Programmation Mathématiques

A. Bazeniar

Enseignant chercheur
Centre universitaire Abdelhafid Boussouf
Mila, Algérie

Se reporter à des manuels de base et à certaines livres de spécialités

Janvier 2024

Chapitre 1

Optimisation multidimensionnelle

1.1 introduction

Dans tout le chapitre, on va étudier le problème d'optimisation sans contraintes qui consiste à minimiser la fonction $f : \mathbb{R}^n \to \mathbb{R}$. On considère donc le problème formulé comme suit,

$$(Pr) \begin{cases} minf(x) \\ x \in \mathbb{R}^n \end{cases} \tag{1.1}$$

Définition 1.1.1. *Soit S une partie de $\mathbb{R}^n$ et $B_r(x)$ une boule ouverte centrée en x .*

1. *$x \in \mathbb{R}^n$ est un point intérieur de S s'il existe $r > 0$ tel que $B_r(x) \in S$.*
2. *L'intérieur de X est l'ensemble de tous les points intérieurs de X. On note $intX \subset S$.*
3. *$V(x)$ est un voisinage de x s'il existe $r > 0$ tel que $B_r(x) \subset V(x)$.*
4. *A est un ensemble ouvert de $\mathbb{R}^n$ pour tout $x \in A$, s'il existe un voisinage $V(x)$ de x tel que $V(x) \subset A$.*
5. *$x \in \mathbb{R}^n$ est un point d'accumulation de S si, pour tout $r > 0, B_r(x) \cap S = \varnothing$.*
6. *F est un ensemble fermé s'il contient tous ses points d'accumulation (si son compliment est ouvert).*
7. *E est compacte si et seulement si E est fermée et bornée.*

Théorème 1.1.1. *(Bolzano–Weierstrass). Soit un espace métrique (S, d) et un sous-ensemble E de S. Les propriétés suivantes sont équivalentes.*

1. *E est compact.*
2. *E est séquentiellement compact, c'est-à-dire, toute suite $\{x_n\}$ dans E possède une sous-suite $\{x_{n_k}\}$ qui converge vers un élément $x \in E$.*

Définition 1.1.2. *On dit que $f : \mathbb{R}^n \to \mathbb{R}$ est coercive si*

$$\lim_{\|x\| \to +\infty} f(x) = +\infty \tag{1.2}$$

$\|.\|$ désigne une norme quelconque de $\mathbb{R}^n$.

Exemple 1.1.1. *Soit la fonction $f(x_1, x_2) = (x_1 + x_2)^2$. Alors f n'est pas croissante à l'infini dans $\mathbb{R}^2$. Il suffit de prendre la suite $(n, -n), n \geq 1$,*

$$f(n, -n) = (n - n)^2 = 0 \nrightarrow +\infty.$$

Remarque 1.1.1. *A une matrice symétrique carrée, Alors la fonction $f(x) = x^T Ax + b^T x + c$ est coercive, ssi A est une matrice définie positive.*

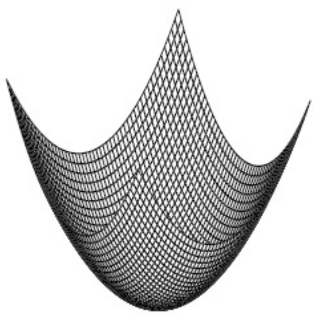


FIGURE 1.1 – Fonction coercive

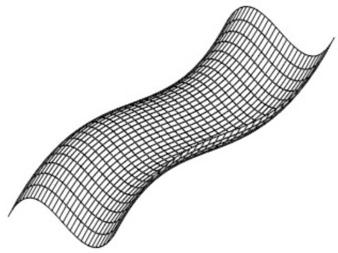


FIGURE 1.2 – Fonction non coercive

Définition 1.1.3. *Le gradient d'une fonction $f : \mathbb{R}^n \to \mathbb{R}$, différentiable est un vecteur de $\mathbb{R}^n$ s'écrivant,*

$$\nabla f(x) = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n})^T, \ \ x \in \mathbb{R}^n$$

Si les dérivées secondes existent et sont continues, alors la matrice Hessienne s'écrit,

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

Exemple 1.1.2. *Soit $f(x) = e^{x_2} - (x_1^2 - 3x_2)^2$. Le gradient et la matrice Hessienne sont :*

$$\nabla f(x) = (-4x_1(x_1^2 - 3x_2), e^{x_2} + 6(x_1^2 - 3x_2))^T,$$

$$\nabla^2 f(x) = \begin{pmatrix} -12x_1^2 + 12x_2 & 12x_1 \\ 12x_1 & e^(x_2) - 18 \end{pmatrix}.$$

1.1.1 Développement de Taylor

Premier ordre :

$$f(y) = f(x) + \nabla f(x)d + o(\| d \|). \tag{1.3}$$

$$f(y) = f(x) + \nabla f(\xi)d \text{ tel que } \xi \in [x, y]. \tag{1.4}$$

Deuxième ordre :

$$f(y) = f(x) + \nabla f(x)d + \frac{1}{2}d^T\nabla^2 f(x)d + o(\| d \|^2). \tag{1.5}$$

$$f(y) = f(x) + \nabla f(x)d + \frac{1}{2}d^T\nabla^2 f(\xi)d \text{ tel que } \xi \in [x, y]. \tag{1.6}$$

Définition 1.1.4. *Pour tout $x \in \mathbb{R}^n$, on appelle la dérivée directionnelle de f en x de direction h,*

$$d_x = \frac{\partial f}{\partial h} =< \nabla f(x), h >; h \in \mathbb{R}^n.$$

Et on note $g(t) = f(x + th)$ tel que, $g(0) = f(x)$.

