

2. Outils mathématiques et conditions d'optimalité

2.1 Notions de base

1. Pour tout $n \in \mathbb{N}^*$, $\mathbb{R}^n$ désigne l'espace **euclidien** $\mathbb{R} \times \mathbb{R} \times \mathbb{R} \cdots \mathbb{R}$ produit n fois. En général un vecteur $x \in \mathbb{R}^n$ sera noté par $x = (x_1, x_2, \dots, x_n)^T$ (vecteur colonne).
2. Pour tous $x, y \in \mathbb{R}^n$, on note par $\langle x, y \rangle \in \mathbb{R}$ le **produit scalaire** de x et y qui est donné par

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i$$

3. Un espace vectoriel $\chi \subset \mathbb{R}^n$ est dit **espace normé** si, sur cet espace est définie la fonctionnelle $\|\cdot\| : x \rightarrow \mathbb{R}$ appelée la norme et qui vérifie les conditions suivantes:
 - $\|x\| \geq 0$, $\forall x \in \chi$ et $\|x\| = 0 \Leftrightarrow x = 0$.
 - $\|\alpha x\| = |\alpha| \cdot \|x\|$, $\forall \alpha \in \mathbb{R}$, $\forall x \in \chi$.
 - $\|x + y\| \leq \|x\| + \|y\|$, $\forall x, y \in \chi$. Cette dernière inégalité s'appelle *inégalité triangulaire*.
- $x \in \mathbb{R}$, $\|x\| = |x|$, donc $(\mathbb{R}, |\cdot|)$ est un espace normé.
- $x \in \mathbb{R}^n$, $\|x\| = \left(\sum_{i=1}^n x_i^2 \right)^{1/2}$, $x = (x_1, x_2, \dots, x_n)^T$, donc $\mathbb{R}^n, \|\cdot\|$ est aussi un espace normé.

On sait aussi que $\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{\sum_{i=1}^n x_i^2}$

2.2 Fonctions différentiables et différentielles

Dans tous ce qui suit, les fonctions sont toujours supposées différentiables au minimum d'ordre deux. Si f est définie sur $U \subset \mathbb{R}^n$ et à valeurs réelles, sa différentiable en $x \in U$ est l'application linéaire notée df_x :

$$f(x+h) = f(x) + df_x(h) + \|h\|\varepsilon(h)$$

avec $df_x(h) = df_x(h_1, h_2, \dots, h_n) = \frac{\partial f}{\partial x_1} h_1 + \frac{\partial f}{\partial x_2} h_2 + \dots + \frac{\partial f}{\partial x_n} h_n = \langle \nabla f(x), h \rangle$

Définition 2.2.1 — Gradient. Une fonction continue $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est dite continûment différentiable si $\frac{\partial f}{\partial x_i}$ existe et est continue. La fonction f est dite aussi de classe C_1 . Le gradient de la fonction f en x est défini par le vecteur:

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right]^T$$

Le gradient indique la direction de plus grande pente. Graphiquement, le gradient donne la direction de la normale au plan tangent à la surface de la fonction de coût.

■ **Exemple 2.1** Soit la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par

$$f(x_1, x_2) = x_1^2 + x_1 \sin(x_2) + 4x_2^3$$

Analytiquement, le gradient de la fonction f s'écrit:

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2x_1 + \sin x_2 \\ x_1 \cos x_2 + 12x_2^2 \end{pmatrix}$$

Il est clair que le gradient de $f \in C^1$.

Définition 2.2.2 — Contour d'une fonction. Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$, alors:

- $z = f(x)$ définit une surface dans $\mathbb{R}^{n+1}$.
- $f(x) = c$, avec c constant définissant les courbes de niveau ou des contours de cette surface. Un contour de niveau c est l'ensemble des points $S(c) = \{x | f(x) = c\}$



Le gradient est perpendiculaire au contour.

