''(s)\|.$$

Le rayon de courbure est  $R(s) = \frac{1}{\kappa(s)}$ .

Si  $\kappa(s) \neq 0$ , on dit que le point  $\psi(s)$  est birégulier.

**Exemple 5.6.** Le cercle de rayon  $R$  admet la paramétrisation normale  $\psi(s) = (R \cos \frac{s}{R}, R \sin \frac{s}{R})$ . La courbure vaut  $\kappa(s) = \|\psi''(s)\| = \frac{1}{R}$ . La courbure d'un cercle de rayon  $R$  est donc constante et égale à  $1/R$ . Le rayon de courbure du cercle est alors constant égale à  $R$ .

Étant donnée une courbe  $\Gamma$ , il n'est pas aisé d'expliciter la paramétrisation normale (en effet, l'intégrale de longueur est souvent difficile à calculer). À partir d'une paramétrisation quelconque  $t \mapsto \varphi(t)$ , on peut trouver la courbure à l'aide de la proposition suivante :

**Proposition 5.4.** *Soit  $\varphi: [a, b] \rightarrow \mathbb{R}^2$  une courbe paramétrée régulière de classe  $C^2$ . Alors pour tout  $t \in [a, b]$  on a*

$$\bar{\kappa}(t) = \frac{\det(\varphi'(t), \varphi''(t))}{\|\varphi'(t)\|^3}, \quad \text{et} \quad \kappa(t) = \frac{|\det(\varphi'(t), \varphi''(t))|}{\|\varphi'(t)\|^3}.$$

*Dém.* Considérons la paramétrisation normale  $\psi(s) = \varphi(t(s))$  de  $\Gamma$ , où  $t(s) = h^{-1}(s)$ ,  $h$  étant le difféomorphisme introduit dans la démonstration de la Proposition 5.2. Nous avons (voir la démonstration de la Proposition 5.2) :

$$\psi'(s) = \frac{\varphi'(t(s))}{\|\varphi'(t(s))\|}.$$

---

\*. Rappelons que si  $\vec{T}$  et  $\vec{N}$  sont deux vecteurs dans le plan orthogonaux, de norme unitaire, alors le déterminant de la matrice  $2 \times 2$   $(\vec{T}, \vec{N})$  est de déterminant  $\pm 1$ . Affirmer que  $(\vec{T}(s), \vec{N}(s))$  est une "base directe" signifie que le déterminant est positif

En dérivant par rapport à  $s$  on trouve

$$\psi''(s) = \frac{\varphi''(t(s))}{\|\varphi'(t(s))\|^2} + \lambda(s)\varphi'(t(s)),$$

où  $\lambda(s)$  est la dérivée de la fonction  $s \mapsto \frac{1}{\|\varphi'(t(s))\|}$ , qu'on n'aura pas besoin de calculer. Donc, en appliquant les propriétés élémentaires du déterminant  $\det(\vec{a}, \vec{b} + \lambda\vec{c}) = \det(\vec{a}, \vec{b}) + \lambda \det(\vec{a}, \vec{c})$  et  $\det(\vec{c}, \vec{c}) = 0$ , on voit que

$$\det(\psi'(s), \psi''(s)) = \frac{\det(\varphi'(t), \varphi''(t))}{\|\varphi'(t)\|^3}.$$

D'autre part,

$$\det(\psi'(s), \psi''(s)) = \det(\vec{T}(s), \bar{\kappa}(s)\vec{N}(s)) = \bar{\kappa}(s).$$

□

## 5.5 Étude locale des courbes planes

Considérons une paramétrisation normale  $\psi: [0, \text{Long}(\Gamma)] \rightarrow \mathbb{R}^2$  d'une courbe  $\Gamma$ , et supposons  $\psi$  de classe  $C^2$ .

Au voisinage de  $s_0 \in [0, \text{Long}(\Gamma)] \rightarrow \mathbb{R}^2$ , en écrivant un développement limité d'ordre 2 on trouve

$$\psi(s) = \psi(s_0) + (s - s_0)\psi'(s_0) + \frac{(s - s_0)^2}{2}\psi''(s_0) + o((s - s_0)^2).$$

On sait déjà que  $s \mapsto \varphi(s_0) + (s - s_0)\varphi'(s_0)$  est la paramétrisation de la droite tangente à la courbe au point  $\psi(s_0)$ . Le terme suivant du développement  $\frac{(s-s_0)^2}{2}\psi''(s_0)$  (son signe, notamment) nous renseigne sur le positionnement de la courbe par rapport à cette droite, au voisinage de  $\psi(s_0)$ . Plus précisément, récrivons le développement précédent comme

$$\psi(s) = \psi(s_0) + (s - s_0)\vec{T}(s_0) + \bar{\kappa}(s_0)\frac{(s - s_0)^2}{2}\vec{N}(s_0) + o((s - s_0)^2).$$

Si  $\psi(s_0)$  est un point bi-régulier, les vecteurs  $\vec{T}(s_0)$  et  $\vec{N}(s_0)$  sont bien définis. Dans le repère orthonormé  $(\psi(s_0), \vec{T}(s_0), \vec{N}(s_0))$ , c'est-à-dire le repère avec pour origine  $\psi(s_0)$  et base  $(\vec{T}(s_0), \vec{N}(s_0))$ , la courbe  $\Gamma$  admet alors pour composantes (en négligeant les termes d'ordre supérieur)

$$(s - s_0, \bar{\kappa}(s_0)\frac{(s - s_0)^2}{2}).$$

Ainsi, dans ce repère la courbe se comporte localement comme une parabole.

**Exemple d'étude d'un point singulier** Si  $\varphi: [a, b] \rightarrow \mathbb{R}^2$  est une paramétrisation d'une courbe admettant un point singulier  $\varphi(t_0)$ , la formule pour la construction de la droite tangente (5.1) ne s'applique plus. Certaines courbes admettent cependant une droite tangente aux points singuliers. Si c'est le cas, la pente de cette droite est donnée par la limite

$$\ell = \lim_{t \rightarrow t_0} \frac{y(t) - y(t_0)}{x(t) - x(t_0)}.$$

La droite est alors celle d'équation  $y = \ell(x - x(t_0)) + y(t_0)$ .

**Exemple 5.7.** Étudions le point singulier de la courbe paramétrée par  $x(t) = 1 + t^2$  et  $y(t) = -1 + t^2 - t^3$ . Le point singulier est  $(x(0), y(0)) = (1, -1)$ . La limite  $\ell = \lim_{t \rightarrow 0} \frac{y(t) - y(0)}{x(t) - x(0)}$  existe et vaut  $\ell = 1$ . La droite tangente au point singulier est donc la droite d'équation  $y = x - 2$ . Pour connaître l'allure de la courbe au voisinage du point singulier, étudions la position de la courbe par rapport au point  $(1, -1)$  : pour cela observons que  $x(t) - 1 = t^2 \geq 0$  et  $y(t) + 1 = t^2 - t^3 \geq 0$  au voisinage de  $t = 0$ , la courbe reste alors à la droite et en haut par rapport au point  $(1, -1)$ , au voisinage de celui-ci. Étudions ensuite la position de la courbe par rapport à sa droite tangente : on a  $y(t) - x(t) + 2 = -t^3$ , qui change de signe en  $t = 0$ . Ainsi la courbe est à la fois au dessus et en dessous de sa droite tangente au voisinage du point  $(1, -1)$ . Ces renseignements nous permettent de dire que la courbe présente en  $(1, -1)$  une singularité de type "cusp".