Définition 1.1.5. *la courbe de niveau k d'une fonction f de deux variables $(x, y) \to f(x, y)$, est l'ensemble des points du plan (x, y) qui vérifient l'équation $f(x, y) = c$.*

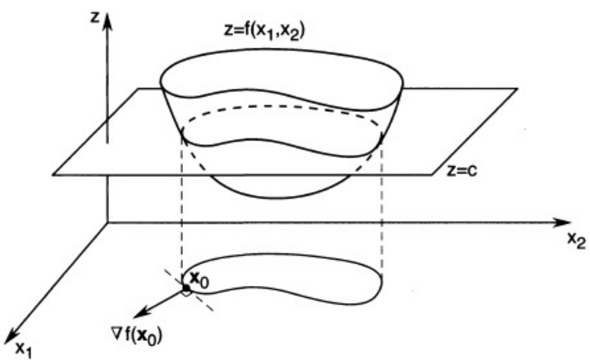


FIGURE 1.3 – courbes de niveau

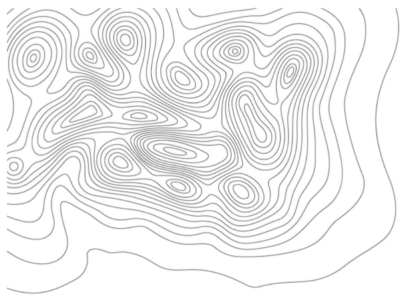


FIGURE 1.4 – Courbes de niveau

Proposition 1.1.1. *Le vecteur gradient $\nabla f(x_0, y_0)$ est orthogonal à la ligne de niveau de f passant au point (x_0, y_0). Il indique la direction de plus forte pente à partir de ce point.*

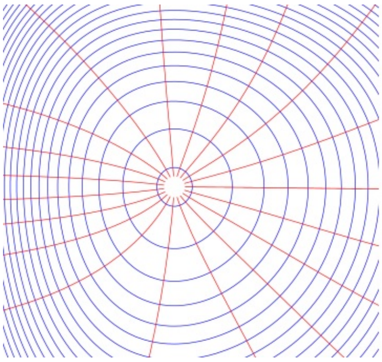


FIGURE 1.5 – Le gradient est orthogonale aux courbes de niveau

1.2 Convexité

Définition 1.2.1. *Un ensemble $S \subseteq \mathbb{R}^n$ est convexe si $x, y \in X \Rightarrow [x, y] \subset X$, c-à-d,*

$$\lambda x + (1-\lambda)y \in S \ \ pour\ tout\ \ x, y \in S et \lambda \in [0,1]. \tag{1.7}$$

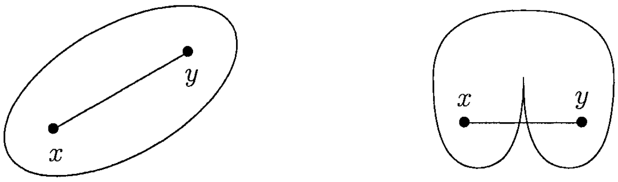


FIGURE 1.6 – Ensemble convexe et non convexe

Définition 1.2.2. *Une fonction $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$ définie sur un ensemble convexe S est convexe sur S ssi*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \ \ \forall x, y \in S \forall \lambda \in (0,1). \tag{1.8}$$

Et dite strictement (fortement) convexe ssi,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{1}{2}c\lambda(1-\lambda) \parallel x - y \parallel^2, \ \ c > 0. \tag{1.9}$$

Théorème 1.2.1. *Soit $S \subset \mathbb{R}^n$ un ensemble convexe ouvert non vide et soit $f : S \subset \mathbb{R}^n \to \mathbb{R}$ soit une fonction différentiable. Alors f est convexe si et seulement si*

$$f(y) \geq f(x) + \nabla f(x)(y - x), \ \ \forall x, y \in S \tag{1.10}$$

Démonstration. 1 :⇒) Soit $f(x)$ une fonction convexe, alors pour tout $0 < \alpha < 1$

$$f(\alpha y + (1-\alpha)x) \leq \alpha f(y) + (1-\alpha)f(x).$$

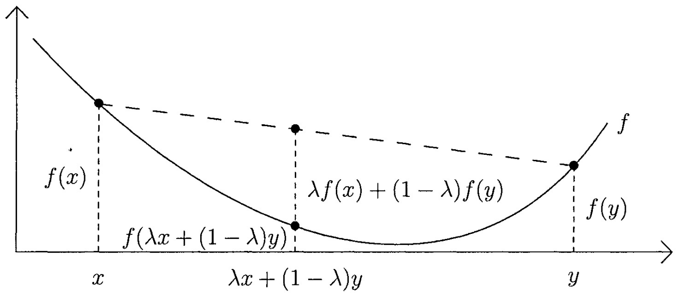


FIGURE 1.7 – Fonction convexe

Donc,

$$\frac{f(x+\alpha(y-x))-f(x)}{\alpha} \leq f(y)-f(x).$$

Posons $\alpha \to 0$, se résulte

$$\nabla f(x)^T(y-x) \leq f(y)-f(x).$$

2 :$\Leftarrow$) Soient $x_1, x_2 \in X$ et $x = \alpha x_1 + (1-\alpha)x_2,\ 0 < \alpha < 1$, alors

$$f(x_1) \geq f(x) + \nabla f(x)^T(x_1 - x),$$

$$f(x_2) \geq f(x) + \nabla f(x)^T(x_2 - x).$$

Donc,

$$\alpha f(x_1) + (1\alpha)f(x_2) \geq f(x) + \nabla f(x)^T(\alpha x_1 + (1-\alpha)x_2 - x) = f(\alpha x_1 + (1-\alpha)x_2),$$