■ **Exemple 2.2** Dans cet exemple, nous présentons l'essentiel des fonctions fournies par Matlab pour la visualisation de la fonction en 3D, la projection sur 2D ainsi que l'évolution du gradient.

Soit la fonction $f(x, y) = \frac{x^3}{3} - x - (\frac{y^3}{3} - y)$. Pour visualiser la fonction en 3D, il faut d'abord créer un maillage à l'aide de la fonction *meshgrid* et puis utiliser la fonction *mesh* ou *surf*. Le code Matlab suivant permet de tracer la fonction f :

```

1 [x, y] = meshgrid(-2:0.25:2, -1:0.2:1);
2 z = x.^3/3 - x - (y.^3/3 - y);
3 figure(1)
4 surf(x, y, z)
```

L'exécution de ce programme permet de visualiser la forme 3D illustrée par Fig 2.1:

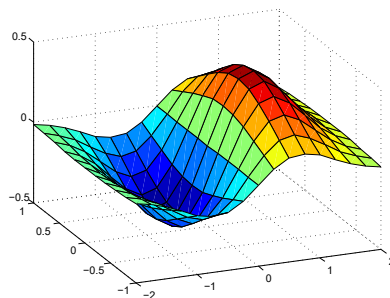


Figure 2.1: Visualisation 3D

Le code suivant permet de visualiser les différents vecteurs gradient sur le tracé 3D (voir Fig 2.2).

```
1 [u,v,w] = surfnorm(x,y,z);
2 hold on
3 quiver3(x,y,z,u,v,w,0.5)
```

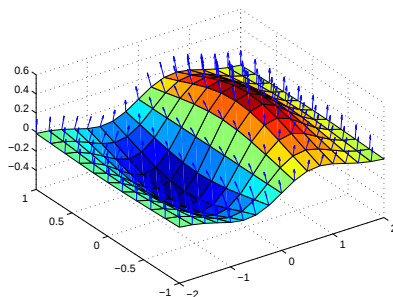


Figure 2.2: Vecteurs gradients sur 3D

La projection de la courbe 3D sur 2D consiste à projeter des coupes réalisées sur z (la valeur de $f(x, y)$) sur le plan (x, y) . Le tracé des contours peut être fait de différents manières: avec le tracé 3D, sur le plan (x, y) ou sous une forme 3D (voir Fig 2.3).

```
1 close all
2 subplot(2,2,1)
3 meshc(x,y,z)
4 subplot(2,2,2)
5 [c,h]=contour(x,y,z,[-0.39 -0.2 -0.1 0 0.1 0.2 0.38])
6 clabel(c,h)
7 grid
8 subplot(2,2,3)
9 [c,h]=contourf(x,y,z,[-0.39 -0.2 -0.1 0 0.1 0.2 0.38], 'k');
10 subplot(2,2,4)
11 [c,h]=contour3(x,y,z,[-0.4 -0.3 -0.2 -0.1 0.1 0.2 0.3 0.4], 'b')
12 axis tight
```

Le tracé des vecteurs gradients a plus de signification lorsqu'il est présenté sur le tracé des contours. Il indique le sens des mouvements (ou bien la vitesse)(voir Fig 2.4).

```
1 [x,y] = meshgrid(-2:.2:2);
2 z = x.*exp(-x.^2 - y.^2);
3 [DX,DY] = gradient(z,.2,.2);
4 contour(x,y,z)
5 hold on
6 quiver(x,y,DX,DY)
```

Définition 2.2.3 — Différentielle seconde et Hessienne. Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est dite deux fois continûment différentiable en $x \in \mathbb{R}^n$ si $\frac{\partial^2 f}{\partial x_i \partial x_j}$ pour $1 \leq i, j \leq n$ existe et est

