

Chapitre I : Rappels mathématiques

I.1 Introduction

Étudier un problème d'optimisation sans contraintes veut dire effectuer une minimisation ou une maximisation d'une fonction f définie de $\mathbb{R}^n \rightarrow \mathbb{R}$ sur un ensemble déterminé. Les problèmes d'optimisation (avec ou sans contraintes) considérés dans ce document s'écrivent sous la forme standard suivante :

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{Ou} \quad \begin{cases} \min_{x \in \mathbb{R}^n} f(x) \\ \text{sc} \quad x \in E \end{cases}$$

Où $x = (x_1, x_2, \dots, x_n)^T$ est un vecteur de $\mathbb{R}^n$, $f: \mathbb{R}^n \rightarrow \mathbb{R}$ est la fonction que l'on désire minimiser (appelée fonction objectif), $E \subset \mathbb{R}^n$ est l'ensemble dans lequel les points doivent appartenir, et sc est l'abréviation de sous la ou les contraintes.

La formulation ci-dessus signifie que l'on cherche à trouver une solution du domaine réalisable ($x^* \in E$) dont la valeur de la fonction objectif est la plus petite.

I.2 Différents types d'optimisation

❖ optimisation linéaire

f est une fonction linéaire : $f(x) = \langle c, x \rangle$ E est défini par des fonctions affines : $ax + b \leq 0$

❖ optimisation linéaire quadratique

f est une fonction convexe quadratique : $f(x) = x^T A x$

A est une matrice symétrique semi-définie positive E est défini par des fonctions affines : $ax + b \geq 0$

❖ optimisation convexe

f est une fonction convexe et E un domaine convexe

❖ optimisation différentiable

f est une fonction différentiable, E est défini par des fonction (= contraintes) différentiables

❖ optimisation non différentiable

Ex : $f(x) = \max \{f_1(x), \dots, f_m(x)\}$

❖ optimisation en dimension infinie

Ex: problèmes variationnels

$J(x) = \int_0^T L(t, x(t), \dot{x}(t)) dt$ Avec x dans un ensemble de fonctions X , $x(0) = x_0$ et $x(T) = x_T$

❖ optimisation discrète

X est un ensemble fini ou dénombrable

❖ optimisation stochastique

X est un ensemble aléatoire

❖ optimisation multicritère

On veut optimiser plusieurs fonctions objectives en même temps (notion d'optimum de Pareto)

I.3 Positivité

I.3.1 Matrice définie positive

En algèbre linéaire, la notion de matrice symétrique définie positive (négative) est analogue à celle de nombre réel strictement positif (strictement négatif). Une matrice définie positive (négative) est une matrice positive (négative) inversible.

(Matrice définie positive) : $A = A^T > 0 \Leftrightarrow x^T A x > 0 \quad \forall x \neq 0$ Ou $\forall x \neq 0, (Ax, x) > 0$

(Matrice définie négative) : $A = A^T < 0 \Leftrightarrow x^T A x < 0 \quad \forall x \neq 0$ Ou $\forall x \neq 0, (Ax, x) < 0$

I.3.2 Propriétés des matrices définies positives

Si $A = [a_{ij}]$ est une matrice de taille $n \times m$ alors ses sous-matrices principales sont :

$$\begin{aligned} A_1 &= [a_{11}] \\ A_2 &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\ A_3 &= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \\ &\vdots \\ A_n &= \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \quad \text{et} \quad A_n = A. \end{aligned}$$

$i = 1, \dots, n$ et $j = 1, \dots, m$

Une matrice **symétrique A est définie positive** si toutes ses valeurs propres sont positives.

Une matrice **symétrique A est définie négative** si toutes ses valeurs propres sont négatives.