## 6 Nappes paramétrées et surfaces de $\mathbb{R}^3$

### 6.1 Rappel sur le produit vectoriel

Soit  $(\mathbf{i}, \mathbf{j}, \mathbf{k})$  la base canonique de  $\mathbb{R}^3$ . Considérons deux vecteurs  $\vec{a}$  et  $\vec{b}$  de  $\mathbb{R}^3$  non co-linéaires. Rappelons que le vecteur

$$\vec{a} \wedge \vec{b} = \det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} \quad (6.1)$$

est le vecteur orthogonal à  $\vec{a}$  et  $\vec{b}$ , de norme  $\|\vec{a} \wedge \vec{b}\| = \|\vec{a}\| \|\vec{b}\| |\sin \theta|$ , où  $\theta$  est l'angle orienté compris entre  $\vec{a}$  et  $\vec{b}$  et tel que  $(\vec{a}, \vec{b}, \vec{a} \wedge \vec{b})$  forme une base directe de  $\mathbb{R}^3$ . La norme  $\|\vec{a} \wedge \vec{b}\|$  est donc égale à l'aire du parallélogramme engendrée par  $\vec{a}$  et  $\vec{b}$ .

Le produit mixte

$$[\vec{a}, \vec{b}, \vec{u}] = (\vec{a} \wedge \vec{b}) \cdot \vec{u} = \det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ u_1 & u_2 & u_3 \end{pmatrix}$$

s'interprète comme le volume (avec signe) du parallélépipède engendré par les vecteurs  $\vec{a}$ ,  $\vec{b}$  et  $\vec{u}$ .

### 6.2 Surfaces et nappes paramétrées

Une *surface* est une partie  $\Sigma \subset \mathbb{R}^3$ , de la forme

$$\Sigma = \{\varphi(u, v) : u, v \in K\} \subset \mathbb{R}^3.$$

où  $\varphi$  est une application  $\varphi : K \rightarrow \mathbb{R}^3$ , et  $K = \overline{U}$  est un compact de  $\mathbb{R}^2$ , égale à l'adhérence d'un ouvert  $U \subset \mathbb{R}^2$ . L'application  $\varphi$  est une *paramétrisation* de  $\Sigma$ . On dit aussi que  $\varphi$  est une nappe paramétrée et que l'ensemble  $\Sigma$  est le support de la nappe.

**Exemple 6.1.** (paramétrisation d'un plan de  $\mathbb{R}^3$ ) Soient  $\vec{a}$  et  $\vec{b}$  deux vecteurs de  $\mathbb{R}^3$  non co-linéaires. Ces deux vecteurs engendrent un plan  $\pi$  passant par l'origine, qui est le sous-espace vectoriel de  $\mathbb{R}^3$  donné par  $\pi = \{u(a_1, a_2, a_3) + v(b_1, b_2, b_3) : u, v \in \mathbb{R}\}$ .

Si  $\vec{P} = (\bar{x}_1, \bar{x}_2, \bar{x}_3) \in \mathbb{R}^3$ , alors le plan passant par  $\vec{P}$  et parallèle à  $\pi$  admet comme équation paramétrique

$$\varphi(u, v) = (\bar{x}_1, \bar{x}_2, \bar{x}_3) + u(a_1, a_2, a_3) + v(b_1, b_2, b_3), \quad u, v \in \mathbb{R}.$$

Ce même plan consiste en l'ensemble points  $P = (x_1, x_2, x_3)$  tels que le vecteur  $\overrightarrow{PP}$  est orthogonal à  $\vec{a} \wedge \vec{b}$ . Ainsi,

$$(x_1, x_2, x_3) \in \pi \iff (\vec{a} \wedge \vec{b}) \cdot \overrightarrow{PP} = 0 \iff \det \begin{pmatrix} (x - \bar{x})_1 & (x - \bar{x})_2 & (x - \bar{x})_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} = 0.$$

La dernière égalité est l'équation cartésienne du plan passant par  $\bar{P}$ , de vecteurs directeurs  $\vec{a}$  et  $\vec{b}$ .

**Exemple 6.2.** La sphère centrée en  $O$  et de rayon  $R$  peut se paramétrer à l'aide de la longitude et de la co-latitude par l'application

$$(\phi, \theta) \mapsto (R \sin \theta \cos \phi, R \sin \theta \sin \phi, R \cos \theta), \quad \phi \in [0, 2\pi[, \quad \theta \in [0, \pi].$$

Posons  $\varphi_u = (\frac{\partial \varphi_1}{\partial u}, \frac{\partial \varphi_2}{\partial u}, \frac{\partial \varphi_3}{\partial u})$  et  $\varphi_v = (\frac{\partial \varphi_1}{\partial v}, \frac{\partial \varphi_2}{\partial v}, \frac{\partial \varphi_3}{\partial v})$ .

**Définition 6.1.** On dit qu'une paramétrisation  $\varphi: K \rightarrow \mathbb{R}^3$  d'une surface  $\Sigma$  est régulière si  $\varphi$  est de classe  $C^1$ , injective dans  $\overset{\circ}{K}$ , et, pour tout  $(u, v)$  à l'intérieur de  $K$ , les deux vecteurs de  $\mathbb{R}^3$   $\varphi_u(u, v)$  et  $\varphi_v(u, v)$  sont linéairement indépendants.

Justifions la définition précédente. Fixons  $v = v_0$  et faisons varier le paramètre  $u$ . L'application  $u \mapsto \varphi(u, v_0)$  définit une courbe paramétrée dans  $\mathbb{R}^3$ , dont le support est contenu sur la surface  $\Sigma$ . De la même manière, si on fixe  $u = u_0$  et on fait varier  $v$ , on obtient une autre courbe paramétrée  $v \mapsto \varphi(u_0, v)$  sur la surface. Ces deux courbes admettent comme vecteurs tangents au point  $P_0 = \varphi(u_0, v_0) \in \mathbb{R}^3$  respectivement les vecteurs  $\vec{a} = \frac{\partial \varphi}{\partial u}(u_0, v_0)$  et  $\vec{b} = \frac{\partial \varphi}{\partial v}(u_0, v_0)$ . Sous l'hypothèse de paramétrisation régulière, ces deux vecteurs sont linéairement indépendants et donc ils engendrent un plan dans  $\mathbb{R}^3$ . Il s'agit du *plan tangent* à  $\Sigma$  au point  $P_0$ .

Le plan tangent à  $\Sigma$  au point  $P_0$  a donc pour équation paramétrique

$$(u, v) \mapsto \varphi(u_0, v_0) + u \varphi_u(u_0, v_0) + v \varphi_v(u_0, v_0), \quad u, v \in \mathbb{R}.$$

Le *verseur normal* à une surface  $\Sigma$  paramétrée par  $(u, v) \mapsto \varphi(u, v)$ , au point  $\varphi(u, v)$  est alors donnée par la formule

$$\vec{\nu}(u_0, v_0) = \frac{(\varphi_u \wedge \varphi_v)(u_0, v_0)}{\|(\varphi_u \wedge \varphi_v)(u_0, v_0)\|}.$$

**Exemple 6.3.** Le cylindre  $\Sigma$  de base le disque  $\{(x, y): x^2 + y^2 \leq R^2\}$  et hauteur infinie peut être paramétrée par

$$(\theta, z) : (\cos \theta, \sin \theta, z), \quad \theta \in [0, 2\pi], \quad z \in \mathbb{R}.$$

Le plan tangent au cylindre au point  $(\cos \theta_0, \sin \theta_0, z_0)$  est le plan paramétré par

$$(\theta, z) \mapsto (\cos \theta_0, \sin \theta_0, z_0) + \theta(-\sin \theta_0, \cos \theta_0, 0) + z(0, 0, 1).$$

L'équation cartésienne de ce plan est

$$0 = \det \begin{pmatrix} x - \cos \theta_0 & y - \sin \theta_0 & z - z_0 \\ -\sin \theta_0 & \cos \theta_0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = x \cos \theta_0 + y \sin \theta_0 - 1$$