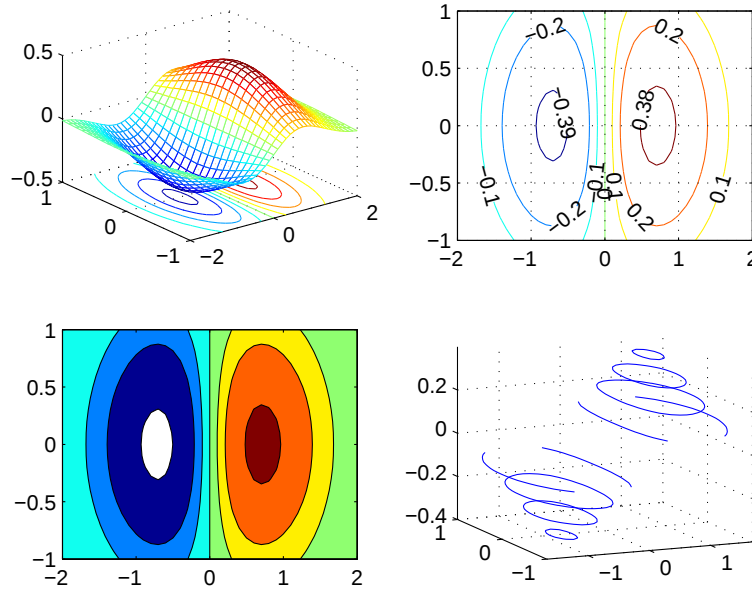


Figure 2.3: Contours

continue. La matrice Hessienne de f en x est une fonction d'ordre n définie par:

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Théorème 2.2.1 — Schwarz. Soit U un ouvert de $\mathbb{R}^n$ et $a \in U$. Si $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est deux fois différentiable en a , alors pour tous $i, j \in \{1, \dots, n\}$,

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a)$$

2.2.1 Définie positivité et négativité

Soit une matrice carrée A de $\mathbb{R}^{n \times n}$, alors on a les définitions suivantes:

Définition 2.2.4 — Définie positive. On dit que la matrice A est *définie positive* si pour n'importe quel vecteur $x \neq 0$, on a la relation:

$$x^T A x > 0, \forall x \neq 0$$

Définition 2.2.5 — Semi-Définie positive (SDP). On dit que la matrice A est *semi-définie positive* si pour n'importe quel vecteur $x \neq 0$, on a la relation:

$$x^T A x \geq 0, \forall x \neq 0$$



Pour des formes carrées **définies négatives** ou **semi-définies négatives (SDN)**, les définitions sont similaires sauf en ce qui concerne les inégalités qui doivent alors

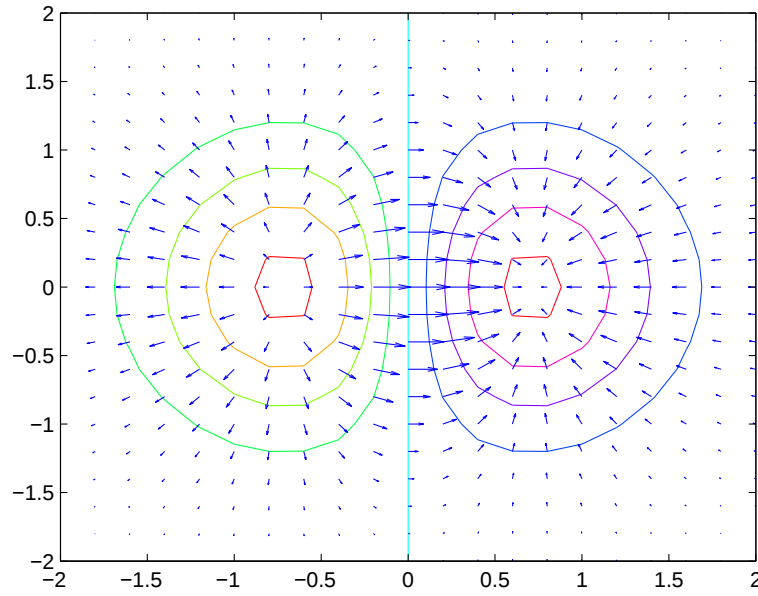


Figure 2.4: Vecteurs gradients

être renversées (< 0).