Pour une matrice symétrique A de taille $n \times n$, les énoncés suivants sont équivalents :

- Les n pivots de A sont positifs.
- Les n déterminants des sous-matrices principales de A sont positifs.
- Les n valeurs propres de A sont strictement positifs.
- $x^T A x > 0$ pour tout vecteur $x \neq 0$.
- $A = L^T L$, ou L a des colonnes linéairement indépendantes.

$$A = A^T > 0 \Leftrightarrow \begin{cases} 1- \lambda_i\{A\} > 0 \\ 2- \det\{A_k\} > 0 \\ 3- A = L^T L \quad \text{Par factorisation de Cholesky avec L inversible et triangulaire} \end{cases}$$

I.3.3 Matrices semi-définies positives

(Matrice définie semi-positive) : $A = A^T \geq 0 \Leftrightarrow x^T A x \geq 0 \quad \forall x$ Ou $\forall x, (Ax, x) \geq 0$

(Matrice définie semi-négative) : $A = A^T \leq 0 \Leftrightarrow x^T A x \leq 0 \quad \forall x$ Ou $\forall x, (Ax, x) \leq 0$

I.4 Convexité

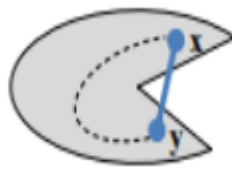
I.4.1 Ensemble convexe

$E \subset \mathcal{R}^n$ est **convexe** si et seulement si :

$$\forall (x, y) \in E, \forall \lambda \in [0, 1], \quad \lambda x + (1 - \lambda)y \in E$$



Non convexe



Non convexe



Convexe

Figure 1 : Ensemble convexe

Exemple :

$E = \{(x_1, x_2) / x_1^2 + x_2^2 \leq 1\} \rightarrow \text{Convexe}$, $E = \{(x_1, x_2) / x_1^2 + x_2^2 \geq 1\} \rightarrow \text{Non convexe}$

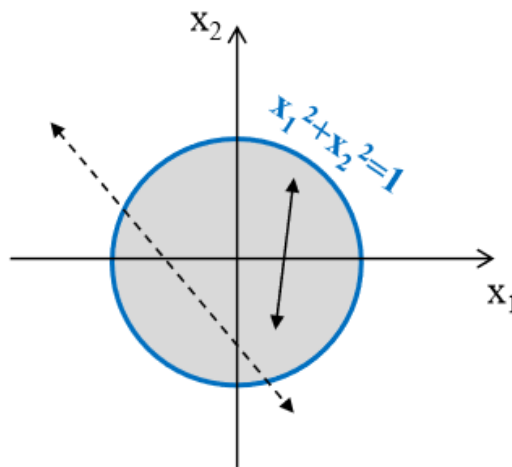


Figure 2 : Exemple d'un ensemble convexe

I.4.2 Propriétés des ensembles convexes

1- Si E est convexe et $\alpha \in \mathbb{R}$ alors $\Omega = \{x \mid x = \alpha \cdot u, u \in E\}$ est convexe

2- Si E_1 et E_2 sont convexes alors

$\Omega = E_1 \cap E_2$ est convexe et $\Omega = \{x \mid x = u_1 + u_2, u_1 \in E_1 \text{ \& } u_2 \in E_2\}$ est convexe

I.4.3 Fonction convexe

a- Définitions

La fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est convexe si et seulement si :

$$\forall (x, y) \in E, \quad \forall \lambda \in [0, 1], \quad f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

f est strictement convexe si et seulement si :

$$\forall (x, y) \in E, \quad x \neq y, \quad \forall \lambda \in]0, 1[, \quad f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y)$$

Toutes fonctions convexes sur $\mathbb{R}$ sont continues

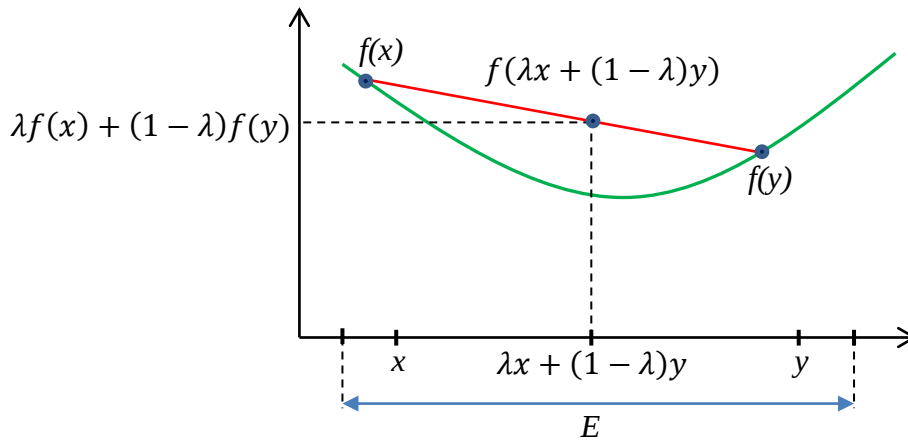


Figure 3 : Fonction convexe

(“ f est en-dessous de ses cordes”).

