
Fonctions de classe $\mathcal{C}^k$ de $\mathbb{R}^n$ dans $\mathbb{R}$.
Points critiques et courbes de niveau. Notes.

Table des matières

1	Prérequis	2
2	Fonctions de classe $\mathcal{C}^k$	2
2.1	Différentielles d'ordre supérieur	2
2.2	Théorème de Schwarz	3
2.3	Matrice hessienne	5
2.4	Développements limités	5
2.5	Convexité	7
3	Optimisation	7
3.1	Existence d'un minimum	7
3.2	Points critiques	8
3.3	Analyse locale des points critiques	9
3.4	Un exemple	10
3.5	Domaines compacts	10
4	Un peu de topographie	11
4.1	Courbes de niveau	11
4.2	Théorème des fonctions implicites	11
4.2.1	Le cercle et la sphère	11
4.2.2	Le cas général	12
4.2.3	Application : racines de polynômes	12
4.3	Tangentes aux courbes de niveau	13
4.3.1	Cas général	13
4.4	Lemme de Morse	14
5	Références	15

1 Prérequis

Connaissances préalables :

- ▷ Équivalence des normes en dimension finie.
- ▷ Différentiabilité de fonctions de plusieurs variables réelles.
- ▷ Intégrales et primitives.

Dans ce qui suit, on se donne :

- ▷ $n \geq 1$ un entier.
- ▷ $U \subset \mathbb{R}^n$ un ouvert non vide.
- ▷ $(\vec{e}_1, \dots, \vec{e}_n)$ la base canonique de $\mathbb{R}^n$.
- ▷ $f : U \rightarrow \mathbb{R}$.

Exceptionnellement, comme en géométrie affine, nous utiliserons la notation $\vec{v}$ afin de mieux distinguer les *points* et les *vecteurs tangents*. Cette notion n'est pas à connaître, et en particulier les flèches peuvent être ignorées.

2 Fonctions de classe $\mathcal{C}^k$

2.1 Différentielles d'ordre supérieur

Soient $f : \mathbb{R}^p \rightarrow \mathbb{R}^q$ et $x \in \mathbb{R}^p$. La différentielle $D_x f$ de f en x est une application linéaire de $\mathbb{R}^p$ dans $\mathbb{R}^q$. En particulier :

- ▷ Si $p = 1$ (fonction d'une variable réelle), alors la donnée de $D_x f : \mathbb{R} \rightarrow \mathbb{R}^q$ est équivalente à la donnée du vecteur $f'(x) = D_x f(1) \in \mathbb{R}^q$. Par conséquent, on peut voir la fonction dérivée de f comme une fonction de $\mathbb{R}$ dans $\mathbb{R}^q$. En itérant ce procédé, les dérivées de f d'ordre supérieur sont des toutes des fonctions de $\mathbb{R}$ dans $\mathbb{R}^q$.
- ▷ Sinon, la fonction $x \mapsto D_x f$ est une fonction de $\mathbb{R}^p$ dans l'espace vectoriel $L(\mathbb{R}^p, \mathbb{R}^q)$. Dans la base canonique de $L(\mathbb{R}^p, \mathbb{R}^q)$, ses coefficients sont données par les dérivées partielles d'ordre 1 (matrice jacobienne). La fonction $x \mapsto D_x f$ est donc continûment différentiable si et seulement si les dérivées partielles d'ordre 1 sont continues, différentiable, et de dérivées partielles continues.

Nous allons ici nous consacrer aux fonctions réelles de plusieurs variables réelles, c'est-à-dire aux fonction de U dans $\mathbb{R}$, où U est un ouvert de $\mathbb{R}^n$.

Commençons par donner un sens à la dérivée seconde de f . Soit $f \in \mathcal{C}^2(U, \mathbb{R})$. Alors, pour tout x , la différentielle $D_x f : \mathbb{R}^n \rightarrow \mathbb{R}$ est une forme linéaire sur $\mathbb{R}^n$. La différentielle de Df est alors une application linéaire de $\mathbb{R}^n$ dans $(\mathbb{R}^n)^*$, ce qui peut être délicat à représenter matriciellement. Nous avons plusieurs choix possibles :

- ▷ Le gradient de f est un vecteur, donc $x \mapsto \nabla_x f$ est une fonction de $\mathbb{R}^n$ dans $\mathbb{R}^n$. Elle est aussi différentiable, sa différentielle est une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^n$, dont la matrice jacobienne est $\left(\frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} \right) \right)_{1 \leq i, j \leq n}$.
- ▷ On écrit la matrice de $D_x f$ dans la base canonique¹ de $(\mathbb{R}^n)^*$. Cette matrice est le vecteur colonne $\left(\frac{\partial f}{\partial x_i} \right)_{1 \leq i \leq n}$, qui est égal aux coordonnées du gradient dans la base canonique. Comme avec le point de vue précédent, il reste à dériver la fonction de $\mathbb{R}^n$ dans $\mathbb{R}^n$ obtenu, dont la matrice jacobienne est $\left(\frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} \right) \right)_{1 \leq i, j \leq n}$. Ces deux points de vue coïncident car la base canonique est orthonormée pour le produit scalaire canonique.
- ▷ Plus conceptuellement : la fonction dérivée Df est à image dans $(\mathbb{R}^n)^*$, que l'on notera $\mathbb{R}^n \rightarrow \mathbb{R}$. La différentielle seconde $D_x^2 f$ est donc une application linéaire $\mathbb{R}^n \rightarrow (\mathbb{R}^n)^*$, donc un élément

1. Plus précisément, la base duale de la base canonique de $\mathbb{R}^n$.

de $\mathbb{R}^n \rightarrow (\mathbb{R}^n \rightarrow \mathbb{R})$. Autrement dit, si l'on se donne deux vecteurs $\vec{u}, \vec{v}$, alors $D^2f(\vec{u}, \vec{v})$ est un réel. De plus, D_x^2f est linéaire en $\vec{u}$ comme en $\vec{v}$, donc est une **forme bilinéaire** sur $\mathbb{R}^n$. Cette forme bilinéaire a une représentation matricielle dans la base canonique, qui est $(D^2f(\vec{e}_i, \vec{e}_j))_{1 \leq i, j \leq n} = \left(\frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_j} \right) \right)_{1 \leq i, j \leq n}$.

Remarque 2.1.

Le troisième point de vue a l'avantage de se généraliser aisément. Par exemple, la différentielle troisième en un point est une application linéaire $D_x^3f : \mathbb{R}^n \rightarrow (\mathbb{R}^n \rightarrow (\mathbb{R}^n \rightarrow \mathbb{R}))$, c'est-à-dire une forme trilinéaire. Plus généralement, la différentielle k -ième est une forme k -linéaire. Pour $k = 1$, on peut la représenter par un vecteur ligne de longueur n ; pour $k = 2$, par une matrice carrée $n \times n$; pour $k = 3$, une représentation naturelle utiliserait un cube $n \times n \times n$. À cause des limitations de l'écriture en deux dimensions, on n'utilise pas de telle représentation.