Ce qui fait de f convexe. □

Théorème 1.2.2. *Si f est différentiable sur l'ensemble convexe S, alors les deux propriétés suivantes sont équivalentes,*

$$f(y) \geq f(x) + \nabla f(x)(y-x),\ \ \forall x, y \in S,\ \ \textit{(inégalité du gradient)}$$

$$\langle \nabla f(x) - \nabla f(y), (x-y)\rangle \geq 0,\ \ \forall x, y \in S,\ \ \textit{(monotonie)}$$

Remarque 1.2.1. *Si l'inégalité est stricte (monotonie stricte), la fonction f est strictement convexe. On écrit*

$$(\nabla f(x) - \nabla f(y))(x-y) \geq \alpha \parallel x-y \parallel^2 \forall x, y \in S,\ \ \textit{(monotonie forte)}$$

Théorème 1.2.3. *Une fonction deux fois différentiable est convexe, ssi*

$$d^T \nabla^2 f(x) d \geq 0\ \ \forall x, y \in X.$$

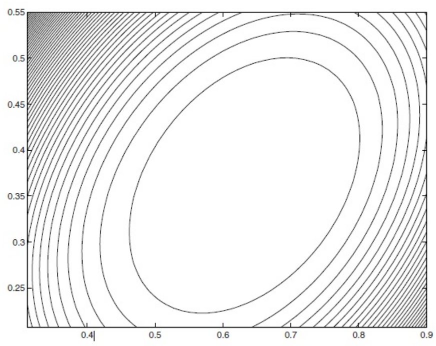


FIGURE 1.8 – courbes de niveau d'une fonction convexe

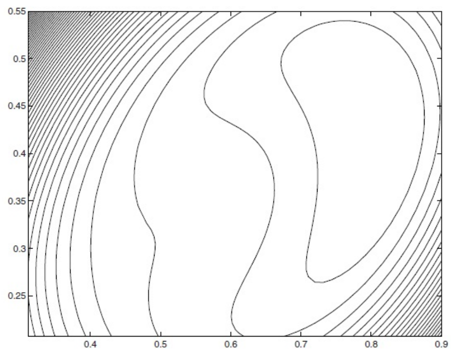


FIGURE 1.9 – Courbes de niveau d'une fonction non convexe

Démonstration. 1 :$\Rightarrow$) Suppose que f est convexe $x \in X$. Dans un voisinage de X,

$$f(x + \lambda d) \geq f(x) + \lambda \nabla f(x)^T d.$$

Et comme f est de classe C^2, alors

$$f(x + \lambda d) = f(x) + \lambda \nabla f(x)^T d + \frac{1}{2}\lambda^2 d^T \nabla^2 f(x) d.$$

Des deux équations au dessus, on trouve

$$\frac{1}{2}\lambda^2 d^T \nabla^2 f(x) d \geq 0.$$

Comme $\lambda \geq 0$, alors

$$d^T \nabla^2 f(x) d \geq 0.$$

2 :$\Leftarrow$) Considerant $x, y \in X$, on a

$$f(y) = f(x) + \lambda \nabla f(x)^T d + \frac{1}{2}\lambda^2 d^T \nabla^2 f(x) d.$$

Mais, $d^T \nabla^2 f(x) d \geq 0$ donc

$$f(y) \geq f(x) + \lambda \nabla f(x)^T d.$$

Ce la fait que f est convexe. □

Définition 1.2.3. *Soit f une fonction définie sur un ensemble S à valeurs dans $\mathbb{R} \cup \{-\infty, +\infty\}$. L'épigraphe de f, notée $epif$, est l'ensemble défini par,*

$$epif = \{(x, \alpha) \in S \times \mathbb{R} \mid f(x) \leq \alpha\}.$$

Autrement dit, l'ensemble des points situés au-dessus du graphe de la fonction dans l'espace $\mathbb{R}^{n+1}$.