Si f est de plus différentiable,

$$f \text{ convexe} \Leftrightarrow \forall (x, y) \in \mathbb{R}^n, f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

$$f \text{ strictement convexe} \Leftrightarrow \forall (x, y) \in \mathbb{R}^n, \quad x \neq y, \quad f(y) > f(x) + \nabla f(x)^T (y - x)$$

(“ f est au-dessus de ses tangentes”).

Si f est deux fois différentiable,

$$f \text{ convexe} \Leftrightarrow \forall x \in \mathbb{R}^n, Hf(x) \geq 0, \text{ C'est-à-dire } \forall d \in \mathbb{R}^n, d^T Hf(x) d \geq 0$$

$$f \text{ strictement convexe} \Leftrightarrow \forall x \in \mathbb{R}^n, Hf(x) > 0, \text{ C'est-à-dire } \forall d \neq 0, d^T Hf(x) d > 0$$

Si f est **convexe**, alors f est **concave**.

Exemple : $f(x) = x^2$ donc $f''(x) = 2 \rightarrow$ convexe sur $\mathbb{R}$

$f(x) = x^3$ donc $f''(x) = 6x \rightarrow$ convexe sur $\mathbb{R}^+$ mais non convexe sur $\mathbb{R}$

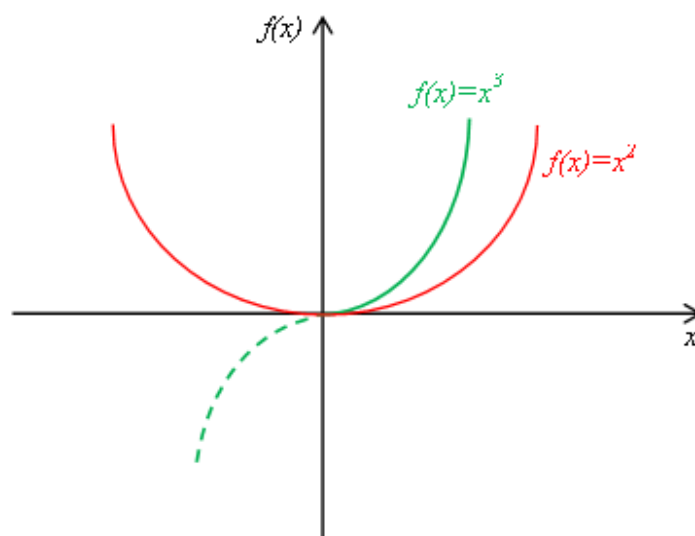


Figure 4 : Fonction convexe et autre non convexe

b- Fonctions coécrites

Soit $\Omega \subset \mathbb{R}^n$ un ensemble non borné et $f: \Omega \rightarrow \mathbb{R}$. On dit que f est coécrive sur Ω si on a

$$\lim_{x \in \Omega, \|x\| \rightarrow +\infty} f(x) = +\infty$$

I.5 Minimum d'une fonction

I.5.1 Définition : Minimum local/Minimum global

Les minima locaux et globaux de f sur $\mathbb{R}^n$ sont définis de la manière suivante :

Soit $f: E \subset \mathbb{R}^n \rightarrow \mathbb{R}$ une fonctionnelle. $x^* \in E$ est un point de minimum local de f sur E si

$$\exists \epsilon > 0 \text{ tel que } f(x^*) < f(x), \forall x \in \mathbb{R}^n, \text{ avec } \|x - x^*\| < \epsilon$$

Un vecteur $x^* \in \mathbb{R}^n$ est un minimum global de f sur $\mathbb{R}^n$ si

$$f(x^*) \leq f(x), \forall x \in \mathbb{R}^n$$

Remarque 1

Les maxima locaux et globaux sont définis de manière similaire. Notons que x^* est un maximum local (respectivement global) de la fonction f sur l'ensemble $\mathbb{R}^n$ si x^* est un minimum local (respectivement global) de la fonction $(-f)$ sur $\mathbb{R}^n$. Il résulte de cette observation que tout problème de maximisation peut être réduit immédiatement à un problème de minimisation (et inversement) en multipliant la fonction objectif par -1.