Si $f : \mathbb{R}^p \rightarrow \mathbb{R}^q$ est à valeurs dans $\mathbb{R}^q$, alors la donnée de f est équivalente à la donnée de q fonctions de $\mathbb{R}^p$ dans $\mathbb{R}$. Sa différentielle k -ième est alors une application k -linéaire de $\mathbb{R}^p$ à valeurs dans $\mathbb{R}^q$. On parle aussi de **tenseur**².

2.2 Théorème de Schwarz

Lors que nous avons calculé la matrice de la dérivée seconde de f , nous avons obtenu deux expressions différentes selon notre point de vue; ou bien $\left(\frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} \right) \right)_{1 \leq i, j \leq n}$, ou bien $\left(\frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_j} \right) \right)_{1 \leq i, j \leq n}$. Heureusement, ces deux expressions sont identiques pour des fonctions de classe $\mathcal{C}^2$.

Théorème 1 (Théorème de Schwarz).

Soit $f \in \mathcal{C}^2(U, \mathbb{R})$. Pour tous $1 \leq i, j \leq n$,

$$\frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_j} \right) = \frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} \right).$$

Démonstration.

Quitte à permuter l'ordre des variables, on peut supposer sans perte de généralité que $i = 1$ et $j = 2$. De plus, les variables x_3 à x_n étant alors fixées, on peut supposer de plus que f est une fonction de $\mathbb{R}^2$ dans $\mathbb{R}$.

Soit $(x, y) \in U$. Soient $h, k \neq 0$ suffisamment petits. On calcule $f(x + h, y + k) - f(x, y)$ de deux façons différentes :

- ▷ En intégrant sa dérivée le long du chemin $(x, y) \rightarrow (x + h, y) \rightarrow (x + h, y + k)$;
- ▷ En intégrant sa dérivée le long du chemin $(x, y) \rightarrow (x, y + k) \rightarrow (x + h, y + k)$.

Le premier calcul donne :

$$\begin{aligned} f(x + h, y + k) - f(x, y) &= [f(x + h, y) - f(x, y)] + [f(x + h, y + k) - f(x + h, y)] \\ &= \int_0^h \frac{\partial f}{\partial x}(x + s, y) \, ds + \int_0^k \frac{\partial f}{\partial y}(x + h, y + t) \, dt. \end{aligned}$$

Le second calcul donne :

$$f(x + h, y + k) - f(x, y) = \int_0^k \frac{\partial f}{\partial y}(x, y + t) \, dt + \int_0^h \frac{\partial f}{\partial x}(x + s, y + k) \, ds.$$

2. Ou, du moins, d'une forme spécifique de tenseur; plus généralement, les tenseurs peuvent aussi admettre des variables qui sont des formes linéaires, et dans un contexte de géométrie différentielle, obéissent à des règles précises de covariance.

Les valeurs obtenues étant identiques,

$$\begin{aligned} \int_0^h \frac{\partial f}{\partial x}(x+s, y) \, ds + \int_0^k \frac{\partial f}{\partial y}(x+h, y+t) \, dt &= \int_0^k \frac{\partial f}{\partial y}(x, y+t) \, dt + \int_0^h \frac{\partial f}{\partial x}(x+s, y+k) \, ds \\ \int_0^h \left[\frac{\partial f}{\partial x}(x+s, y+k) - \frac{\partial f}{\partial x}(x+s, y) \right] \, ds &= \int_0^k \left[\frac{\partial f}{\partial y}(x+h, y+t) - \frac{\partial f}{\partial y}(x, y+t) \right] \, dt \\ \int_0^h \int_0^k \frac{\partial^2 f}{\partial y \partial x}(x+s, y+t) \, dt \, ds &= \int_0^k \int_0^h \frac{\partial^2 f}{\partial x \partial y}(x+s, y+t) \, ds \, dt. \end{aligned}$$

Soit $\varepsilon > 0$. La fonction f étant supposée de classe $\mathcal{C}^2$, ses dérivées partielles d'ordre 2 sont continues. Soit $\delta > 0$ tel que $\left| \frac{\partial^2 f}{\partial x \partial y}(x+s, y+t) - \frac{\partial^2 f}{\partial x \partial y}(x, y) \right| \leq \varepsilon$ dès que $\|(s, t)\| \leq \delta$. Supposons de plus que $\|(h, k)\| \leq \delta$. Alors

$$\left| \frac{1}{hk} \int_0^h \int_0^k \frac{\partial^2 f}{\partial y \partial x}(x+s, y+t) \, dt \, ds - \frac{\partial^2 f}{\partial y \partial x}(x, y) \right| \leq \varepsilon,$$

et de même pour le membre de droite. Par inégalité triangulaire,

$$\left| \frac{\partial^2 f}{\partial y \partial x}(x, y) - \frac{\partial^2 f}{\partial x \partial y}(x, y) \right| \leq 2\varepsilon.$$

Ceci étant vrai pour tout $\varepsilon > 0$, on en déduit le théorème de Schwarz. $\square$

Remarque 2.2.

Le plus souvent, les démonstrations du théorème de Schwarz utilise le théorème des accroissements finis au lieu d'intégrales.

Remarque 2.3.

L'argument en ε - δ peut être évité à l'aide d'une utilisation astucieuse du théorème de convergence dominée. La première partie du calcul donne :

$$\begin{aligned} (hk)^{-1} \int_0^h \int_0^k \frac{\partial^2 f}{\partial y \partial x}(x+s, y+t) \, dt \, ds &= (hk)^{-1} \int_0^k \int_0^h \frac{\partial^2 f}{\partial x \partial y}(x+s, y+t) \, ds \, dt \\ \int_0^1 \int_0^1 \frac{\partial^2 f}{\partial y \partial x}(x+hs, y+kt) \, dt \, ds &= \int_0^1 \int_0^1 \frac{\partial^2 f}{\partial x \partial y}(x+hs, y+kt) \, ds \, dt. \end{aligned}$$

La famille de fonctions $(s, t) \mapsto \frac{\partial^2 f}{\partial y \partial x}(x+hs, y+kt)$, paramétrée par (h, k) , est une famille de fonction continues uniformément bornée si $|h|, |k| \leq 1$. De plus, elle converge simplement vers la fonction constante égale à $\frac{\partial^2 f}{\partial y \partial x}(x, y)$ quand (h, k) tend vers $(0, 0)$. Par le théorème de convergence dominée,

$$\lim_{(h,k) \rightarrow (0,0)} \int_0^1 \int_0^1 \frac{\partial^2 f}{\partial y \partial x}(x+hs, y+kt) \, dt \, ds = \int_0^1 \int_0^1 \frac{\partial^2 f}{\partial y \partial x}(x, y) \, dt \, ds = \frac{\partial^2 f}{\partial y \partial x}(x, y),$$

et de même pour le membre de droite.