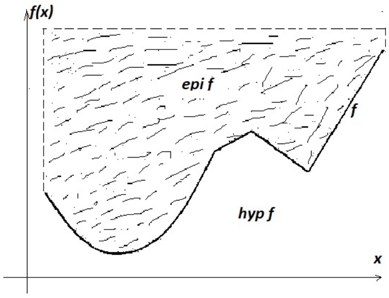


FIGURE 1.10 – L'épigraphe d'une fonction f

Théorème 1.2.4. *Soit $S \subseteq \mathbb{R}^n$ un ensemble convexe non vide. Soit $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$. Alors f est convexe si et seulement si $epif$ est un ensemble convexe.*

Démonstration. Supposons que f soit convexe. Soit $x_1, x_2 \in S$ et $(x_1, \alpha_1), (x_2, \alpha_2)$ sont dans $epif$. Alors, il résulte de la définition (1.2.2) et de la définition (1.2.3) que

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \leq \lambda \alpha_1 + (1-\lambda)\alpha_2.$$

pour tout $\lambda \in (0, 1)$. Puisque S est un ensemble convexe, $\lambda x_1 + (1-\lambda)x_2 \in S$. Donc $(\lambda x_1 + (1-\lambda)x_2, \lambda \alpha_1 + (1-\lambda)\alpha_2) \in epif$, ce qui signifie que $epif$ est convexe.

réciproquement, supposons que $epif$ est convexe, et que $x_1, x_2 \in S$ et $(x_1, f(x_1)), (x_2, f(x_2)) \in epif$. On a alors par la convexité de $epif$ que

$$(\lambda x_1 + (1-\lambda)x_2, \lambda f(x_1) + (1-\lambda) f(x_2)) \in epif, \quad \lambda \in (0, 1).$$

Cela signifie que

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

pour chaque $\lambda \in (0, 1)$. Par conséquent, f est convexe. □

Théorème 1.2.5. *Soit $S \subset \mathbb{R}^n$ un ensemble convexe non vide et $f : S \subset \mathbb{R}^n \to \mathbb{R}$. Soit $x^* \in S$ un minimum de f. Si f est convexe, alors x^* est aussi un minimum global.*

Théorème 1.2.6. *Soit $f : S \subset \mathbb{R}^n \to \mathbb{R}$ une fonction convexe différentiable. Alors x^* est un minimum global si et seulement si $\nabla f(x^*) = 0$.*

Démonstration. 1. $\Rightarrow$) : Soit f une fonction convexe différentiable dans $\mathbb{R}^n$ et $\nabla f(x^*) = 0$, alors

$$f(x) \geq f(x^*) + \nabla f(x^*)(x - x^*) = f(x^*), \ \forall x \in \mathbb{R}^n$$

ce qui indique que x^* est un minimum global de f.

2. $\Leftarrow$) : C'est évident car le minimum global est aussi un minimum local, il est aussi un point stationnaire. □

1.3 Fonction quadratique

Dans cette section, on étudie une classe particulière des fonctions. En général, une fonction quadratique $f : \mathbb{R}^n \to \mathbb{R}$ peut s'écrire comme suit

$$f(x) = x^T Ax + b^T x + c, \tag{1.11}$$

Avec A est une matrice symétrique d'ordre n et b un vecteur de $\mathbb{R}^n$. explicitement, le terme $x^T Ax$ peut s'écrire

$$x^T Ax = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j.$$

Exemple 1.3.1. *Soit* $A = \begin{pmatrix} 2 & -1 \\ -1 & 4 \end{pmatrix}$, *alors* $x^T Ax = 2x_1^2 - 2x_1 x_2 + 4x_2^2$.

Définition 1.3.1. *Soit A une matrice symétrique d'ordre n,*

1. *Si $x^T Ax > 0$ (resp. ≥ 0) alors, A est définie positive et f admet un min local unique (resp. semi-défini),*
2. *Si $x^T Ax < 0$ (resp. ≤ 0), alors A est définie négative et f admet un max local unique (resp. semi-défini),*
3. *Si $x_1^T Ax_2 > 0$ et $x_1^T Ax_2 < 0$ $\forall x_1, x_2 \in \mathbb{R}^n$, alors A est indéfini et f admet un point sell.*

Théorème 1.3.1. *Soit A une matrice symétrique d'ordre n,*

$$A \text{est défini positive} \Leftrightarrow \text{toutes les valeurs propres de } A \text{ sont positives.} \tag{1.12}$$

Remarque 1.3.1. *Les vecteurs propres correspondants au valeurs propres sont orthogonaux aux courbes de niveaux de $f(x)$.*

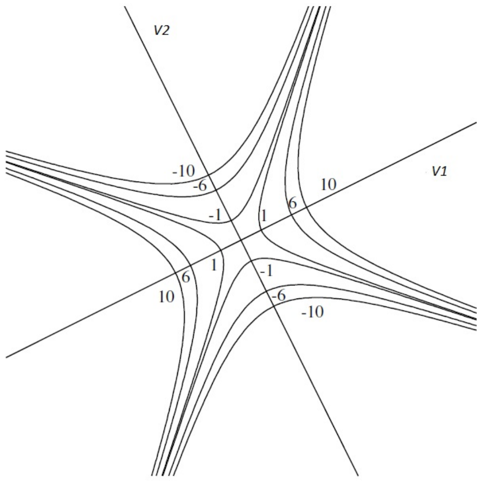


FIGURE 1.11 – Courbes de niveaux de $f = 3x_1^2 - 3x_2^2 + 8x_1x_2$.