**Exemple 6.4.** Le graphe d'une fonction de 2 variables  $f$  peut être paramétré par

$$(x, y) \mapsto (x, y, f(x, y)).$$

Le *verseur normal* au graphe est alors donné par la formule

$$\nu = \frac{1}{\sqrt{1 + |\nabla f|^2}}(-f_x, -f_y, 1).$$



### 6.3 Aire d'une surface paramétrée et intégrale de surface

Considérons la surface  $\Sigma$  paramétrée par  $(u, v) \mapsto \varphi(u, v)$ , où  $(u, v) \in K$ . Considérons un grillage de  $K$ , le subdivisant en petits rectangles d'aire  $du dv$ , et soit  $R_0$  l'un de ces rectangles, de sommets  $(u_0, v_0)$ ,  $(u_0 + du, v_0)$ ,  $(u_0, v_0 + dv)$  et  $(u_0 + du, v_0 + dv)$ . L'application  $\varphi$  envoie  $R_0$  en une petite portion de la surface  $\varphi(R_0)$ . Mais en appliquant la formule de Taylor d'ordre 1 (en négligeant le reste),

$$\varphi(u_0 + du, v_0) \simeq \varphi(u_0, v_0) + \varphi_u(u_0, v_0) \quad \text{et} \quad \varphi(u_0, v_0 + dv) \simeq \varphi(u_0, v_0) + \varphi_v(u_0, v_0).$$

Ainsi, la portion de surface  $\varphi(R_0)$  est approximativement le parallélogramme construit à l'aide de vecteurs  $\varphi_u(u_0, v_0)$  et  $\varphi_v(u_0, v_0)$ , pointés en  $\varphi(u_0, v_0)$ . Ce parallélogramme est d'aire  $\|(\varphi_u \wedge \varphi_v)(u_0, v_0)\| du dv$ . Ainsi la quantité

$$d\sigma \stackrel{\text{déf}}{=} \|(\varphi_u \wedge \varphi_v)(u, v)\| du dv$$

s'interprète comme un "élément infinitesimal d'aire" sur la surface. Cela conduit naturellement à la définition suivante :

$$\text{Aire}(\Sigma) = \int_K \|(\varphi_u \wedge \varphi_v)(u, v)\| du dv.$$

**Exemple 6.5.** La formule précédente, appliquée à la sphère de rayon  $R$ , donne l'aire  $A = 4\pi R^2$ .

**Définition 6.2.** Soit  $f: \Sigma \rightarrow \mathbb{R}$  une fonction continue, définie sur une surface  $\Sigma$  régulière, paramétrée par  $(u, v) \mapsto \varphi(u, v)$ , où  $(u, v) \in K$ . La quantité

$$\int_K f d\sigma \stackrel{\text{déf}}{=} \int_K f(\varphi(u, v)) \|(\varphi_u \wedge \varphi_v)(u, v)\| du dv$$

est l'intégrale de surface de  $f$  sur  $\Sigma$ .

Pour les nappes paramétrées, tout comme pour les courbes, on peut introduire la notion de paramétrisations équivalentes. On montre alors que l'intégrale de surface est indépendante de la paramétrisation choisie. En particulier (en considérant la fonction constante  $f = 1$ ), l'aire d'une surface est indépendante de la paramétrisation choisie.

## 7 Le théorème des fonctions implicites

### 7.1 Le cas des fonctions de deux variables

Une courbe dans le plan est souvent donnée à l'aide d'une équation de la forme

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

Le but de cette section est d'étudier les propriétés de telles courbes. À partir de l'équation  $F(x, y) = 0$  il est parfois possible d'exprimer une variable en fonction de l'autre, par exemple  $y = f(x)$ . Dans ce cas la courbe d'équation  $F(x, y) = 0$  est le graphe de la fonction  $f$ . Il peut être plus aisé d'exprimer  $x$  comme fonction de  $y$  : si l'on peut écrire  $x = g(y)$  cela signifie que la courbe est le graphe de la fonction  $g$ .

Mais il y a des courbes qui ne sont pas (globalement) de graphes : c'est le cas, par exemple, du cercle centré en  $O$  de rayon 1, qui correspond aux solutions de l'équation  $x^2 + y^2 - 1 = 0$ . Localement la situation est différente : si on fixe un point du cercle  $(x_0, y_0)$  il est possible de trouver un voisinage de ce point où le cercle est bien le graphe d'une fonction  $y = f(x)$  et/ou  $x = g(y)$  : il s'agit de prendre la fonction  $y = \sqrt{1 - x^2}$  (si  $y_0 > 0$ ) ou la fonction  $y = -\sqrt{1 - x^2}$  (si  $y_0 < 0$ ) ; ou encore  $x = \sqrt{1 - y^2}$  (si  $x_0 > 0$ ) ou  $x = -\sqrt{1 - y^2}$  (si  $x_0 < 0$ ).

Étant donnée une courbe d'équation  $F(x, y) = 0$ , et un point  $(x_0, y_0)$  de cette courbe, quand est-ce qu'on peut dire que la courbe est (au moins localement) le graphe d'une fonction  $y = f(x)$ ? La réponse est fournie par le théorème ci-dessous :

**Théorème 7.1** (des fonctions implicites pour les fonctions de 2 variables). *Soit  $U$  un ouvert de  $\mathbb{R}^2$  et  $F: U \rightarrow \mathbb{R}$  une fonction de classe  $C^1$ . Soit  $(x_0, y_0) \in U$  tel que*

$$F(x_0, y_0) = 0, \quad \frac{\partial F}{\partial y}(x_0, y_0) \neq 0.$$

*Alors il existe un voisinage  $I$  de  $x_0$ , un voisinage  $J$  de  $y_0$  et une fonction  $f: I \rightarrow J$  tels que*

$$\forall (x, y) \in I \times J: \quad F(x, y) = 0 \iff y = f(x).$$

*Dans ce cas la fonction  $x \mapsto f(x)$  est de classe  $C^1$  et*

$$f'(x) = -\frac{F_x(x, f(x))}{F_y(x, f(x))}, \quad (7.1)$$

où  $F_x = \frac{\partial F}{\partial x}$  et  $F_y = \frac{\partial F}{\partial y}$ .

Revenons au cas de l'équation  $x^2 + y^2 - 1 = 0$ . Les hypothèses sont satisfaites en tout point  $(x_0, y_0)$  du cercle, sauf aux points  $(-1, 0)$  et  $(1, 0)$  : au voisinage de ces deux points il n'est manifestement pas possible d'exprimer  $y$  comme fonction de  $x$ , mais il est possible d'exprimer  $x$  en fonction de  $y$ .

**Remarque 7.1.** En échangeant les rôles des variables  $x$  et  $y$  on trouve ceci : Si  $(x_0, y_0) \in U$  est un point tel que

$$F(x_0, y_0) = 0, \quad \frac{\partial F}{\partial x}(x_0, y_0) \neq 0.$$

*Alors il existe un voisinage  $I$  de  $x_0$  et un voisinage  $J$  de  $y_0$  et une fonction  $g: J \rightarrow I$  tels que*

$$\forall (x, y) \in I \times J: \quad F(x, y) = 0 \iff x = g(y).$$

*Dans ce cas la fonction  $y \mapsto g(y)$  est de classe  $C^1$  et  $g'(y) = -\frac{F_y(g(y), y)}{F_x(g(y), y)}$ .*

*Dém. du théorème.* Nous pouvons supposer, par exemple  $F_y(x_0, y_0) > 0$ . Par la continuité de  $F_y$ , on a  $F_y(x, y) > 0$  dans un rectangle  $I \times J$ . Posons  $J = [y_0 - h, y_0 + h]$ . La fonction  $y \mapsto F(x_0, y)$  étant croissante dans  $J$  on a  $F(x_0, y_0 - h) < 0$  et  $F(x_0, y_0 + h) > 0$ . Considérons les fonctions continues  $x \mapsto F(x, y_0 - h)$  et  $x \mapsto F(x, y_0 + h)$ . Quitte à remplacer  $I$  par un intervalle centré en  $x_0$  plus petit, pour  $x \in I$  on a :  $F(x, y_0 - h) < 0$  et  $F(x, y_0 + h) > 0$ . Par le théorème des valeurs intermédiaires, appliqué à la fonction croissante  $y \mapsto F(x, y)$  (avec  $x$  fixé), il existe un et un seul  $y = f(x)$  tel que  $F(x, f(x)) = 0$ . La fonction  $f$  est ainsi construite.