Autrement dit, $\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial^2 f}{\partial x_i \partial x_j}$. L'ordre dans lequel on calcule les dérivées partielles n'a aucune importance, et par récurrence, cela reste vrai pour les dérivées d'ordre supérieur :

$$\frac{\partial^3 f}{\partial x \partial y \partial x} = \frac{\partial^3 f}{\partial x \partial x \partial y} = \frac{\partial^3 f}{\partial x^2 \partial y}.$$

On écrira donc en général les dérivées partielles dans le même ordre que les variables, ce qui donne lieu à des expressions de la forme

$$\frac{\partial^k f}{\partial x_1^{k_1} \dots \partial x_n^{k_n}}, \quad \text{avec} \quad k = k_1 + \dots + k_n.$$

2.3 Matrice hessienne

Corollaire 2.4.

Soit $f \in \mathcal{C}^2(U, \mathbb{R})$. Alors, pour tout $x \in U$, la forme bilinéaire $D_x^2 f$ est symétrique.

Définition 2.5.

On note $\text{Hess}(f) = \left(\frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} \right) \right)_{1 \leq i, j \leq n}$ sa représentation matricielle, elle aussi symétrique, appelée **matrice hessienne**.

La donnée d'une forme bilinéaire symétrique est équivalente à la donnée d'une forme quadratique. Dans la suite, je noterai $\mathcal{Q}_x$ la forme quadratique associée à $D_x^2 f$, c'est-à-dire la forme quadratique de matrice $\text{Hess}(f)$, ou encore définie par $\mathcal{Q}_x(\vec{u}) = D_x^2 f(\vec{u}, \vec{u})$.

Plus généralement, si f est de classe $\mathcal{C}^k$, alors $D_x^k f$ est une forme k -linéaire symétrique, au sens où sa valeur ne dépend pas de l'ordre des variables (invariance par permutations).

Exemple 2.6.

Soit $\gamma : \mathbb{R} \rightarrow U$ une courbe paramétrée. Calculons les dérivées de $f \circ \gamma$. La dérivée première est donnée par

$$(f \circ \gamma)'(t) = D_{\gamma(t)} f(\gamma'(t)) = \langle \nabla_{\gamma(t)} f, \gamma'(t) \rangle = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\gamma(t)) \cdot \gamma'_i(t).$$

En dérivant cette expression, on obtient

$$(f \circ \gamma)''(t) = D_{\gamma(t)}^2 f(\gamma'(t), \gamma'(t)) + D_{\gamma(t)} f(\gamma''(t)). \quad (1)$$

En particulier, si $\gamma(t)$ paramètre un segment à vitesse constante (i.e. $\gamma(t) = (1-t)p + tq$ pour deux points p, q), alors $\gamma'' = 0$, et donc

$$(f \circ \gamma)''(t) = D_{\gamma(t)}^2 f(\gamma'(t), \gamma'(t)) = \mathcal{Q}_{\gamma(t)}(\gamma'(t)).$$

Exemple 2.7.

Soit $f : x \mapsto \ell(x)$, où ℓ est une forme linéaire. Alors Df est constante égale à ℓ , donc $D^2 f = 0$.

Exemple 2.8.

Soit $f : x \mapsto Q(x)$, où Q est une forme quadratique. Soit φ la forme bilinéaire symétrique associée à Q , c'est-à-dire telle que $Q(x) = \varphi(x, x)$ pour tout x . Alors $D_x f = 2\varphi(x, \cdot)$ et $D^2 f = 2\varphi$. Par conséquent, $\mathcal{Q} = 2Q$.

2.4 Développements limités

Comme en dimension 1, la dérivée seconde permet d'exprimer des développements limités d'ordre 2.

Théorème 2.9.

Soit $f \in \mathcal{C}^2(U, \mathbb{R})$. Alors, pour tout $x \in U$,

$$\begin{aligned} f(x + \vec{h}) &= f(x) + D_x f(\vec{h}) + \frac{1}{2} \mathcal{Q}_x(\vec{h}) + o(\|\vec{h}\|^2) \\ &= f(x) + \sum_{i=1}^n \frac{\partial f}{\partial x_i} h_i + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j} h_i h_j + o(\|\vec{h}\|^2). \end{aligned}$$

Démonstration.

Nous utilisons la formule de dérivation en chaîne (1) couplée à la formule de Taylor à l'ordre 1 sous forme intégrale. Choisissons $\gamma(t) = x + t\vec{h}$, et posons $g := f \circ \gamma$. Alors

$$\begin{aligned} f(x + \vec{h}) &= g(1) \\ &= g(0) + g'(0) + \int_0^1 (1-t)g''(t) dt \\ &= g(0) + g'(0) + \frac{1}{2}g''(0) + \int_0^1 (1-t)(g''(t) - g''(0)) dt \\ &= D_x f(\vec{h}) + \frac{1}{2}\mathcal{Q}_x(\vec{h}) + \int_0^1 (1-t) \left[\mathcal{Q}_{x+t\vec{h}}(\vec{h}) - \mathcal{Q}_x(\vec{h}) \right] dt. \end{aligned}$$

Soit $\varepsilon > 0$. La fonction f étant supposée de classe $\mathcal{C}^1$, il existe $\delta > 0$ tel que, si $\|\vec{k}\| \leq \delta$, alors $\|\mathcal{Q}_{x+\vec{k}} - \mathcal{Q}_x\| \leq \varepsilon$. Alors $|\mathcal{Q}_{x+\vec{k}}(\vec{h}) - \mathcal{Q}_x(\vec{h})| \leq C\varepsilon \|\vec{h}\|^2$. Mais alors, si $\|\vec{h}\| \leq \delta$,

$$\left| \frac{1}{2} \int_0^1 (1-t) \left[\mathcal{Q}_{x+t\vec{h}}(\vec{h}) - \mathcal{Q}_x(\vec{h}) \right] dt \right| \leq \frac{C}{2} \varepsilon \|\vec{h}\|^2. \quad \square$$

Plus généralement, si f est de classe $\mathcal{C}^k$, le même argument implique que

$$f(x + \vec{h}) = f(x) + \sum_{j=1}^k \frac{1}{j!} D_x^j f(\vec{h}, \dots, \vec{h}) + o(\|\vec{h}\|^k),$$

avec, en coordonnées,

$$D_x^j f(\vec{h}, \dots, \vec{h}) = \sum_{i_1, \dots, i_j=1}^n \frac{\partial^j f}{\partial x_{i_1} \dots \partial x_{i_j}} h_{i_1} \dots h_{i_j}.$$

Exemple 2.10. Soit $f : (x, y) \mapsto e^{x^2 + \sin(y)}$ définie sur $\mathbb{R}^2$. Alors

$$\begin{aligned} Jac_{(x,y)}(f) &= \begin{pmatrix} 2xe^{x^2 + \sin(y)} & \cos(y)e^{x^2 + \sin(y)} \end{pmatrix}, \\ Hess_{(x,y)}(f) &= \begin{pmatrix} 2(1+x^2)e^{x^2 + \sin(y)} & 2x \cos(y)e^{x^2 + \sin(y)} \\ 2x \cos(y)e^{x^2 + \sin(y)} & -\sin(y)e^{x^2 + \sin(y)} \end{pmatrix}. \end{aligned}$$

En particulier, en choisissant $(x, y) = (1, 0)$,

$$\begin{aligned} Jac_{(1,0)}(f) &= (2e \quad e), \\ Hess_{(1,0)}(f) &= \begin{pmatrix} 4e & 2e \\ 2e & 0 \end{pmatrix}. \end{aligned}$$

On obtient ainsi le développement limité

$$f(1+h, k) =_{(0,0)} e + 2eh + ek + 2eh^2 + 2ehk + o(h^2 + k^2).$$

Ce dernier résultat se retrouve aussi plus directement à l'aide de développements limités.