Exemple 1.3.2. *Soit $f(x) = 3x_1^2 - 3x_2^2 + 8x_1x_2$.*
Les valeur propres de A sont $\lambda_1 = 5$ et $\lambda_2 = -5$, alors f n'admet pas de minimum.
Les vecteurs qui correspond sont $V_1 = \frac{1}{\sqrt{5}}(2,1)$ et $V_1 = \frac{1}{\sqrt{5}}(1,-2)$. Comme elle montre la figure (1.11),
- *Dans la direction de V_1 les courbes de niveaux sont positives et f admet un min,*
- *Dans la direction de V_2 les courbes de niveaux sont négative et f admet un max,*
- *Dans la direction $(3,1)$ les courbes de niveaux sont constantes.*
- *Donc f admet au point d'intersection de V_1 et V_2 un point sell.*

Remarque 1.3.2. *Si la fonction quadratique contient le terme $b^T x$, alors le centre des courbes de niveaux va se déplacer vers la droite, et la fonction s'écrite*

$$x^* = \frac{-1}{2} A^{-1} b,$$

$$f(x) = x^T A x + b^T x + c = (x - x^*)^T A (x - x^*) + cst,$$

$$cst = c - \frac{1}{4} b^T A b.$$

Exemple 1.3.3. *Si on ajoute $b = (1,1)$ à la fonction $f(x) = 3x_1^2 - 3x_2^2 + 8x_1x_2$, alors le centre des courbe de niveaux devient,*

$$x^* = \frac{-1}{2} A^{-1} b = \frac{-1}{50}(7,1),$$

$$cst = 2, \text{ et } f(x) = \frac{-1}{50}(x-7)A(x-1) + 2.$$

Exercice 1.3.1. *Écrire un programme en Python qui transforme la forme quadratique $f(x) = x^T A x + b^T x + c$ à la forme suivante,*

$$f(x) = (x - x^*)^T A (x - x^*) + cst, \quad x, x^* \in \mathbb{R}^n.$$

1.3.1 Dérivation d'une forme quadratique

1. Le Gradient et le Hessien de la forme $f(x) = x^T Ax + b^T x + c$, sont
$$\nabla f(x) = 2Ax + b \text{ et } \nabla^2 f(x) = 2A.$$
2. Le Gradient et le Hessien de la forme $f(x) = (b^T x)^2$, sont
$$\nabla f(x) = 2b^T xb \text{ et } \nabla^2 f(x) = 2b^T b.$$
3. Le Gradient et le Hessien de la forme $f(x) = (x^T Ax)^2$, sont
$$\nabla f(x) = 4(x^T Ax)Ax \text{ et } \nabla^2 f(x) = 8(x^T Ax).$$
4. Le Gradient et le Hessien de la forme $f(x) = (b^T x)(x^T Ax)^2$, sont
$$\nabla f(x) = (x^T Ax)b + 2(b^T x)Ax \text{ et } \nabla^2 f(x) = 2Ax + 4b^T Ax.$$

Forme hyperbolique : sa figure de courbes de niveau au voisinage de l'origine (point critique) présente un point selle.
Forme d'ellipses : sa figure de courbes de niveau centrées à l'origine.

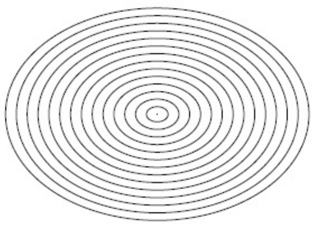


FIGURE 1.12 – Courbes de niveau d'une forme définie

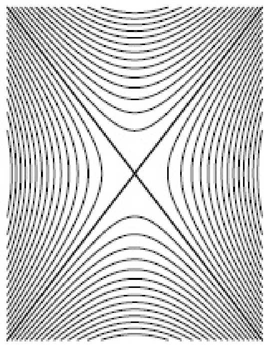


FIGURE 1.13 – Courbes de niveau d'une forme indéfinie

1.4 Existence et unicité

Théorème 1.4.1. *(Existence)*
Soit $f : \mathbb{R}^n \to \mathbb{R}$, continue et coercive. Alors le problème (Pr) admet au moins une solution.

Théorème 1.4.2. *(Unicité)*
Soit $f : \mathbb{R}^n \to \mathbb{R}$ de classe C^1. On suppose qu'il existe $\alpha > 0$ tel que

$$\forall x, y \in \mathbb{R}^n \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \alpha \parallel x - y \parallel^2 . \tag{1.13}$$

Alors f est strictement convexe et coercive et (Pr) admet une solution unique .

Définition 1.4.1. *Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable en $x \in \mathbb{R}^n$. S'il existe un vecteur $d \in \mathbb{R}^n$ tel que*

$$\nabla f(x) d < 0, \tag{1.14}$$

alors d est appelé une direction de descente de f en x.

Utilisant le développement de Taylor,

$$f(x_k + \lambda d) = f(x_k) + t \nabla f(x_k)^T d + o(\lambda),$$

il est alors facile de voir que $\exists \delta > 0$ telle que $f(x_k + \lambda d) < f(x_k), \forall \lambda \in (0, \delta)$ si et seulement si d est une direction de descente de f à x_k.

Théorème 1.4.3. *(conditions nécessaires d'optimalité du premier ordre : Euler)*
Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable sur $\mathbb{R}^n$. Si x^* un minimum (global ou local) de f,alors

$$\nabla f(x^*) = 0 \tag{1.15}$$

Remarque 1.4.1. *Un point x^* vérifiant $\nabla f(x^*) = 0$ est appelé* point critique *ou* point stationnaire.