Démontrons que  $f$  est de classe  $C^1$  : Prenons  $x, x_1 \in I$ . En appliquant le théorème des accroissements finis à la fonction

$$\Phi(t) = F((1-t)x_1 + tx, (1-t)f(x_1) - tf(x)), \quad 0 \leq t \leq 1,$$

on trouve

$$\exists \tau \in [0, 1]: \quad \Phi(1) - \Phi(0) = \Phi'(\tau).$$

Considérons le point  $P_\tau = ((1-\tau)x_1 + \tau x, (1-\tau)f(x_1) - \tau f(x))$ , qui est un point sur le segment de  $(x, f(x))$  à  $(x_1, f(x_1))$ . L'égalité précédente se réécrit

$$F(x, f(x)) - F(x_1, f(x_1)) = \Phi'(\tau) = (x - x_1)F_x(P_\tau) + (f(x) - f(x_1))F_y(P_\tau).$$

D'autre part,  $F(x, f(x)) = F(x_1, f(x_1)) = 0$ . Donc,

$$\frac{f(x) - f(x_1)}{x - x_1} = -\frac{F_x(P_\tau)}{F_y(P_\tau)}.$$

Si l'on prend  $x_1 \rightarrow x$ , on a  $P_\tau \rightarrow (x, f(x))$ . Mais comme  $F$  est de classe  $C^1$ , on voit que  $f$  est dérivable et

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

L'expression à droite étant une fonction continue de  $x$ , on trouve que  $f$  est de classe  $C^1$ .  $\square$

La démonstration ci-dessus montre que si  $f$  est de classe  $C^k$ , alors  $f$  est de classe  $C^k$ .

**Corollaire 7.2.** *Si  $F$  est une fonction de classe  $C^1$  au voisinage de  $(x_0, y_0) \in \mathbb{R}^2$  et  $\nabla F(x_0, y_0) \neq (0, 0)$ , alors la courbe d'équation  $F(x, y) = 0$  possède une droite tangente au point  $(x_0, y_0)$  et  $\nu = \frac{\nabla F(x_0, y_0)}{\|\nabla F(x_0, y_0)\|}$  est le vecteur orthogonal la courbe d'équation  $F(x, y) = 0$ , au point  $(x_0, y_0)$ .*

*Dém.* En effet, au voisinage de  $(x_0, y_0)$ , la courbe est le graphe d'une fonction dérivable de la forme  $y = f(x)$ , ou éventuellement de la forme  $x = g(y)$ . Traitons, par exemple, le premier cas. Un point  $(x, y)$  appartient à la droite tangente si et seulement si  $y - y_0 = f'(x_0)(x - x_0)$ , c'est à dire si et seulement si (d'après (7.1))  $(x - x_0)F_x(x_0, y_0) + (y - y_0)F_y(x_0, y_0) = 0$ . Cela dit précisément que le vecteur  $\nabla F(x_0, y_0)$  est orthogonale à  $(x - x_0, y - y_0)$ , c'est à dire orthogonale à la droite tangente.  $\square$

**Corollaire 7.3.** *Si  $g: U \rightarrow \mathbb{R}$  est une fonction de classe  $C^1$ , où  $U$  est un ouvert de  $\mathbb{R}^2$ , alors en tout point  $(x_0, y_0) \in U$ ,  $\nabla g(x_0, y_0)$  est orthogonal à la ligne de niveau de  $g$  passant par  $(x_0, y_0)$ .*

*Dém.* Si  $\nabla g(x_0, y_0) = 0$  il n'y a rien à démontrer (le vecteur nul est orthogonal à tous les vecteurs). Si  $\nabla g(x_0, y_0) \neq 0$ , considérons la fonction  $F(x, y) = g(x, y) - c$  où  $c = g(x_0, y_0)$ . On a  $\nabla F(x_0, y_0) = \nabla g(x_0, y_0)$  et la ligne de niveau de  $g$  passant par  $(x_0, y_0)$  est l'ensemble d'équation  $F(x, y) = 0$ . Le résultat est alors une conséquence immédiate du corollaire précédent.  $\square$

## 7.2 Le cas des fonctions de 3 variables

Si  $F: U \rightarrow \mathbb{R}$  est une fonction de 3 variables de classe  $C^1$ , définie sur un ouvert  $U \subset \mathbb{R}^3$  et telle que, en un point  $(x_0, y_0, z_0) \in U$  on a

$$F(x_0, y_0, z_0) = 0, \quad F_z(x_0, y_0, z_0) \neq 0,$$

alors il est possible, localement, d'exprimer  $z$  comme fonction de  $x$  et de  $y$ . Autrement dit, il existe un voisinage  $V$  de  $(x_0, y_0)$ , un voisinage  $J$  de  $z_0$  et une fonction, de classe  $C^1$ ,  $f: V \rightarrow J$  tels que

$$\forall (x, y, z) \in V \times J: \quad F(x, y, z) = 0 \iff z = f(x, y)$$

et dans ce cas

$$f_x(x, y) = -\frac{F_x(x, y, f(x, y))}{F_z(x, y, f(x, y))}, \quad f_y(x, y) = -\frac{F_y(x, y, f(x, y))}{F_z(x, y, f(x, y))}.$$

La démonstration est semblable à celle du théorème précédent et omise ici.

Dans ce cas la surface représentative du graphe de la fonction  $z = f(x, y)$  possède, au point  $(x, y, f(x, y))$ , un vecteur normal donné par (voir l'exemple 6.4)

$$\nu = \frac{(-f_x, -f_y, 1)}{\sqrt{1 + \|\nabla f\|^2}},$$

où toutes les dérivées sont calculées au point  $(x, y)$ . Supposons maintenant, par exemple, que  $F_z(x_0, y_0, z_0) > 0$ . Compte tenu des formules précédentes pour  $f_x$  et  $f_y$ , on trouve

$$\nu = \left( \frac{F_x}{\|\nabla F\|}, \frac{F_y}{\|\nabla F\|}, \frac{F_z}{\|\nabla F\|} \right) = \frac{\nabla F}{\|\nabla F\|}. \quad (7.2)$$

**Exemple 7.2.** La fonction  $F(x, y, z) = x^2 e^z + z e^y + y^2$  s'annule à l'origine. De plus  $F_z(0, 0, 0) = 1$ . Par conséquent, l'équation  $F(x, y, z) = 0$  est, au voisinage de l'origine, le graphe d'une fonction  $z = f(x, y)$ . Cette fonction  $f$  est solution du système différentiel  $f_x(x, y) = -\frac{2x e^{f(x, y)}}{x^2 e^{f(x, y)} + e^y}$  et  $f_y(x, y) = -\frac{f(x, y) e^y + 2f(x, y)}{x^2 e^{f(x, y)} + e^y}$ . De plus, on a  $f(0, 0) = 0$  et  $f_x(0, 0) = 0$ ,  $f_y(0, 0) = 0$ .