2.5 Convexité

Une première application de la dérivée seconde consiste à étendre la caractérisation de la convexité aux fonctions de plusieurs variables.

Définition 2.11 (Convexité).

Supposons U convexe. Une fonction $f : U \rightarrow \mathbb{R}$ est dite **convexe** si, pour tous $p, q \in U$ et tout $t \in [0, 1]$,

$$f((1-t)p + tq) \leq (1-t)f(p) + tf(q) ;$$

autrement dit, si l'épigraphe de f est convexe.

Théorème 2.12.

Supposons U convexe. Soit $f \in \mathcal{C}^2(U, \mathbb{R})$. Alors il y a équivalence entre :

- ▷ f est convexe ;
- ▷ la forme quadratique $\mathcal{Q}_x$ est positive pour tout $x \in U$;
- ▷ les valeurs propres de $\text{Hess}_x(f)$ sont positives (ou nulles) pour tout $x \in U$.

Démonstration.

L'équivalence entre les deux derniers points provient de la théorie des formes quadratiques réelles. Concentrons-nous sur l'équivalence entre les deux premiers points.

Prenons des tranches de U . La fonction f est convexe si et seulement si, pour tous $p, q \in U$, la fonction $g : t \mapsto f((1-t)p + tq)$ est convexe sur $[0, 1]$. La fonction g étant de classe $\mathcal{C}^2$, la fonction f est convexe si et seulement si $g''(t) \geq 0$ pour tout $t \in [0, 1]$. Or, nous avons déjà calculé

$$g''(t) = \mathcal{Q}_{tp+(1-t)q}(\vec{p}\vec{q}).$$

Si $\mathcal{Q}_x$ est positive pour tout x , alors g est convexe pour tous $p, q \in U$, et donc f est convexe. Réciproquement, supposons f convexe, et choisissons $x \in U$. Soit $\vec{h} \in \mathbb{R}^n$. Alors en appliquant ce qui précède à $\vec{p} = x$, $\vec{q} = x + \varepsilon \vec{h}$ et $t = 0$, un choix de $\varepsilon > 0$ garantit que $x + \varepsilon \vec{h} \in U$. Mais alors $g''(0) = \varepsilon^2 \mathcal{Q}_x(\vec{h}) \geq 0$, donc $\mathcal{Q}_x$ est bien positive pour tout x . $\square$

3 Optimisation

Soit $U \subset \mathbb{R}^n$ et $f : U \rightarrow \mathbb{R}$. Le but est de trouver des points minimisants ou maximisant f , et *a fortiori* le minimum ou le maximum de f . En dimension plus grande que 1, les tableaux de variation ne sont plus disponibles, ce qui nécessite une autre stratégie. Celle-ci se déroule en deux temps :

- ▷ On utilise des résultats généraux pour garantir l'existence d'un minimum ou d'un maximum ;
- ▷ Une analyse des points critiques permet de restreindre les candidats potentiels à la minimalité ou maximalité.

Dans ce qui suit, nous chercherons le minimum d'une fonction ; nous laissons à la lectrice le loisir d'adapter ces résultats à la recherche d'un maximum.

3.1 Existence d'un minimum

Dans ce qui suit, U étant ouvert, la compacité ne peut pas être utilisée directement. Nous donnons en exemple deux critères pouvant potentiellement servir à montrer l'existence d'un minimum. Le premier est issu du cours sur la compacité.

Théorème 3.1.

Soit $f \in \mathcal{C}(\mathbb{R}^n, \mathbb{R})$. Si $\lim_{x \rightarrow \infty} f(x) = +\infty$, alors f admet un minimum sur $\mathbb{R}^n$.

Démonstration.

Comme $\lim_{x \rightarrow \infty} f(x) = +\infty$, il existe $R > 0$ tel que $f(x) \geq f(0) + 1$ pour tout $x \notin B(0, R)$. Mais alors $\inf f = \min\{\inf_{\overline{B}(0, R)} f, \inf_{\overline{B}(0, R)^c} f\} = \inf_{\overline{B}(0, R)} f$. Or $\overline{B}(0, R)$ est compact, donc f atteint son minimum sur $\overline{B}(0, R)$, donc son minimum sur $\mathbb{R}^n$. $\square$

Théorème 3.2.

Soit $f \in \mathcal{C}(\mathbb{R}^n, \mathbb{R})$. Si f est périodique en chaque variable, alors f admet un minimum sur $\mathbb{R}^n$.

Démonstration.

Soit $T_i > 0$ une période pour la i -ième variable, de telle sorte que $f((x_i + T_i)_{1 \leq i \leq n}) = f(x)$ pour tout x . Alors $\inf f = \inf_{[0, T_1] \times \dots \times [0, T_n]} f$, et le minimum est atteint par compacité. $\square$

3.2 Points critiques

Supposons que l'on sache que f admette un minimum ; comment le localiser ? Le premier théorème, analogue de la dimension 1, est le suivant :

Théorème 2. *Soit $f \in \mathcal{C}^1(U, \mathbb{R})$. Soit x un extremum local de f . Alors $D_x f = 0$. En particulier, si x minimise f , alors $D_x f = 0$.*

Démonstration. Soit x un extremum local de f ; sans perte de généralité, supposons que x minimise f . Soit $\vec{h} \in \mathbb{R}^n$. Alors, pour tout ε suffisamment petit,

$$f(x + \varepsilon \vec{h}) = f(x) + \varepsilon D_x f(\vec{h}) + o(\varepsilon) \geq f(x).$$

Par conséquent, pour tout ε suffisamment petit,

$$D_x f(\vec{h}) + o(1) \geq 0.$$

En prenant la limite quand ε tend vers 0, on trouve $D_x f(\vec{h}) \geq 0$. En appliquant cela au vecteur $-\vec{h}$, on trouve aussi $D_x f(\vec{h}) \leq 0$, et donc $D_x f(\vec{h}) = 0$. Ceci étant vrai pour tout $\vec{h}$, on a bien $D_x f = 0$. $\square$

Définition 3.3.

*Soit $f \in \mathcal{C}^1(U, \mathbb{R})$. Un point $x \in U$ est appelé **point critique** de f si $D_x f = 0$.*

Les points minimisants sont donc à chercher parmi les points critiques. La fonction $x \mapsto D_x f$ va de $U \subset \mathbb{R}^n$ dans $(\mathbb{R}^n)^*$; les espaces de départ et d'arrivée étant de même dimension, on s'attend typiquement à ce que les points critiques soient isolés. C'est le cas génériquement.

Si f n'est pas définie sur un ouvert, mais par exemple sur une hypersurface X , l'argument précédent implique seulement qu'un point est critique si l'espace tangent à X en x est inclus dans $\text{Ker}(D_x f)$. Ce principe est derrière la méthode dite des multiplicateurs de Lagrange.