Théorème 2.2.2 — Sylvester. La condition nécessaire et suffisante pour qu'une matrice symétrique soit strictement définie positive est que tous les principaux mineurs successifs soient strictement positifs.

Illustration du théorème

Soit $A \in \mathbb{R}^{3 \times 3}$ une matrice symétrique définie comme suit:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Les principaux mineurs successifs de A sont les trois déterminants suivants:

$$|a_{11}| > 0, \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0$$



Les principaux mineurs successifs de A sont les déterminants des sous-matrices de A obtenues en enlevant successivement la dernière colonne de droite et la rangée d'en bas.

Corollaire 2.2.3 Dans le cas d'une matrice quelconque et qui n'est pas forcément symétrique, alors une matrice définie positive implique que tous les mineurs principaux soient strictement positifs. Rappelons que les mineurs de la matrice précédente

sont:

$$|a_{11}| > 0, |a_{22}| > 0, |a_{33}| > 0, \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} > 0, \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} > 0, \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0$$

Elle sera définie négative si ces déterminants alternent en signe:

$$|a_{11}| < 0, |a_{22}| < 0, |a_{33}| < 0, \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} > 0, \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} > 0, \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} < 0$$

Corollaire 2.2.4 Cependant, pour déterminer si une matrice est semi-définie positive, il ne suffit pas que les principaux mineurs successifs soient non négatifs (i.e. ≥ 0). Il faut que tous les mineurs principaux soient non négatifs, ce qui implique le calcul de plusieurs déterminants. Par exemple, dans le cas de A la matrice 3×3 vue plus haut, il faut vérifier que :

$$|a_{11}| \geq 0, |a_{22}| \geq 0, |a_{33}| \geq 0, \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \geq 0, \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} \geq 0, \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} \geq 0, \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \geq 0$$

Pour déterminer si une matrice est semi-définie négative, il faut que tous les mineurs principaux d'ordre impair soient non positifs (i.e. ≤ 0) et que tous les mineurs principaux d'ordre pair soient non négatifs (i.e. ≥ 0). Par exemple, dans le cas de A la matrice 3×3 vue plus haut, il faut vérifier que:

$$|a_{11}| \leq 0, |a_{22}| \leq 0, |a_{33}| \leq 0, \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \geq 0, \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} \geq 0, \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} \geq 0, \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \leq 0$$

Exercice 2.1 Montrer par deux méthodes la définie positivité de la matrice :

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

2.3 Fonctions convexes et concaves

Définition 2.3.1 — Domaine convexe. Un sous ensemble χ de $\mathbb{R}^n$ est dit convexe, si et seulement si:

$$\forall x, y \in \chi, \forall \lambda \in [0, 1] \text{ on a : } \lambda x + (1 - \lambda)y \in \chi$$

■ **Exemple 2.3** Dans la figure suivante, on présente deux exemples d'un ensemble convexe (gauche) et non convexe (droite) ■

Définition 2.3.2 — Fonction convexe. Étant donnée une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ définie sur une partie convexe.

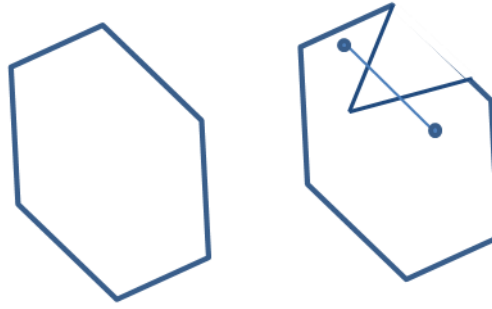


Figure 2.5: Convexité des ensembles

- f est dite fonction convexe si et seulement si: $\forall x_1, x_2 \in \chi, \forall \lambda \in [0, 1]$ on a:

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$$

- f est dite strictement convexe si et seulement si: $\forall x_1, x_2 \in \chi, \forall \lambda \in [0, 1]$ on a:

$$f(\lambda x_1 + (1 - \lambda)x_2) < \lambda f(x_1) + (1 - \lambda)f(x_2)$$

- On dit que f est concave (respectivement strictement concave si l'inégalité est stricte) si :

$$f(\lambda x_1 + (1 - \lambda)x_2) \geq \lambda f(x_1) + (1 - \lambda)f(x_2)$$

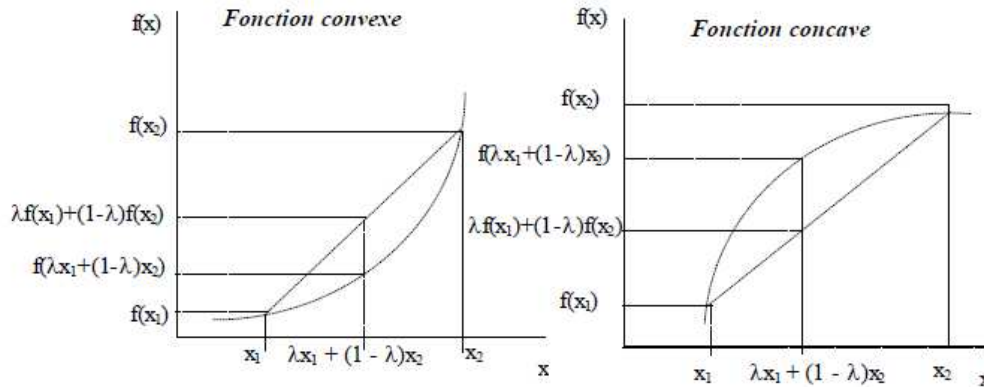


Figure 2.6: Convexité des fonctions

La convexité peut être exprimée via le gradient et le Hessian de la fonction.

Théorème 2.3.1 Étant donnée une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ définie sur une partie convexe χ .

A- Caractérisation de la convexité avec le gradient Supposons que f est de classe C^1 , alors:

1. f est convexe sur χ si et seulement si:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \forall x, y \in \chi$$

2. f est strictement convexe sur χ si et seulement si:

$$f(y) > f(x) + \langle \nabla f(x), y - x \rangle, \forall x, y \in \chi$$

3. f est strictement convexe sur χ , si ∇f est monotone sur χ

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0, \forall x, y \in \chi$$

B- Caractérisation de la convexité avec le Hessien Supposons que f est de classe C^2 , alors:

1. f est convexe sur χ si et seulement si:

$$\langle \nabla^2 f(y) \cdot (y - x), y - x \rangle \geq 0, \forall x, y \in \chi$$

2. f est strictement convexe sur χ si et seulement si:

$$\langle \nabla^2 f(y) \cdot (y - x), y - x \rangle > 0, \forall x, y \in \chi$$

Proposition 2.3.2 Nous pouvons déterminer la convexité/concavité d'une fonction en déterminant si le Hessien associé est semi-défini positif ou négatif, comme suit

- f est convexe, si et seulement si le Hessien F est semi-défini positif pour $x \in \chi$.
- Si F est défini positif pour tous $x \in \chi$, alors f est strictement convexe.
- f est concave, si et seulement si le Hessien F est semi-défini négatif pour $x \in \chi$.
- Si F est défini négatif pour tous $x \in \chi$, alors f est strictement concave.

Corollaire 2.3.3 Lorsque f est convexe, elle admet un minimum global et lorsqu'elle est concave, elle admet un maximum global. Si la convexité ou la concavité n'est observée que localement, l'extremum est bien entendu local. L'étude de la convexité permet donc de connaître la nature d'un extremum. De nombreux problèmes d'optimisation consistent à rechercher ces points. La procédure consiste à d'abord déterminer un ou des points candidats par annulation des dérivées partielles premières puis à en connaître la nature grâce à la matrice Hessienne (maximum, minimum ou point selle).

Exercice 2.2 Démontrer que la fonction $f(x) = x^2$ est une fonction convexe. ■