Observons que la formule (7.2) implique que, pour une fonction de 3 variables, le vecteur gradient est orthogonal, en tout point, à la surface de niveau passant par ce point. Cette affirmation se démontre comme dans le cas des fonctions de 2 variables (voir le corollaire 7.3). En généralisant à  $n$  variables, il n'est pas difficile de se convaincre alors de la validité de la proposition suivante :

**Proposition 7.4.** Soit  $\Sigma$  un ensemble de niveau d'une fonction  $g$  de  $n$  variables de classe  $C^1$ , et  $x_0 \in \Sigma$  un point tel que  $g(x_0) = 0$ ,  $\nabla g(x_0) \neq 0$ . Alors le vecteur  $\nabla g(x_0)$  est orthogonal à la surface  $\Sigma$  au point  $x_0$ .

Traisons maintenant le cas de deux équations de la forme

$$\begin{cases} F(x, y, z) = 0 \\ G(x, y, z) = 0, \end{cases} \quad (7.3)$$

où les fonctions  $F$  et  $G$  sont de classe  $C^1$  au voisinage de  $P_0 = (x_0, y_0, z_0)$ . On se propose d'exprimer deux variables en fonction de la troisième, par exemple  $y = f(x)$  et  $z = g(x)$ . Pour cela, commençons par supposer

$$F(P_0) = G(P_0) = 0, \quad F_z(P_0) \neq 0.$$

En appliquant ce qui a été vu au début de cette section on peut écrire, au voisinage de  $P_0$ ,

$$F(x, y, z) = 0 \iff z = \varphi(x, y).$$

Considérons maintenant la fonction

$$\Phi(x, y) = G(x, y, \varphi(x, y)).$$

Au voisinage de  $P_0$  on a

$$\begin{cases} F(x, y, z) = 0 \\ G(x, y, z) = 0 \end{cases} \iff \begin{cases} z = \varphi(x, y) \\ G(x, y, \varphi(x, y)) = 0 \end{cases} \iff \begin{cases} z = \varphi(x, y) \\ \Phi(x, y) = 0 \end{cases}.$$

Observons que

$$\Phi_y = G_y + G_z \varphi_y = G_y - G_z \frac{F_y}{F_z}. \quad (7.4)$$

Maintenant, supposons que  $0 \neq \Phi_y(x_0, y_0)$ .

Dans ce cas nous pouvons exprimer  $y$  en fonction de  $x$  et écrire, au voisinage de  $(x_0, y_0)$ ,

$$\Phi(x, y) = 0 \iff y = f(x).$$

Cela donne alors, au voisinage de  $P_0$ ,

$$\begin{cases} F(x, y, z) = 0 \\ G(x, y, z) = 0 \end{cases} \iff y = f(x) \quad \text{et} \quad z = \varphi(x, f(x)) = g(x).$$

Pour arriver à cette conclusion nous avons eu besoin de supposer :

$$F(P_0) = G(P_0) = 0, \quad F_z(P_0) \neq 0 \quad \text{et} \quad \Phi_y(x_0, y_0) \neq 0. \quad (7.5)$$

Mais la formule précédemment donnée pour  $\Phi_y$  montre que

$$\Phi_y(x_0, y_0) \neq 0 \iff \det \frac{\partial(F, G)}{\partial(y, z)} \neq 0,$$

où

$$\frac{\partial(F, G)}{\partial(y, z)} = \begin{pmatrix} F_y & F_z \\ G_y & G_z \end{pmatrix}.$$

Observons par ailleurs que lorsque  $\det \frac{\partial(F, G)}{\partial(y, z)} \neq 0$ , on ne peut avoir simultanément  $F_z(P_0) = 0$  et  $G_z(P_0) = 0$ . Or, on a déjà vu que si  $F_z(P_0) \neq 0$  on peut exprimer les solutions du système (7.3) par les relations  $y = f(x)$  et  $z = g(x)$ . Si on revanche  $F_z(P_0) = 0$  on doit avoir  $G_z(P_0) \neq 0$  et en interchangeant les rôles de  $F$  et  $G$  on parvient à la même conclusion. Autrement dit, dans (7.5), la condition  $F_z(P_0) \neq 0$  est superflue.

Les discussions précédentes nous conduisent à énoncer le théorème suivant.

**Théorème 7.5** (Des fonctions implicites – 3 variables). *Soit  $F$  et  $G$  deux fonctions de classe  $C^1$  au voisinage d'un point  $P_0 = (x_0, y_0, z_0) \in \mathbb{R}^3$ , tel que  $F(P_0) = G(P_0) = 0$ . Si la matrice*

$$\frac{\partial(F, G)}{\partial(x, y, z)} = \begin{pmatrix} F_x(P_0) & F_y(P_0) & F_z(P_0) \\ G_x(P_0) & G_y(P_0) & G_z(P_0) \end{pmatrix}.$$

*est de rang 2, alors il existe un voisinage de  $P_0$  où le système (7.3) définit deux variables en fonction de la troisième. L'ensemble des solutions de (7.3) dans ce voisinage est alors le support d'une courbe régulière de  $\mathbb{R}^3$ .*\*

Rappelons que les lignes d'une matrice  $2 \times 3$  de rang 2 sont linéairement indépendantes. Sous les hypothèses du théorème précédent, on voit que les vecteurs  $\nabla F(P_0)$  et  $\nabla G(P_0)$  ne sont pas co-linéaires. Autrement dit,  $\nabla F(P_0) \wedge \nabla G(P_0) \neq 0$ . Mais le vecteur  $\nabla F(P_0)$  est orthogonal en  $P_0$  à la surface de niveau  $F(x, y, z) = 0$ . Et le vecteur  $\nabla G(P_0)$  est orthogonal en  $P_0$  à la surface de niveau  $G(x, y, z) = 0$ . L'hypothèse que la matrice est de rang 2 traduit alors le fait que ces deux surfaces n'ont pas le même plan tangent. La conclusion du théorème est que l'intersection des ces deux surfaces est bien une courbe.

### 7.3 Le théorème de l'inversibilité locale

Si  $U$  et  $V$  sont deux ouverts de  $\mathbb{R}^n$  et  $x_0 \in U$ . On dit que  $f$  est un difféomorphisme local en  $x_0$  s'il existe un voisinage ouvert  $W$  de  $x_0$  tel que  $f: W \rightarrow F(W)$  soit un  $C^1$ -difféomorphisme, c'est-à-dire une application bijective de classe  $C^1$  telle que l'application inverse est de classe  $C^1$ .

---

\*. Par exemple, la courbe paramétrée par  $x = x, y = f(x), z = g(x)$  si les deux dernière colonnes de la matrice sont linéairement indépendantes. Si le deux premières colonnes sont linéairement indépendantes on pourra expliciter  $x$  et  $y$  en fonction de  $z$ . Si la première et dernière colonne sont linéairement indépendantes, on écrira  $x$  et  $z$  comme fonction de  $y$ .

Pour qu'une fonction  $f$  soit un difféomorphisme local en  $x_0$  il est nécessaire que la matrice jacobienne en  $x_0$  soit inversible (c'est à dire de déterminant non nul, s'agissant d'une matrice carrée) En effet, on nous avons déjà vu que, en posant  $y_0 = f(x_0)$ ,

$$J_{f^{-1}}(y_0) = \left( J_f(x_0) \right)^{-1}.$$

Le théorème suivant affirme que cette condition est aussi suffisante.

**Exemple 7.3.** La fonction  $f: \mathbb{R} \rightarrow \mathbb{R}$  définie par  $f(x) = x^3$  est bijective et de classe  $C^1$ . Mais cette fonction n'est pas un difféomorphisme local en  $x = 0$ . En effet, la matrice jacobienne est simplement le singleton  $f'(0)$  et l'on a  $f'(0) = 0$ . Observons que l'application inverse est  $f^{-1}(y) = y^{1/3}$  et  $f^{-1}$  n'est pas de classe  $C^1$  au voisinage de  $y = 0$ .