De même, si f est défini sur un compact par exemple un simplexe), on peut se retrouver à séparer les cas, en utilisant le théorème précédent pour localiser les points critiques à l'intérieur du domaine, et en étudiant à part le bord du domaine (potentiellement à l'aide des multiplicateurs de Lagrange).

3.3 Analyse locale des points critiques

Soit $x \in U$ un point critique. Le développement limité à l'ordre 2 devient :

$$f(x + \vec{h}) =_0 f(x) + \frac{1}{2} \mathcal{Q}_x(\vec{h}) + o(\|\vec{h}\|^2).$$

Par conséquent, en utilisant la même stratégie que dans la partie précédente :

- ▷ Si x est un minimum local, alors $\mathcal{Q}_x(\vec{h}) \geq 0$ pour tout $\vec{h}$, donc $\mathcal{Q}_x$ est positive.
- ▷ Si x est un maximum local, alors $\mathcal{Q}_x(\vec{h}) \leq 0$ pour tout $\vec{h}$, donc $\mathcal{Q}_x$ est négative.
- ▷ Si $\mathcal{Q}_x$ est définie positive, alors il existe $\lambda_1 > 0$ telle que $\mathcal{Q}_x(\vec{h}) \geq \lambda_1 \|\vec{h}\|^2$ pour tout $\vec{h}$, donc $f(x + \vec{h}) > f(x)$ pour tout $\vec{h}$ suffisamment petit et non nul, donc x est un minimum local.
- ▷ De même, si $\mathcal{Q}_x$ est définie négative, alors x est un maximum local.

Si l'on cherche à déterminer les points critiques qui sont des minima locaux ou des maxima locaux, il suffit donc de déterminer les points pour lesquels les valeurs propres de n sont toutes positives ou toutes négatives respectivement.

Définition 3.4.

Soit $f \in \mathcal{C}^2(U, \mathbb{R})$. Un point critique $x \in U$ est dit **non dégénéré** si $\text{Hess}_x(f)$ est non dégénérée (i.e. de rang n).

En dimension 1, on retrouve les critères habituels portant sur la dérivée seconde de f .

Rappelons qu'en dimension 2, le signe des valeurs propres se détermine facilement à l'aide du déterminant et de la trace. Si x est un extremum local, alors les valeurs propres de $\text{Hess}_x(f)$ sont de même signe, donc $\det(\text{Hess}_x(f)) \geq 0$; réciproquement, si $\det(\text{Hess}_x(f)) > 0$, alors $\text{Hess}_x(f)$ est ou bien définie positive, ou bien définie négative, donc est un extremum local. La trace de f permet alors de distinguer ces deux situations : positive pour un minimum local, négative pour un maximum local.

Plus généralement, étant donné un point critique non dégénéré x , la signature de $\text{Hess}_x(f)$ détermine le comportement de f au voisinage de x :

- ▷ Si $\text{Hess}_x(f)$ est de signature $(n, 0)$, alors x est un minimum local de f .
- ▷ Si $\text{Hess}_x(f)$ est de signature $(0, n)$, alors x est un maximum local de f .
- ▷ Sinon, on dit que x est un **point selle**.

Un point selle n'est ni un minimum ni un maximum local. Plus précisément, on peut trouver deux vecteurs non nuls $\vec{h}_+$ et $\vec{h}_-$, ainsi que $0 < \lambda_-$, λ_+ tels que

$$\begin{aligned} f(x + t\vec{h}_-) &=_0 f(x) - \frac{\lambda_-}{2} t^2 + o(t^2), \\ f(x + t\vec{h}_+) &=_0 f(x) + \frac{\lambda_+}{2} t^2 + o(t^2). \end{aligned}$$

Pour aller plus loin : on peut trouver un système de coordonnées $(H_i)_{1 \leq i \leq n}$ orthonormé en 0 tel que

$$f(x + \vec{h}) =_0 f(x) + \frac{1}{2} \sum_{i=1}^n \lambda_i H_i^2 + o(\|\vec{h}\|^2),$$

et les λ_i ne sont pas tous de même signe.

3.4 Un exemple

Cherchons les minima de

$$f : \begin{cases} \mathbb{R}^2 & \rightarrow \mathbb{R} \\ (x, y) & \mapsto (2 + \cos(y)) \sin(x) \end{cases}.$$

La fonction f est 2π -périodique en chacun de ses paramètres ; elle admet donc un minimum et un maximum globaux. On calcule :

$$\text{Jac}_{(x,y)}(f) = ((2 + \cos(y)) \cos(x) \quad -\sin(y) \sin(x)).$$

Par conséquent, les points critiques de f sont les points (x, y) tels que $(2 + \cos(y)) \cos(x) = -\sin(y) \sin(x) = 0$; ce système est équivalent à $\cos(x) = \sin(y) = 0$, donc les points critiques sont exactement les éléments de $\{(\pi/2 + k\pi, \ell\pi) : (k, \ell) \in \mathbb{Z}^2\}$.

Ici, il serait suffisant de calculer les valeurs atteintes par f en ces points critiques pour conclure. Cherchons cependant à déterminer si ces points sont des minima locaux, des maxima locaux, ou des points selle. On calcule :

$$\text{Hess}_{(x,y)}(f) = \begin{pmatrix} -(2 + \cos(y)) \sin(x) & -\sin(y) \cos(x) \\ -\sin(y) \cos(x) & -\cos(y) \sin(x) \end{pmatrix}.$$

Si (x, y) est un point critique, alors $\cos(x) = \sin(y) = 0$, et donc l'expression de la matrice hessienne se simplifie :

$$\text{Hess}_{(x,y)}(f) = \begin{pmatrix} -(2 + \cos(y)) \sin(x) & 0 \\ 0 & -\cos(y) \sin(x) \end{pmatrix}.$$

- ▷ Si $x \equiv \pi/2 \pmod{2\pi}$ et $y \equiv 0 \pmod{2\pi}$ alors $\text{Hess}_{(x,y)}(f)$ a pour valeurs propres -3 et -1 , donc (x, y) est un maximum local et $f(x, y) = 3$.
- ▷ Si $x \equiv \pi/2 \pmod{2\pi}$ et $y \equiv \pi \pmod{2\pi}$ alors $\text{Hess}_{(x,y)}(f)$ a pour valeurs propres -1 et 1 , donc (x, y) est un point selle et $f(x, y) = 1$.
- ▷ Si $x \equiv 3\pi/2 \pmod{2\pi}$ et $y \equiv 0 \pmod{2\pi}$ alors $\text{Hess}_{(x,y)}(f)$ a pour valeurs propres 3 et 1 , donc (x, y) est un minimum local et $f(x, y) = -3$.
- ▷ Si $x \equiv 3\pi/2 \pmod{2\pi}$ et $y \equiv \pi \pmod{2\pi}$ alors $\text{Hess}_{(x,y)}(f)$ a pour valeurs propres 1 et -1 , donc (x, y) est un point selle et $f(x, y) = -1$.

En particulier, $\min f = -3$ et $\max f = 3$.