Le problème d'optimisation peut admettre une infinité de solutions optimales, comme dans la figure suivante :

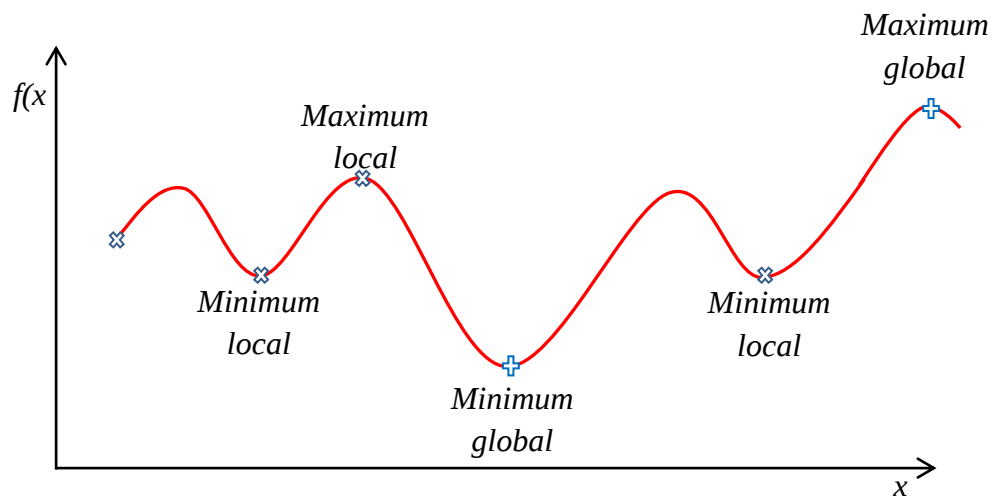


Figure 5 : Minima locaux et globaux

Remarque 2 : Minima des fonctions convexes

Dans le cas d'une fonction objectif convexe, il n'y a pas de distinction entre minimum local et global : tout minimum local est également global, comme l'établit le théorème suivant :

Soit une fonction $f: \Omega \rightarrow \mathbb{R}$ $\Omega \subset \mathbb{R}^n$

Si f est convexe sur Ω alors tout minimum local (au sens strict) est un minimum global (au sens strict)

I.6 Le gradient

Dans une fonction à plusieurs variables, le gradient est un vecteur qui pointe dans la direction selon laquelle la fonction augmente le plus rapidement. Par construction, le vecteur gradient est perpendiculaire à la courbe isovaleur de la fonction au point considéré. Les coordonnées du vecteur gradient associées à une fonction sont les dérivées partielles de cette fonction en ce point. Ainsi pour, $f(x_i)$ les coordonnées du vecteur gradient en x_i , sont :

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \dots \quad \frac{\partial f}{\partial x_n} \right]^T$$

I.7 Le Hessien

On définit la matrice Hessienne, ou Hessien, comme la matrice contenant l'ensemble des dérivées secondes d'une fonction. La matrice Hessienne est par construction une matrice carrée symétrique ; par exemple pour une fonction à n variables, x_i :

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad H(x) = \nabla(\nabla^T f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

a- Dérivée directionnelle

La dérivée directionnelle de f en $x \in \mathbb{R}^n$ dans la direction unitaire $d \in \mathbb{R}^n$ est :

$$f'_d(x) = \frac{\partial f}{\partial d}(x) = \lim_{h \rightarrow 0^+} \frac{f(x + hd) - f(x)}{h} = \nabla^T f(x) d = d \nabla f(x)$$

Si $\|d\| = 1$, la dérivée directionnelle est le **taux d'accroissement** de f dans la direction d au point x

Le taux d'accroissement est **maximal** dans la direction du gradient

Le gradient indique la direction de **plus grande pente**

b- Conditions nécessaires d'existence de minimum

Condition nécessaire du 1^{er} ordre

Soit $f : E \rightarrow \mathbb{R}$, $E \subset \mathbb{R}^n$, $f \in C^1$

Si f a un minimum local en $x^* \in E$.

Alors $\forall d$ admissible en x^*

$d^T \nabla f(x^*) \geq 0 \Leftrightarrow$ Le taux d'accroissement est positif.