**Théorème 7.6** (de l'inversibilité locale). *Soient  $U$  et  $V$  deux ouverts de  $\mathbb{R}^n$ ,  $x_0 \in U$  et  $f: U \rightarrow V$ . Si  $\det(J_f(x_0)) \neq 0$ , alors  $f$  est un difféomorphisme local en  $x_0$ .*

Donnons juste l'idée de la démonstration pour les fonctions de deux variables : La formule de Taylor au premier ordre affirme que  $f(x) - f(x_0)$  est approximativement égale à la différentielle  $df_{x_0}$  évalué en  $(x - x_0)$ , ou encore au produit de la matrice Jacobienne  $J_{x_0}(f)$  par le vecteur  $x - x_0$ . Ainsi, et en négligeant le reste dans la formule de Taylor,  $f(x) \simeq f(x_0) + df_{x_0}(x - x_0) = f(x_0) + J_f(x_0)(x - x_0)$ . Mais le terme à droite est inversible si et seulement si  $\det(J_f(x_0)) \neq 0$ . On s'attent alors que dans un voisinage suffisamment petit de  $x_0$  on puisse inverser  $f$  et que  $f^{-1}(y) \simeq y_0 + (J_f(x_0))^{-1}(y - y_0)$ .

La démonstration rigoureuse repose sur le théorème des fonctions implicites. Une démonstration alternative fait appel au théorème des contractions.

## 7.4 Extrema liés. Optimisation sous contrainte

Nous avons déjà étudié les problèmes d'optimisation de type

$$\min\{f(x): x \in U\} \quad \text{et} \quad \max\{f(x): x \in U\},$$

où l'ensemble  $U \subset \mathbb{R}^n$  était un ouvert.

Dans cette section on s'intéresse aux problèmes analogues d'optimisation sur des ensembles fermés, par exemple,

$$\min\{f(x): x \in \Sigma\}, \quad \max\{f(x): x \in \Sigma\},$$

où  $\Sigma$  est un fermé de  $\mathbb{R}^n$ .

Nous considérons le cas particulier important où  $\Sigma$  est une surface de niveau d'une fonction continue  $g: \mathbb{R}^n \rightarrow \mathbb{R}$  (ou une ligne de niveau si  $n = 2$ ). Sans perte de généralité on pourra alors supposer que

$$\Sigma = \{x \in \mathbb{R}^n: g(x) = 0\}.$$

Le calcul de

$$\min\{f(x): g(x) = 0\}, \quad \text{et} \quad \max\{f(x): g(x) = 0\}, \quad (\text{P})$$

sont deux "problèmes d'optimisation avec contrainte égalité". Pour simplifier la présentation nous supposons que  $g$  est une fonction  $C^1$ . Commençons par établir une propriété géométrique importante de l'ensemble  $\Sigma$ .

La condition  $\nabla g(x_0) \neq 0$  exprime qu'en  $x_0$  la contrainte n'est pas dégénérée. Voici un exemple de contrainte dégénéré :  $xy = 0$  est une contrainte dégénéré à l'origine (observer que cette contrainte n'est pas une courbe mais plutôt l'intersection de deux courbes) à l'origine.

S'il n'y a avait pas la contrainte  $g(x) = 0$ , les solutions de ces problèmes d'optimisation seraient à chercher parmi les points stationnaires de la fonction  $f$ . Dans le cas d'une optimisation avec contrainte, comme c'est le cas pour les problèmes (P), la situation est différente :

**Théorème 7.7** (des multiplicateurs de Lagrange). Soient  $f$  et  $g$  deux fonctions de classe  $C^1$  au voisinage de  $P_0 \in \mathbb{R}^n$ . Si  $P_0$  est un point tel que  $g(P_0) = 0$ ,  $\nabla g(P_0) \neq 0$  et  $P_0$  est un minimiseur ou un maximiseur pour  $f(x)$  sous la contrainte  $g(x) = 0$ , c'est à dire

$$f(P_0) = \min\{f(x) : g(x) = 0\}, \quad \text{ou} \quad f(P_0) = \max\{f(x) : g(x) = 0\},$$

alors  $\nabla f(P_0)$  est parallèle à  $\nabla g(P_0)$ . Autrement dit,

$$\exists \lambda_0 \in \mathbb{R} \text{ tel que } \nabla f(P_0) + \lambda_0 \nabla g(P_0) = 0.$$

On appelle le réel  $\lambda_0$  le multiplicateur de Lagrange associé au problème d'optimisation.

**Remarque 7.4.** Dans le cas  $\nabla f(P_0) \neq 0$ , et pour les fonctions de 2 variables, la conclusion s'interprète ainsi : les lignes de niveau de  $f$  (la fonction à optimiser) et de  $g$  (la fonction qui donne la contrainte) tangentes au point  $P_0$ . En effet,  $\nabla f(P_0)$  et  $\nabla g(P_0)$  sont, respectivement, orthogonaux à leurs lignes de niveau. De même, pour les fonctions de trois variables, les surfaces de niveau de  $f$  et de  $g$  sont tangentes en  $P_0$ .

*Dém.* Nous donnons la démonstration uniquement pour les fonctions de 2 ou 3 variables.

- Le cas des fonction de deux variables. Soit  $(x_0, y_0) \in \mathbb{R}^2$  un point où  $g(x_0, y_0) = 0$  et  $\nabla g(x_0, y_0) \neq 0$ . Supposons, par exemple  $\frac{\partial g}{\partial y}(x_0, y_0) \neq 0$  (si  $\frac{\partial g}{\partial y}(x_0, y_0) = 0$ , alors, par l'hypothèse sur  $g$ ,  $\frac{\partial g}{\partial x}(x_0, y_0) \neq 0$  et l'on raisonne en intervertissant les variables  $x$  et  $y$ ). Par le théorème des fonctions implicites, au voisinage de  $(x_0, y_0)$  on a  $g(x, y) = 0 \iff y = \phi(x)$ , où  $\phi: I \rightarrow J$  et  $I$  et  $J$  sont des intervalles de  $\mathbb{R}$ , voisinages de  $x_0$  et  $y_0$  respectivement. En particulier,  $\phi(x_0) = y_0$ . Si  $f(x_0, y_0) = \min\{f(x, y) : g(x, y) = 0\}$ , ou si  $f(x_0, y_0) = \max\{f(x, y) : g(x, y) = 0\}$ , alors la fonction  $\Psi(x) = f(x, \phi(x))$  possède respectivement un minimum ou un maximum local en  $x_0$ . Dans les deux cas,  $\Psi'(x_0) = 0$  par le principe de Fermat. Mais alors,

$$0 = \Psi'(x_0) = f_x(x_0, \phi(x_0)) + f_y(x_0, \phi(x_0))\phi'(x_0) = \nabla f(x_0, y_0) \cdot (1, \phi'(x_0)).$$

Le vecteur  $\nabla f(x_0, y_0)$  est donc orthogonal au vecteur  $(1, \phi'(x_0))$ . Mais le vecteur  $(1, \phi'(x_0))$  est tangent en  $(x_0, y_0)$  au graphe de la fonction  $\phi$ , c'est-à-dire à la courbe  $\{(x, y) : y = \phi(x)\} = \{(x, y) : g(x, y) = 0\}$ . Ainsi,  $\nabla f(x_0, y_0)$  est orthogonal à la courbe  $\{(x, y) : g(x, y) = 0\}$ . Mais  $\nabla g(x_0, y_0)$  est lui même orthogonale à cette courbe par la proposition 7.4. Ainsi,  $\nabla f(x_0, y_0)$  et  $\nabla g(x_0, y_0)$  sont deux vecteurs parallèles.