Exercice : La fonction f ci-dessus donne l'ordonnée d'une paramétrisation standard du tore dans l'espace ; interprétez géométriquement ses points critiques.

3.5 Domaines compacts

Si f est définie non pas sur un ouvert mais sur un compact, l'existence d'un minimum est directement garantie par compacité ; cependant, la localisation de ce minimum (la seconde étape de notre stratégie) devient beaucoup plus délicate. Par exemple, si f est définie sur le tétraèdre $\{(x, y, z) : x, y, z \geq 0, x + y + z \leq 1\}$, on peut utiliser la stratégie précédente en :

- ▷ cherchant et analysant les points critiques à l'intérieur du tétraèdre $\{(x, y, z) : x, y, z > 0, x + y + z < 1\}$;
- ▷ cherchant et analysant les points critiques sur chacune des 4 faces du tétraèdre $\{(0, y, z) : y, z > 0, y + z < 1\}$, $\{(x, y, z) : x, y, z > 0, x + y + z = 1\}$... ;
- ▷ cherchant et analysant les points critiques sur chacune des 6 arêtes du tétraèdre $\{(x, 0, 0) : 0 < x < 1\}$, $\{(0, y, 1 - y) : x, y, z > 0, x + y + z = 1\}$... ;
- ▷ cherchant et analysant les points critiques sur chacun des 4 sommets du tétraèdre $(0, 0, 0)$, $(1, 0, 0)$...

soit, en tout, 15 problèmes d'optimisation à résoudre, dont 4 triviaux (dimension 0) et 5 en dimension plus grande que 1. Si cela est faisable, les calculs peuvent prendre beaucoup de temps. Dans certains cas, des arguments spécifiques permettent de simplifier les calculs ; par exemple, si f est concave, alors son minimum est atteint en l'un des quatre sommets du tétraèdre.

4 Un peu de topographie

4.1 Courbes de niveau

Définition 4.1. Soit $f : U \rightarrow \mathbb{R}$ et $h \in \mathbb{R}$. La **courbe de niveau** h est l'ensemble $\{x \in U : f(x) = h\} = f^{-1}(\{h\})$. On le notera parfois $C_f(h)$.

Les courbes de niveau sont représentées dans de nombreuses cartes : topographiques (courbes de même altitude), météorologiques (isobares et isothermes)... Le but de cette partie est d'en comprendre la structure géométrique.

Si f est continue, alors toutes les courbes de niveau sont des fermés (en tant que préimages de fermés par une fonction continue). Réciproquement, tout fermé F est courbe de niveau pour une fonction continue, par exemple courbe de niveau 0 pour la fonction $f : x \mapsto d(x, F)$, qui est 1-lipschitzienne. Avec plus d'efforts, on peut montrer que tout fermé est courbe de niveau d'une fonction C^∞ . Cependant, ces situations sont exceptionnelles : on a choisi la fonction f et le niveau 0 pour obtenir les résultats voulu. Le but ici est d'étudier les ensemble de niveau sous des conditions générales.

4.2 Théorème des fonctions implicites

Le cas le plus simple est celui d'une fonction qui est une forme linéaire non nulle ; en ce cas, les ensembles de niveau sont des hyperplans. Par développement limité, si $D_x f \neq 0$, on peut s'attendre à ce que les ensembles de niveau ressemblent près de x à des hyperplans. Nous allons montrer dans un premier temps que ces ensembles de niveau ressemblent à des graphes de fonctions.

4.2.1 Le cercle et la sphère

Concentrons-nous sur le cercle et la sphère, qui présentent l'essentiel des difficultés possibles. Soit

$$f : \begin{cases} \mathbb{R}^2 & \rightarrow \mathbb{R} \\ (x, y) & \mapsto x^2 + y^2 \end{cases}.$$

L'ensemble de niveau 1 de f est le cercle unité de centre 0. Ce cercle n'est pas le graphe d'une fonction ; cependant, on peut exprimer certains morceaux du cercle comme graphes de fonctions. Par exemple, la moitié supérieure du cercle est le graphe de la fonction $x \mapsto \sqrt{1 - x^2}$, tandis que sa partie inférieure est le graphe de la fonction $x \mapsto -\sqrt{1 - x^2}$. Quitte à enlever les points $(1, 0)$ et $(-1, 0)$, ces deux fonctions sont de classe C^∞ .

Formalisons un peu mieux ces assertions :

- ▷ La fonction $x \mapsto \sqrt{1 - x^2}$ n'est pas définie en-dehors de $[-1, 1]$, ce qui est normal, car le cercle est inclus dans la bande $[-1, 1] \times \mathbb{R}$.
- ▷ Deux points sont gênants, les points $(1, 0)$ et $(-1, 0)$, en lesquels les tangentes au cercle sont verticales (et $D_x f$ est proportionnel à $(1 \ 0)$).
- ▷ Enlevons ces points gênant en nous restreignant à l'intervalle $(-1, 1)$. Alors un point p appartient à $\{q : f(q) = 1\} \cap (-1, 1) \times (0, +\infty)$ si et seulement si p appartient au graphe de la fonction $x \mapsto \sqrt{1 - x^2}$.

- ▷ De même, un point p appartient à $\{q : f(q) = 1\} \cap (-1, 1) \times (-\infty, 0)$ si et seulement si p appartient au graphe de la fonction $x \mapsto -\sqrt{1-x^2}$.

Autrement dit, on peut exprimer l'ensemble de niveau comme graphe d'une fonction $\mathcal{C}^\infty$ quitte à enlever les points en lesquels le graphe aurait une tangente verticale (c'est-à-dire les points où la dérivée de f est "horizontale"), et éventuellement en restreignant l'ensemble de départ et l'ensemble d'arrivée pour garantir l'existence et l'unicité d'un tel graphe.

Tout cela reste valable pour la sphère, en utilisant la fonction

$$g : \begin{cases} \mathbb{R}^3 & \rightarrow \mathbb{R} \\ (x, y, z) & \mapsto x^2 + y^2 + z^2 \end{cases}.$$

Un point p appartient à $\{q : g(q) = 1\} \cap B(0, 1) \times (0, +\infty)$ si et seulement si p appartient au graphe de la fonction $(x, y) \mapsto \sqrt{1-x^2-y^2}$. De même, un point p appartient à $\{q : g(q) = 1\} \cap B(0, 1) \times (-\infty, 0)$ si et seulement si p appartient au graphe de la fonction $(x, y) \mapsto -\sqrt{1-x^2-y^2}$.

Dans ce cas, les points gênants sont ceux où le graphe de g est vertical, c'est-à-dire tel que $(0, 0, 1) \in \text{Ker}(D_x g)$, ou plus simplement $D_x g(0, 0, 1) = 0$.

4.2.2 Le cas général

Cette formulation restant valable dans un cadre beaucoup plus général.

Théorème 3 (Théorème des fonctions implicites).

Soit $f \in \mathcal{C}^k(U, \mathbb{R})$ et $h \in \mathbb{R}$. Notons $\mathcal{C}_f(h)$ la courbe de niveau h , et soit $x \in \mathcal{C}_f(h)$.