Démonstration. (Par contradiction). Supposons que $\nabla f(x^*) \neq 0$.
En prenant $d = -\nabla f(x^*)$, on obtient

$$d^T \nabla f(x^*) = -\nabla \parallel f(x^*) \parallel^2 < 0.$$

Ainsi, d est une direction de descente et il existe $\delta > 0$ tel que

$$f(x^* + \alpha d) < f(x^*), \forall \alpha \in (0, \delta),$$

ce qui est contradictoire avec l'hypothèse : x^* est un minimum local. □

Théorème 1.4.4. *(conditions nécessaires d'optimalité du deuxième ordre)*
Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable sur $\mathbb{R}^n$. Si x^* un minimum (resp. maximum) de f,alors

$$\nabla f(x^*) = 0, \;\; \nabla^2 f(x^*) > 0, \;\; (resp. < 0) \tag{1.16}$$

Démonstration. Exercice. □

Théorème 1.4.5. *(conditions suffisante du deuxième ordre)*
Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable sur $\mathbb{R}^n$. Si

$$\nabla f(x^*) = 0, \;\; \nabla^2 f(x^*) > 0, \;\; (resp. < 0),$$

alors x^* est un minimum (resp. maximum) de f.

Démonstration. Supposons que $\nabla f(x^*) = 0$ et que $\nabla^2 f(x^*)$ est définie positive. Par le développement de Taylor, pour tout vecteur $d \in \mathbb{R}^n$ tel que $x^* + d$ se trouve dans un voisinage de x^* dans lequel $\nabla^2 f(x^* + d)$ est définie positive, on a

$$f(x^* + d) = f(x^*) + \frac{1}{2} d^T \nabla^2 f(x^* + td)d, \;\; t \in (0, 1).$$

On peut alors choisir $r > 0$ tel que $x^* + d \in B_r(x^*)$ et $d^T \nabla^2 f(x^* + td)d > 0$. Donc ,

$$f(x^* + d) > f(x^*).$$

□

Exemple 1.4.1. *Considérons la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie comme suit,*

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

On calcul le gradient et le Hessien,

$$\nabla f(x, y) = 0 \Rightarrow \begin{pmatrix} 2x + 2y \\ 2x + 2y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

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

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

Les valeurs propre de $\nabla^2 f$ sont tous positifs $(\lambda_1 = 0, \lambda_2 = 4))$, alors $(0, 0)$ est un min local.

1.4.1 Critère de Sylvester

Le critère de Sylvester et les mineurs principaux sont en analyse matricielle pour déterminer les propriétés des matrices symétriques et leur définitivité.

Définition 1.4.2. *Soit A, une matrice symétrique réelle de taille n. On appelle* ***mineurs principaux successifs*** *les déterminants des n matrices $A_p = (a_{ij})_{(1\leq i,j\leq p)}$, pour $p = 1, \ldots, n$.*

Théorème 1.4.6. *(Critère de Sylvester)*

1. *Une forme quadratique A est définie positive (semi-définie positive) si et seulement si tous les mineurs principaux successifs de la matrice A sont positifs (≥ 0).*
2. *Une forme quadratique A est définie négative (semi-définie négative) si et seulement si tous les mineurs principaux successifs de la matrice A alternent en signe, le premier étant négatif (≤ 0).*
3. ***Sinon**. La forme quadratique change de signe.*

Exemple 1.4.2. *Considérons la matrice symétrique suivante :*

$$A = \begin{pmatrix} 4 & 2 & -1 \\ 2 & 5 & 3 \\ -1 & 3 & 6 \end{pmatrix}.$$

On calcul les mineurs principaux :

$$d_1 = |4| = 4,\ d_2 = det\begin{pmatrix} 4 & 2 \\ 2 & 5 \end{pmatrix} = 12 \ et\ d_3 = det\begin{pmatrix} 4 & 2 & -1 \\ 2 & 5 & 3 \\ -1 & 3 & 6 \end{pmatrix} = 113$$

Puisque tous les mineurs principaux de la matrice sont strictement positifs, selon le critère de Sylvester, la matrice A est définie positive.

Définition 1.4.3. *(Convergence d'un algorithme) On dit que l'algorithme Aconverge si la suite $(x_k)_{k\in\mathbb{N}}$ engendrée par l'algorithme converge vers une limite x^*.*

1.4.2 Taux de convergence d'un algorithme

Le taux de convergence mesure la rapidité à laquelle l'algorithme, suivant une suite $(x_k)_{k\in\mathbb{N}}$, converge vers la solution optimale x^*.
On dit que la convergence est ($e_k = \| x_k - x^* \|$),

— Linéaire si la distance entre la solution x^k et la solution optimale diminue d'une quantité constante à chaque itération :

$$\exists C \in [0,1[,\ \ \exists k_0,\ \ \forall k \geq k_0,\ \ e_{k+1} \leq Ce_k, \quad (C \text{ est une constante}).$$

— Super-linéaire si la distance entre la solution actuelle et la solution optimale diminue plus rapidement qu'une convergence linéaire :

$$e_{k+1} \leq \alpha_k ek,$$

où α_k est une suite positive convergente vers 0. Si α_k est une suite géométrique, la convergence de l'algorithme est dite géométrique.

— D'ordre p si l'erreur e_k décroît de la manière suivante :

$$\exists C \in [0,1[,\ \ \exists k_0,\ \ \forall k \geq k_0,\ \ e_{k+1} \leq C \parallel e_k \parallel^p .$$

Si $p = 2$, la convergence de l'algorithme est dite quadratique.