- Le cas des fonctions de 3 variables. Soit  $P_0 = (x_0, y_0, z_0)$ . L'hypothèse sur  $g$  est  $\nabla g(P_0) \neq 0$ . Supposons, par exemple,  $g_z(P_0) \neq 0$ . On a  $g(x, y, z) = 0$  au voisinage de  $P_0$  si et seulement si  $z = \phi(x, y)$  dans ce voisinage, où  $\phi$  est une fonction de classe  $C^1$ . La fonction  $\Psi(x, y) := f(x, y, \phi(x, y))$  possède alors un extremum libre en  $(x_0, y_0)$ . Mais  $(x_0, y_0)$  est alors un point stationnaire pour  $\Psi$  et  $\nabla \Psi(x_0, y_0) = 0$ . En appliquant à  $\phi$  des formules analogues (7.4) on trouve

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \nabla \Psi(x_0, y_0) = \begin{pmatrix} (f_x - f_z g_x / g_z)(P_0) \\ (f_y - f_z g_y / g_z)(P_0) \end{pmatrix} = \frac{1}{g_z(P_0)} \begin{pmatrix} (f_x g_z - f_z g_x)(P_0) \\ (f_y g_z - f_z g_y)(P_0) \end{pmatrix}.$$

D'après la formule du produit vectoriel (6.1), cela implique que la première et la deuxième composante du vecteur  $(\nabla f \wedge \nabla g)(P_0)$  s'annulent. Maintenant, si  $g_x(P_0) = 0$  et  $g_y(P_0) = 0$  alors la troisième composante de  $(\nabla f \wedge \nabla g)(P_0)$  sera nulle également. Sinon,  $g_x(P_0) \neq 0$  ou  $g_y(P_0) \neq 0$  et on parvient à la même conclusion en raisonnant comme ci-dessus mais en échangeant les rôles des variables  $x$ ,  $y$  et  $z$ . En conclusion  $\nabla f \wedge \nabla g$  s'annule au point  $P_0$ , ce qui se produit précisément lorsque  $\nabla f(P_0)$  et  $\nabla g(P_0)$  sont parallèles. □

Le théorème précédent admet la reformulation suivante : introduisons la Lagrangienne du problème d'optimisation, qui est la fonction de  $n + 1$  variables

$$\mathcal{L}(x, \lambda) = f(x) + \lambda g(x).$$

On a

$$\begin{cases} \nabla f(P_0) + \lambda_0 \nabla g(P_0) = 0 \\ g(P_0) = 0 \end{cases} \iff \nabla \mathcal{L}(x_0, \lambda_0) = 0,$$

où  $\nabla_{x,\lambda} \mathcal{L}$  est le vecteur de  $n + 1$  composantes  $\frac{\partial \mathcal{L}}{\partial x_1}, \dots, \frac{\partial \mathcal{L}}{\partial x_n}$  et  $\frac{\partial \mathcal{L}}{\partial \lambda}$ .

Autrement dit, les solutions  $P_0 \in \mathbb{R}^n$  du problème d'optimisation sous contrainte

$$\min\{f(x) : g(x) = 0\}, \quad \text{et} \quad \min\{f(x) : g(x) = 0\}, \quad (\text{P})$$

où  $f$  et  $g$  sont de classe  $C^1$  au voisinage de  $P_0$  et à valeur dans  $\mathbb{R}$ , sont à chercher parmi les points  $P_0$  tels qu'il existe  $\lambda_0 \in \mathbb{R}$  tel que  $(P_0, \lambda_0) \in \mathbb{R}^{n+1}$  est un point stationnaire de la Lagrangienne (vue comme fonction de  $n + 1$ -variables). Les solutions du problème (P) sont aussi à chercher parmi les points où la contrainte  $g(x) = 0$  est dégénérée. En conclusion, il s'agit alors de chercher les solutions de

$$\nabla_{x,\lambda} \mathcal{L}(x, \lambda) = 0, \quad \text{ou} \quad \begin{cases} \nabla g(x) = 0 \\ g(x) = 0. \end{cases}$$

Observer que ce théorème ne fournit qu'une condition nécessaire : ce théorème s'avère très utile pour trouver les points  $P_0$  qui sont les bons candidats à être les solutions des problèmes d'optimisation avec contrainte. Mais sous les hypothèses du théorème, il peut arriver que  $(P_0, \lambda_0)$  soit un point stationnaire de  $\mathcal{L}$ , ou que  $g(P_0) = 0$  et  $\nabla g(P_0) = 0$ , sans que  $P_0$  soit ni un minimum, ni un maximum du problème d'optimisation.

Dispose-t-on de conditions suffisantes pour l'existence d'extrema avec contraintes ? La réponse est affirmative, mais n'insistons pas sur ce point. Bien souvent, on applique le théorème de Weierstrass pour démontrer que le minimum et/ou le maximum existent. Cela est possible notamment si la contrainte  $\Sigma$  définit un ensemble compact. \*

**Exemple 7.5.** Trouvons le maximum et le minimum de la fonction  $f(x, y, z) = x + y - \sqrt{6}z$  sur la sphère  $\Sigma$  de centre  $O$  et de rayon 1. Tout d'abord,  $f$  est une fonction continue et  $\Sigma = \{(x, y, z) : x^2 + y^2 + z^2 - 1 = 0\}$  est un compact de  $\mathbb{R}^3$ . Donc, par le théorème de Weierstrass, les problèmes de minimisation et de maximisation posés possèdent bien des solutions. Ici,  $g(x, y, z) = x^2 + y^2 + z^2 - 1$ . Observons que la contrainte  $g(x, y, z) = 0$  n'est jamais dégénérée, puisque le seul

\*. Rappelons aussi la variante suivante du théorème de Weierstrass, qui s'applique (parfois) quand  $\Sigma$  n'est pas compact :

**Théorème 7.8** (Weierstrass - variante). *Soit  $f : \Sigma \subset \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction continue. Si  $\bar{x} \in \Sigma$  et si l'ensemble de sous-niveau  $K = \{x \in \Sigma : f(x) \leq f(\bar{x})\}$  est compact, alors le problème de minimisation  $\min_{x \in \Sigma} f(x)$  possède une solution. Autrement dit, il existe  $x^* \in D$  tel que  $f(x^*) = \min_{x \in \Sigma} f(x)$ .*

Le cas typique d'application est celui d'une fonction  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  telle que  $\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty$ . Une telle fonction a tous les ensembles de sous-niveau bornés (pourquoi?). Si de plus  $f$  est continue, ses ensembles de sous-niveau sont fermés (pourquoi?) et donc compacts. La fonction possède alors un minimum absolu.

Bien entendu on peut établir un théorème analogue pour l'existence d'un maximum absolu : il s'agit cette fois-ci de supposer que l'ensemble de "sur-niveau"  $\{x : f(x) \geq f(\bar{x})\}$  est borné. Le cas typique d'application est celui d'une fonction continue telle que  $\lim_{\|x\| \rightarrow +\infty} f(x) = -\infty$ .



point où  $\nabla g$  s'annule est l'origine, mais l'origine ne vérifie pas la contrainte. Ainsi, les points de minimum et maximum sont à chercher parmi les points stationnaires de la Lagrangienne. La Lagrangienne du système est la fonction  $\mathcal{L}(x, y, z, \lambda) = x + y - \sqrt{6}z + \lambda(x^2 + y^2 + z^2 - 1)$ . Ses points stationnaires sont les solutions du système

$$\begin{cases} 1 + 2\lambda x = 0 \\ 1 + 2\lambda y = 0 \\ -\sqrt{6} + 2\lambda z = 0 \\ x^2 + y^2 + z^2 - 1 = 0. \end{cases}$$

En exprimant  $x, y, z$  en fonction de  $\lambda$ , la quatrième équation donne  $\lambda = \pm\sqrt{2}$  et ensuite  $(x_0, y_0, z_0, \lambda_0) = \pm(1/2\sqrt{2}, 1/2\sqrt{2}, -\sqrt{6}/2\sqrt{2}, \sqrt{2})$ . Les solutions des problèmes de minimisations et maximisation (dont l'existence a été établie avant via le théorème de Weierstrass) sont alors à chercher parmi les points  $P = (1/2\sqrt{2}, 1/2\sqrt{2}, -\sqrt{6}/2\sqrt{2})$  et  $Q = (-1/2\sqrt{2}, -1/2\sqrt{2}, +\sqrt{6}/2\sqrt{2})$ . Mais un calcul direct montre que  $f(P) > f(Q)$ . Ceci permet de dire que  $f$  atteint sur la sphère son maximum en  $P$  et son minimum en  $Q$ .