Supposons que $D_x f(0, \dots, 0, 1) \neq 0$. Alors on peut trouver :

▷ *Un ouvert V de $\mathbb{R}^{n-1}$;*

▷ *Un intervalle ouvert I ;*

▷ *Une fonction $g \in \mathcal{C}^k(V, I)$;*

telles que $x \in V \times W$, et $\mathcal{C}_f(h) \cap V \times I$ coïncide avec le graphe de g (ou autrement dit, pour tous $(y, z) \in V \times I$, on ait $(y, z) \in \mathcal{C}_f(h)$ si et seulement si $z = g(y)$).

Dans le cas du cercle, on peut de plus "tourner le plan d'un quart de tour" pour pouvoir exprimer le cercle comme le graphe des fonctions $y \mapsto \sqrt{1-y^2}$ et $y \mapsto -\sqrt{1-y^2}$, c'est-à-dire comme union d'ensembles de la forme $\{(\sqrt{1-y^2}, y) : y \in (-1, 1)\}$ et $\{(-\sqrt{1-y^2}, y) : y \in (-1, 1)\}$. Cette représentation a l'avantage de mieux convenir autour des points $(1, 0)$ et $(-1, 0)$, mais a l'inconvénient d'être inadaptée aux points $(0, 1)$ et $(0, -1)$.

De même, si $D_x f \neq 0$, alors on peut trouver un vecteur de la base canonique $(0, \dots, 0, 1, 0, \dots, 0)$ tel que $D_x f(0, \dots, 0, 1, 0, \dots, 0) \neq 0$, et ainsi écrire près de x l'ensemble de niveau comme graphe d'une fonction.

Si $D_x f \neq 0$ pour tout $x \in \mathcal{C}_f(h)$, alors, quitte à tourner l'espace, $\mathcal{C}_f(h)$ peut être partout décrite comme le graphe d'une fonction. On dit alors que h est une **valeur régulière**³ de f , et $\mathcal{C}_f(h)$ est alors une **sous-variété**. C'est le cas, par exemple, des sphères.

4.2.3 Application : racines de polynômes

Une application intéressante du théorème des fonctions implicites est la paramétrisation des racines de polynômes. Donnons-nous par exemple $P \in \mathbb{R}_5[X]$, et supposons que P a une racine réelle *simple* x_0 .

3. D'après le théorème de Sard, si f est de classe $\mathcal{C}^\infty$, presque toutes les valeurs de h sont des valeurs régulières.

Soit

$$f : \begin{cases} \mathbb{R}^2 & \rightarrow \mathbb{R} \\ (x, y) & \mapsto x + P(y) \end{cases} ,$$

une fonction $\mathcal{C}^\infty$. Alors $(0, x_0) \in \mathcal{C}_f(0)$, et $D_{(0, x_0)}f = (1 \quad P'(x_0))$. En particulier, $D_{(0, x_0)}f(0, 1) = P'(x_0)$, et $P'(x_0) \neq 0$ car x_0 est racine simple. Par le théorème des fonctions implicites, on peut trouver deux intervalles $0 \in I$, $x_0 \in J$ et une fonction $\lambda : I \rightarrow J$ tels que, pour tous $(y, x) \in I \times J$, on ait $P(x) + y = 0$ si et seulement si $x = \lambda(y)$. Autrement dit, $(y + P)_{y \in I}$ est une famille à un paramètre de polynômes, et la racine simple x_0 dépend continûment du paramètre y .

Plus généralement, soit

$$g : \begin{cases} \mathbb{R}_5[X] \times \mathbb{R} \simeq \mathbb{R}^7 & \rightarrow \mathbb{R} \\ (P, x) & \mapsto P(x) \end{cases} .$$

Alors $(P, x_0) \in \mathcal{C}_g(0)$ et $D_{(P, x_0)}g(0, \dots, 0, 1) = P'(x_0)$. Par le même argument, on peut trouver un ouvert V de $\mathbb{R}_5[X]$ contenant P , un intervalle ouvert I contenant x_0 et une fonction $\lambda : \mathbb{R}_5[X] \rightarrow \mathbb{R}$ de classe $\mathcal{C}^\infty$ telles que, pour tout $Q \in V$ et $x \in I$, on ait $Q(x) = 0$ si et seulement si $x = \lambda(Q)$.

Autrement dit, dans chaque espace $\mathbb{R}_n[X]$, les racines simples dépendent de façon lisse du polynôme ! Ces fonctions ne peuvent cependant pas être écrites à l'aide de radicaux si $n \geq 5$. De plus :

- ▷ Ces fonctions ne sont définies que localement. Si l'on travaille dans $\mathbb{C}_n[X]$, il y a des obstructions à l'existence de telles fonctions dans tout $\mathbb{C}_n[X]$, même en enlevant les polynômes ayant des racines doubles : ces obstructions sont de même nature que celles empêchant l'existence d'une fonction logarithme sur $\mathbb{C}^*$.
- ▷ Ce résultat est faux si x_0 n'est pas une racine simple ; on pourra par exemple considérer la famille de polynômes $(y + X^2)_{y \in \mathbb{R}}$ près de $y = 0$.
- ▷ Ce résultat s'étend aux valeurs propres de matrices : si λ est une valeur propre simple d'une matrice A , alors λ dépend de façon lisse des coefficients de la matrices près de A .

4.3 Tangentes aux courbes de niveau

4.3.1 Cas général

Comme vu précédemment, si $D_x f \neq 0$, alors la courbe de niveau à laquelle x appartient, c'est-à-dire $\mathcal{C}_f(f(x))$ est près de x le graphe d'une fonction. On peut donc approcher cette courbe de niveau par un hyperplan tangent. Pour déterminer ce graphe, il suffit de remplacer f par son approximation linéaire.

Rappelons que $\mathcal{C}_f(f(x)) = \{y \in U : f(y) = f(x)\}$. De plus, au voisinage de x , la fonction f peut être approchée par son développement limité à l'ordre 1 :

$$f(y) =_x f(x) + D_x f(\vec{xy}) + o(\|\vec{xy}\|) .$$

En injectant cette expression dans la courbe de niveau, on obtient

$$\begin{aligned} T_x \mathcal{C}_f &= \{y \in \mathbb{R}^n : f(x) + D_x f(\vec{xy}) = f(x)\} \\ &= \{y \in \mathbb{R}^n : D_x f(\vec{xy}) = 0\} \\ &= y + \text{Ker}(D_x f) ; \end{aligned}$$

c'est l'équation d'un hyperplan affine.

Cet hyperplan a la propriété suivante (que nous ne chercherons pas à démontrer) : pour tout $y \in T_x \mathcal{C}_f$,

$$d(y, \mathcal{C}_f) =_x o(\|\vec{xy}\|) ,$$

et $T_x \mathcal{C}_f$ est le plus grand sous-espace vectoriel sur lequel cette propriété est valable. Autrement dit, l'hyperplan tangent approche la courbe de niveau à l'ordre 1.

Rappelons au passage que $\text{Ker}(D_x f) = (\nabla_x f)^\perp$, ce qui justifie que *les courbes de niveau sont orthogonales au gradient*.