Remarque 1.4.2. *Parfois, le taux de convergence n'est pas mesuré en termes de distance entre la solution x^k et la solution optimale, mais en termes du nombre d'itérations nécessaires pour atteindre une certaine précision. Ce taux peut dépendre de,*

— *Choix de la méthode d'optimisation,*

— *Paramètres de l'algorithme,*

— *La fonction objectif,*

— *La condition initiale.*

1.5 Méthodes de résolution

Soient x_k, d_k la direction et le pas α_k. La k-ième itération est alors

$$x_{k+1} = x^k + \alpha_k d_k. \tag{1.17}$$

Les différentes longueurs de pas α_k et les différentes directions d_k construisent les différentes méthodes.

Critère d'arrêt : En général les critère d'arrêt les plus utilisées sont

$$\|\nabla f(x^k)\| \leq \varepsilon.$$

$$\|x^{k+1} - x^k\| \leq \varepsilon.$$

$$|f(x^{k+1}) - f(x^k)| \leq \varepsilon.$$

Dans le premier teste l'algorithme converge moins rapidement, alors que dans les autre converge rapidement.

Cette section traite des différentes méthodes de résolution du problème de minimisation multidimensionnelle.

1.5.1 Méthode de descente de gradient

L'une des méthodes les plus anciennes et les plus connues pour minimiser une fonction de plusieurs variables est la méthode de descente de gradient. Le principe général de la méthode est basée sur le procédé suivant.

D'une part, On sait auparavant que les itérations sont générées par,

$$x = x_k + \alpha_k d_k.$$

Cela fait que,

$$x - x^k = \alpha_k d_k.$$

D'autre part, l'approximation de la fonction déffirentiable f au voisinage de x^k est le développement de Taylor,

$$f(x) = f(x^k) + (x - x^k)\nabla f(x^k) + o(\|x - x^k\|).$$

Par consequent,

$$f(x) = f(x^k) + \alpha_k d_k \nabla f(x^k) + o(\|x - x^k\|).$$

Cependant, le problème est de minimisation, ce qui fait $f(x) - f(x^k) < 0$, alors d_k est une direction de descente.
Finalement, pour un pas α_k fixe, on prend

$$d_k = -\nabla f(x^k).$$

Les itérations de la méthode de descente de gradient s'effectuent par le principe suivant,

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k), \tag{1.18}$$

Le pas α_k est un scalaire non négatif qui minimise $f(x^k - \alpha \nabla f(x^k))$. En d'autres termes, nous cherchons la suite (x^k) le long de la direction du gradient $-\nabla f(x^k)$ jusqu'à le point minimum.

Proposition 1.5.1. *Si α^* est le pas optimal, alors*

$$-\nabla f(x^k)\nabla f(x^k - \alpha_k \nabla f(x^k)) = 0$$

Démonstration. En exercice. □

Remarque 1.5.1. *Au point x^{k+1}, le gradient de f est orthogonal aux courbes de niveau de f, donc on peut en quelque sorte dire que la direction d_k est « tangente » à l'hypersurface de niveau de f passant par x^{k+1}.*

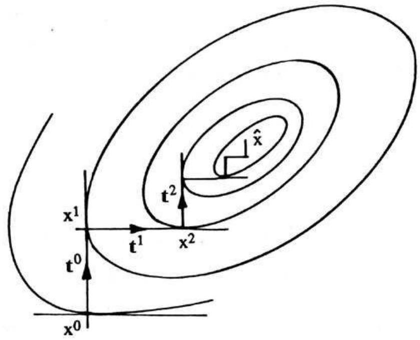


FIGURE 1.14 – "zigzag" de la méthode du gradient à pas optimal.

Théorème 1.5.1. *(Convergence cas : fonctions quadratiques)*
Soit A symétrique et définie positive et x^* une solution du problème quadratique,

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle.$$

Soit λ_1 et λ_n les valeurs propres les plus grandes et les plus petites de A respectivement. La convergence est au moins linéaire, et les limites sont

$$\frac{f(x^{k+1}) - f(x^*)}{f(x^k) - f(x^*)} \leq \frac{(\lambda_1 - \lambda_n)^2}{(\lambda_1 + \lambda_n)^2}. \tag{1.19}$$

$$\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} \leq \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}. \tag{1.20}$$

Dans ce cas, la formule du pas optimal est donnée par,

$$\alpha^* = \frac{2}{\lambda_1 + \lambda_n}.$$

Proposition 1.5.2. *(Convergence cas : fonctions non quadratiques).*
Soit $f(x)$ une fonction deux fois continûment différentiable et bornée. Soit également sa matrice hessienne bornée et pour tout $x \in \mathbb{R}^n$,

$$x^T \nabla^2 f(x) x \leq \frac{M-m}{M} ||x||^2.$$

Alors, la suite $\{x^k\}$ converge et vérifie,

$$x^{k+1} = x^k - \frac{M-m}{M} \nabla f(x^k).$$

Avec M et m satisfont,

$$0 < m \leq \lambda_n \leq \lambda_1 \leq M.$$

L'inconvénient majeur de cette méthode réside dans la détermination de la constante M. Même si on peut déterminer le pas $\alpha = M^{-1}(M-m)$, la convergence serait lente car le pas n'est pas parfait.