Même si le théorème des multiplicateurs de Lagrange est un outil puissant, il convient parfois ne pas l'utiliser, notamment s'il est possible d'éliminer la contrainte en exprimant une variable en fonction des autres :

**Exemple 7.6.** Optimiser  $x^2 + y^2 + z^2$  sous la contrainte  $x + 2y + 3z = 1$ . Interpréter le résultat. Solution. La lagrangienne du système est  $\mathcal{L}(x, y, z, \lambda) = x^2 + y^2 + z^2 + \lambda(x + 2y + 3z - 1)$ . Son unique point stationnaire, obtenu avec le multiplicateur  $\lambda = -1/7$ , est  $(1/14, 1/7, 3/14)$ . La contrainte n'étant pas compacte le théorème de Weierstrass ne s'applique pas. Appliquons la variante du théorème de Weierstrass. Choisissons un point arbitraire sur la contrainte, par exemple  $(0, 0, 1/3)$ . L'ensemble de sous-niveau  $\{(x, y, z) : x^2 + y^2 + z^2 \leq 1/9\}$  est manifestement compact (c'est la boule fermée de rayon  $1/3$  et centre  $1/3$ ). Le problème posé possède alors un minimum absolu. Le point  $(1/14, 1/7, 3/14)$  est donc l'optimum demandé et il s'agit d'un minimum. Il s'agit du point du plan d'équation  $x + 2y + 3z = 1$  à distance minimale de l'origine.

**Exemple 7.7.** On cherche à construire un caisson de  $20\text{m}^3$ . Le matériau pour le fond coûte 3 euros/ $\text{m}^2$ , pour le couvercle 2 euros/ $\text{m}^2$  et pour les côtés 1 euro/ $\text{m}^2$ . Quel est le caisson le moins cher ? Et le plus cher ?

Réponse :

- *Modélisation* : Notons  $x$ (=longueur),  $y$ (=largeur),  $z$ (=hauteur) les mesures du caisson en mètres. Le coût de construction est  $C(x, y, z) = 3xy + 2xy + 2(xz + yz) = 5xy + 2xz + 2yz$ . Il s'agit de résoudre les problèmes de minimisation et maximisation pour  $C$  "avec contrainte" :  $\min\{C(x, y, z) : xyz = 20\}$  et  $\max\{C(x, y, z) : xyz = 20\}$ .
- *Élimination de la contrainte et recherche des points stationnaires* : On pourrait introduire la Lagrangienne et en déterminer les points stationnaires (cela conduit à résoudre un système de 4 équations et 4 inconnues). Mais il est plus aisé de poser  $z = 20/(xy)$  et d'étudier la fonction

$$f(x, y) = C(x, y, \frac{20}{xy}) = 5xy + 40(\frac{1}{y} + \frac{1}{x}), \quad x > 0, y > 0.$$

La fonction  $f$  étant définie sur un ouvert, il s'agit d'en trouver les points stationnaires : Cela conduit à résoudre le système (de deux équations et 2 inconnues)  $\nabla f(x, y) = 0$ . Ce système possède une seule solution pour  $x > 0$  et  $y > 0$ . Elle est donnée par  $x = y = 2$  (et donc  $z = 5$ ).

- *Synthèse* : On construit donc un caisson de mesures  $2 \times 2 \times 5$ . Ce choix correspond-t-il au caisson de coût minimum, maximum, ou ni l'un ni l'autre ? Le problème de minimisation est-il *bien posé* ? Et celui de maximisation ? Pour répondre à ces questions appliquons la variante du théorème de Weierstrass avec  $(\bar{x}, \bar{y}) = (1, 1)$  (ce choix est arbitraire). Observons que l'ensemble de sous-niveau  $K = \{(x, y) : f(x, y) \leq f(1, 1)\} = 5xy + 40(\frac{1}{y} + \frac{1}{x}) \leq 85\}$  est compact (en effet, il est manifestement fermé et il est borné, puisque  $x \geq 40/85$ ,  $y \geq 40/85$  et  $5xy \leq 85 \Rightarrow x \leq 17 \cdot 85/40$  et  $y \leq 17 \cdot 85/40$ ). Mais alors le problème de minimisation de l'exemple 7.7 est bien posé, c'est-à-dire que le caisson de coût minimum existe : c'est bien le caisson de mesures  $2 \times 2 \times 5$  trouvé avant. Le problème de maximisation est mal posé : le caisson de coût maximum n'existe pas. On le voit en observant que des caissons de mesures  $x, x, 20/x^2$  ont un coût qui tend à l'infini si  $x \rightarrow +\infty$ .

**Optimisation sous plusieurs contraintes.** Le théorème des multiplicateurs de Lagrange se généralise au cas où il y a plusieurs contraintes à satisfaire. Considérons par exemple le problème

$$\min\{f(x) : g_1(x) = 0, g_2(x) = 0\}, \quad \text{ou} \quad \max\{f(x) : g_1(x) = 0, g_2(x) = 0\} \quad (x \in \mathbb{R}^n, n \geq 3).$$

où  $f$ ,  $g_1$  et  $g_2$  sont des fonctions de classe  $C^1$ . On introduit dans ce cas la Lagrangienne de  $(n + 2)$ -variables  $\mathcal{L}(x, \lambda_1, \lambda_2)$ , où

$$\mathcal{L}(x, \lambda_1, \lambda_2) = f(x) + \lambda_1 g_1(x) + \lambda_2 g_2(x).$$

On peut démontrer que les points  $x_0$  de minimum ou maximum de  $f$  sous les contraintes  $g_1(x) = g_2(x) = 0$  sont à chercher parmi les points  $x \in \mathbb{R}^n$  tels que  $(x, \lambda_1, \lambda_2)$  est un point stationnaires de la Lagrangienne. Ainsi, pour trouver ces points (ou du moins des points candidats à être des solutions du problème d'optimisation), on commence par trouver les solutions du système de  $n + 2$  équations

$$\nabla_{x, \lambda_1, \lambda_2} \mathcal{L}(x, \lambda_1, \lambda_2) = 0. \tag{7.6a}$$

Mais comme on l'a vu avant, un autre cas de figure est possible : celui où la contrainte  $g_1(x) = g_2(x) = 0$  est dégénérée : ainsi les solutions du problèmes d'optimisation sont aussi à chercher parmi les éventuelles solutions des systèmes

$$\begin{cases} \nabla g_1(x) \text{ et } \nabla g_2(x) \text{ sont colinéaires} \\ g_1(x) = g_2(x) = 0 \end{cases} \tag{7.6b}$$

Dans la plupart des cas la contrainte ne sera pas dégénérée et les systèmes (7.6b) n'ont pas de solution.

Bien entendu, ces considérations se généralisent à un nombre arbitraire de contraintes. Dans ce cas, la contrainte est dégénéré lorsque les gradients de  $g_1, \dots, g_n$  ne forment pas un système libre.