Par exemple, reprenons

$$g : \begin{cases} \mathbb{R}^3 & \rightarrow \mathbb{R} \\ (x, y, z) & \mapsto x^2 + y^2 + z^2 \end{cases}.$$

On calcule $D_X g = 2(x \ y \ z)$, donc $\nabla_X g = 2X$. En particulier, si $X \in \mathbb{S}_2 = \mathcal{C}_g(1)$, alors le plan tangent à la sphère unité est $T_X \mathbb{S}_2 = X + (\text{Vect}(X))^\perp$, de direction $X^\perp$.

Exemple 4.2 (Graphe d'une fonction).

Soit $f \in \mathcal{C}^1(U, \mathbb{R})$. Le graphe de f est l'ensemble $\{(x, f(x)) : x \in U\}$. C'est donc l'ensemble de niveau 0 de la fonction

$$g : \begin{cases} U \times \mathbb{R} & \rightarrow \mathbb{R} \\ (x, y) & \mapsto y - f(x) \end{cases}.$$

On calcule $D_{(x,y)} g = (-D_x f \ 1)$. Le plan tangent a donc pour direction $\{(\vec{u}, v) : -D_x f(\vec{u}) + v = 0\} = \{(\vec{u}, v) : v = D_x f(\vec{u})\}$; c'est le graphe de $D_x f$. Le plan tangent a finalement pour équation

$$v = f(x) + D_x f(\vec{x}\vec{v}).$$

4.4 Lemme de Morse

Si la plupart des courbes de niveau sont des courbes de niveau de valeurs régulières, il existe en général certaines courbes de niveau contenant des points critiques; pensons par exemple à l'ensemble de niveau 0 de la fonction $(x, y) \mapsto x^2 + y^2$. Pour une fonction f générique, ces points critiques sont isolés et non dégénérés.

Soit $f \in \mathcal{C}^2(U, \mathbb{R}^n)$. Soit $x \in U$ un point critique non dégénéré. On sait qu'il existe une base orthonormée dans laquelle

$$\mathcal{Q}_x(\vec{h}) = \sum_{i=1}^s H_i^2 - \sum_{i=s+1}^n H_i^2,$$

où s est la signature de $\mathcal{Q}_x$. Autrement dit, la dérivée seconde de f admet, dans un bon système de coordonnées, une forme standard. Le lemme de Morse affirme que f elle-même peut être mise sous cette forme dans un bon système de coordonnées.

Théorème 4 (Lemme de Morse).

Soit $f \in \mathcal{C}^2(U, \mathbb{R}^n)$. Soit $x \in U$ un point critique non dégénéré de signature s . Alors il existe un voisinage V de x et un difféomorphisme^a $\psi : V \rightarrow W \subset \mathbb{R}^n$ tels que

$$f(\psi^{-1}(z)) = f(x) + \sum_{i=1}^s z_i^2 - \sum_{i=s+1}^n z_i^2 \quad \forall z \in W.$$

a. C'est-à-dire une nouvelle paramétrisation d'un voisinage de x .

En particulier,

$$\begin{aligned} y \in \mathcal{C}_f(h) \cap V &\iff f(y) = h \\ &\iff \exists z \in W, f(\psi^{-1}(z)) = h \\ &\iff \exists z \in W, \sum_{i=1}^s z_i^2 - \sum_{i=s+1}^n z_i^2 = h - f(x). \end{aligned}$$

Autrement dit, $\mathcal{C}_f(h) \cap V$ est l'image par ψ d'un ensemble de niveau de $z \mapsto \sum_{i=1}^s z_i^2 - \sum_{i=s+1}^n z_i^2$ de 0. Il suffit donc de comprendre ces derniers !

Si $n = 1$, alors les ensembles de niveau sont :

- ▷ Si $s = 1$: vides pour $h < f(x)$, réduits à $\{x\}$ pour $h = f(x)$, et deux points presque équidistants de x si $h > f(x)$.
- ▷ Si $s = -1$: vides pour $h > f(x)$, réduits à $\{x\}$ pour $h = f(x)$, et deux points presque équidistants de x si $h < f(x)$.

Les choses deviennent intéressantes pour $n = 2$. Les ensembles de niveau sont alors :

- ▷ Si $s = 2$: vides pour $h < f(x)$, réduits à $\{x\}$ pour $h = f(x)$, et une ellipse déformée si $h > f(x)$; la famille de presque-ellipses ainsi obtenues sont concentriques, et les petits et grands axes sont les sous-espaces propres de $\text{Hess}_x(f)$.
- ▷ Si $s = 0$: vides pour $h > f(x)$, réduits à $\{x\}$ pour $h = f(x)$, et une ellipse déformée si $h < f(x)$; la famille de presque-ellipses ainsi obtenues sont concentriques, et les petits et grands axes sont les sous-espaces propres de $\text{Hess}_x(f)$.
- ▷ Si $s = 1$: des courbes (hyperboles déformées) pour $h \neq f(x)$, et l'union de deux courbes transverses si $h = f(x)$. Les sous-espaces propres de $\text{Hess}_x(f)$ sont les bissectrices de ces deux courbes à leur point sécant.

Pour $n = 3$, on retrouve des courbes de niveau qui vont ressembler à des ellipsoïdes, à des hyperboloïdes à une nappe ou à des hyperboloïdes à deux nappes.

Finalement, tout cela a un lien avec la topologie ! Le nombre de points critiques et leur signature dépend de la “forme” de la surface sur laquelle on se place. Par exemple, la donnée d'une fonction de $\mathbb{R}^2$ dans $\mathbb{R}$ qui soit 2π -périodique en chaque variable est la même chose que la donnée d'une fonction définie sur le tore. Soit $f \in \mathcal{C}^2(\mathbb{R}^2, \mathbb{R})$ une fonction 2π -périodique en chaque variable et telle que tous ses points critiques soient non dégénérés. Notons :

- ▷ n_2 le nombre de points critiques dans $[0, 2\pi)^2$ de signature 2 (c'est-à-dire le nombre de maxima locaux dans ce carré) ;
- ▷ n_1 le nombre de points critiques dans $[0, 2\pi)^2$ de signature 1 (c'est-à-dire le nombre de points selle dans ce carré) ;
- ▷ n_0 le nombre de points critiques dans $[0, 2\pi)^2$ de signature 0 (c'est-à-dire le nombre de minima locaux dans ce carré).

Alors, d'après le théorème du point fixe de Lefschetz-Hopf,

$$n_2 - n_1 + n_0 = \chi(\mathbb{T}^2) = 0,$$

où $\chi(\mathbb{T}^2)$ est la caractéristique d'Euler du tore. C'est le cas par exemple de la fonction $(x, y) \mapsto (2 + \cos(y)) \sin(x)$ donnée en exemple plus tôt, qui avait un maximum local, un minimum local et deux points selles. Notons en particulier qu'une telle fonction aura toujours au moins un maximum local et un minimum local, donc au moins deux points selle dans le carré $[0, 2\pi)^2$.

5 Références

Le matériel ci-dessus est traité partiellement (mais plus que suffisamment au niveau de l'agrégation) dans le livre de **Skandalis**.