Si x^* minimise f localement et si x^* est un point intérieur de E , alors : $\nabla f(x^*) = 0$

Condition nécessaire du 2^{ème} ordre

Soit $f : E \rightarrow \mathbb{R}$, $E \subset \mathbb{R}^n$, $f \in C^2$

Si f atteint un minimum local en $x^* \in E$ et d est une direction admissible

Si $d^T \nabla f(x^*) = 0$ alors $d^T H(x^*) d \geq 0$

Si x^* un point intérieur de E et x^* minimise f

Alors : $\nabla f(x^*) = 0$ et $H(x^*) \geq 0$

c- Conditions suffisantes pour un minimum

Soit $f : E \rightarrow \mathbb{R}$, $E \subset \mathbb{R}^n$, $f \in C^2$ et x^* est un point intérieur de E

Si : $\nabla f(x^*) = 0$ et $H(x^*) > 0$ alors f a un minimum local au sens strict en x^*

I.8. Les points critiques ou stationnaires

Soit $M_0(x_0, y_0)$ un point stationnaire d'une fonction f de classe C^2 sur un ouvert E Posons :

$$\det(A) = \det[H(x_0, y_0)] = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2}(x_0, y_0) & \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) \\ \frac{\partial^2 f}{\partial y \partial x}(x_0, y_0) & \frac{\partial^2 f}{\partial y^2}(x_0, y_0) \end{bmatrix} = \begin{bmatrix} r & s \\ s & t \end{bmatrix}$$

Si $rt - s^2 < 0$ alors f admet un point selle (de cheval) ou col (d'une montagne) en M_0

Si $rt - s^2 > 0$ et $r > 0$ alors f admet un minimum local en M_0

Si $rt - s^2 > 0$ et $r < 0$ alors f admet un maximum local en M_0

Si $rt - s^2 = 0$ alors on ne peut pas conclure par cette méthode

Cas des fonctions convexes

Soit $f \in C^2$, f convexe, $f : E \rightarrow \mathbb{R}$, $E \subset \mathbb{R}^n$ convexe

Alors f a un minimum global en $x^* \in E$

* ssi $\nabla f^T(x^*)(y - x^*) \geq 0 \quad \forall y \in E \Leftrightarrow \nabla f^T(x^*)d \geq 0 \quad \forall$ la direction admissible d

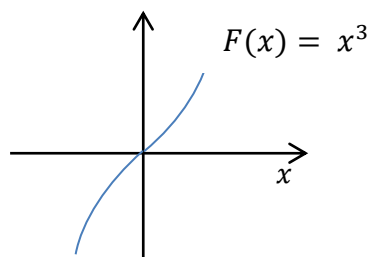
* ssi $\nabla f(x^*) = 0$ si x^* est un point intérieur de E

Exemples

$$1^\circ f(x) = x^3$$

$$f'(0) = 0$$

$$f''(0) = 0$$



Les conditions nécessaires sont vérifiées, mais $x = 0$ ne minimise pas f

$$2^{\circ} f(x, y) = x^2 - y^2$$

$$\nabla f(x, y) = \begin{bmatrix} 2x \\ -2y \end{bmatrix} \quad \nabla f(x, y) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{en } (x, y) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$H(x, y) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$ $\det(H) = -4 < 0$ donc Les conditions nécessaires du 2^{ème} ordre ne sont pas vérifiées $\Rightarrow$ point selle comme la montre la figure suivante :