Définition 1.5.1. *(Critères de Armijo : recherche linéaire)*
Dû à Armijo (1966). La condition de Armijo ou condition de décroissance linéaire est donnée par,

$$f(x + \alpha d) \leq f(x) + c\alpha (\nabla f(x)^t d), \quad 0 < c < 1. \tag{1.21}$$

En pratique la constante c est choisie très petite, typiquement : $c = 10^{-4}$.

Algorithme de descente de gradient dans $\mathbb{R}^n$

1. Initialisation
   $k = 0$: choix de $x^0 \in \mathbb{R}^n$ dans un voisinage de x^*.
2. Itération k

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k).$$

3. Trouver le pas α_k

$$f(x_k - \alpha_k \nabla f(x^k)) = \min_{\alpha \geq 0} f(x_k - \alpha \nabla f(x^k)).$$

4. Critère d'arrêt
   Si $\| \nabla f(x^k) \| < \varepsilon$, stop
   Sinon, on pose $k = k + 1$ et on retourne à 2.

1.5.2 Méthode de Newton

La méthode de Newton pour l'optimisation dans $\mathbb{R}^n$ est une extension de la méthode de Newton pour l'optimisation unidimensionnelle à des problèmes multivariés. La méthode utilise des pas optimaux dans la direction de descente pour minimiser une fonction sur $\mathbb{R}^n$ en se basant sur le gradient et la hessienne de la fonction.

Algorithme de Newton dans $\mathbb{R}^n$

1. Initialisation
   $k = 0$: choix de $x^0 \in \mathbb{R}^n$ dans un voisinage de x^*.
2. Itération k

$$x^{k+1} = x^k - [\nabla^2 f(x^k)]^{-1} \nabla f(x^k).$$

3. Critère d'arrêt
   Si $\| \nabla f(x^k) \| < \varepsilon$, Stop
   Sinon, on pose $k = k + 1$ et on retourne à 2.

Les propriétés remarquables de cet algorithme sont,

1. Sa convergence quadratique (le nombre de décimales exactes est multiplié par 2 à chaque itération).
2. les difficultés et le coût de calcul de la hessienne $\nabla^2 f(x^k)$: l'expression analytique des dérivées secondes est rarement disponible dans les applications.
3. le coût de résolution du système linéaire $\nabla^2 f(x^k)(x^{k+1} - x^k) = \nabla f(x^k)$.
4. l'absence de convergence si le premier itéré est trop loin de la solution, ou si la hessienne est singulière.
5. Pas de distinction entre minima, maxima et points stationnaires.

Théorème 1.5.2. *(Convergence de la Méthode de Newton)*
Soit $f \in C^2$ et supposons que x^k soit suffisamment proche de la solution x^* du problème de minimisation où $\nabla f(x^*) = 0$. Si la hessienne $\nabla^2 f(x^*)$ est définie positive et que $\nabla^2 f(x^*)$ satisfait la condition de Lipschitz :

$$\|\nabla^2 f(x^{k+1}) - \nabla^2 f(x^k)\| \leq L||x^{k+1} - x^k||, \quad \textit{pour un certain } L. \tag{1.22}$$

Alors, la suite générée $\{x_k\}$ converge vers x^* avec un taux quadratique.

Exemple 1.5.1. *On étudie la convergence de la méthode de Newton pour la fonction*

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

entre les points $(1, 1)$ *et* $(3, 3)$

Après des calculs, on a

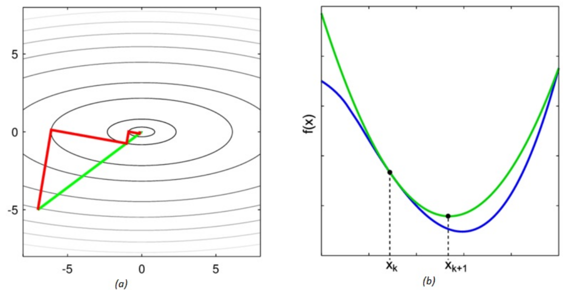
$$\nabla^2 f(1,1) = \begin{pmatrix} 6 & 3 \\ 3 & 6 \end{pmatrix} \quad et \quad \nabla^2 f(3,3) = \begin{pmatrix} 18 & 3 \\ 3 & 18 \end{pmatrix}$$

On obtient,

$$\frac{\|\nabla^2 f((1,1)) - \nabla^2 f(3,3)\|}{\|(1,1) - ((3,3)\|} = 6.$$

Donc, la méthode de Newton devrait converger vers un minimum local dans la région entourant ces points.

1.5.2.1 Comparaison des deux méthodes



La figure (a) compare la convergence de la méthode de descente de gradient (ligne rouge) avec la méthode de Newton (ligne verte) sur l'ellipsoïde quadratique $f(x,y) = 1/2x^2 + 2y^2$. L'optimisation démarre en $(-7,-5)$. Dans cet exemple, la méthode de descente de gradient nécessite théoriquement un nombre infini d'itérations pour atteindre l'optimum en $(0,0)$. En pratique, la solution est trouvée après 5 itérations. En revanche, l'étape de Newton conduit instantanément à la solution exacte.
La figure (b) illustre comment le prochain itéré est trouvé dans la méthode de Newton : au point courant x^k, un modèle quadratique (vert) de la fonction objectif (bleu) est construit. La position de son minimum définit le prochain itéré x^{k+1}.