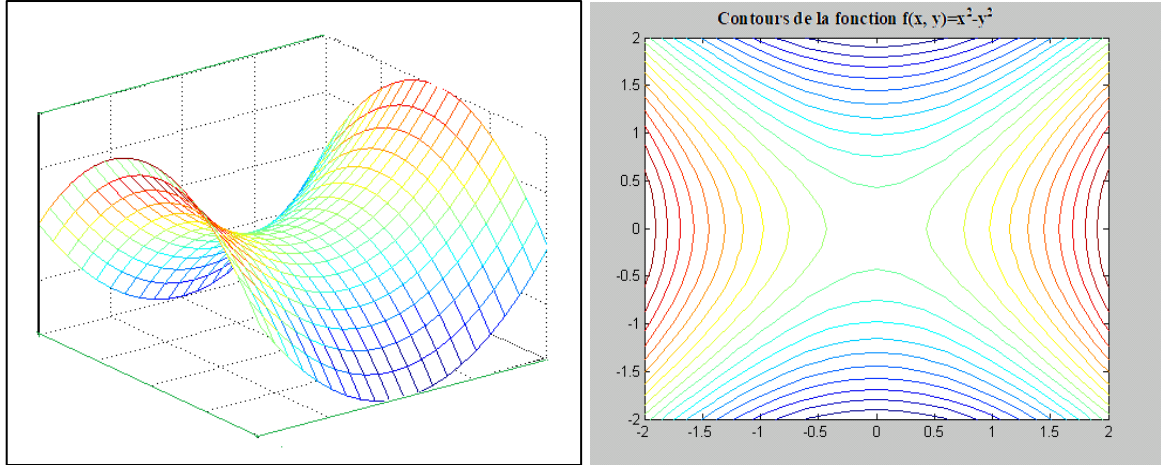


Figure 6 : Point selle ou col

$$3^{\circ} f(x, y) = x^2 + y^2$$

$$\nabla f(x, y) = \begin{bmatrix} 2x \\ 2y \end{bmatrix} \quad \nabla f(x, y) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{en } (x, y) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$H(x, y) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad \det(H) = 4 > 0 \quad \text{donc le point } (x, y) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ est minimum local au sens strict}$$

D'où :

$f(x, y)$ est convexe $\Rightarrow$ minimum global au sens strict comme la montre la figure ci-dessous

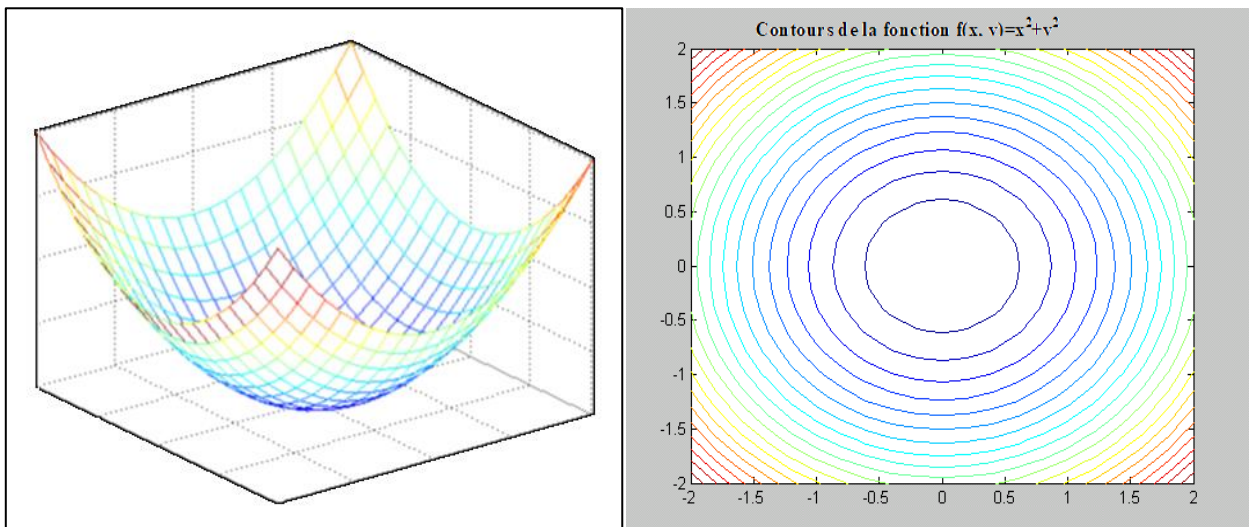


Figure 7 : Minimum local

Forme quadratique

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ donnée par :

$$f(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle + c \text{ ou } f(x) = \frac{1}{2} x^T A x - b^T x + c \quad \forall x \in \mathbb{R}^n \text{ et } n \in \mathbb{N}$$

Avec A une matrice carrée réelle et symétrique de taille n , et $b \in \mathbb{R}^n$ un vecteur et $c \in \mathbb{R}$ un scalaire On calcule facilement :

$$\nabla f(x) = Ax - b \quad \text{et} \quad \nabla^2 f(x) = H(x) = A$$

- f est une fonction convexe $\Leftrightarrow A$ est une matrice semi-définie positive
- f est fortement convexe $\Leftrightarrow f$ est strictement convexe $\Leftrightarrow A$ est une matrice définie positive (matrice SDP).