Chapitre I

Rappels Mathématiques

1.1 Rappels sur le Calcul Différentiel

1.1.1Gradient et Hessienne

Définition 1.1.1 Soit Ω un ouvert de R^n et $f: \Omega \to$ R et $\{e_1, e_2, \cdots, e_n\}$ la base canonique de R^n (toutes les composantes de e_i sont nulles sauf sa $i^{ème}$composante est égale à 1). On appelle dérivée partielle de f par rapport à sa $i^{ème}$ variable x_i la limite, si elle existe, du rapport

$$\frac{f(x+te_i)-f(x)}{t}$$, lorsque t tend vers 0.

On la note $\frac{\partial f(x)}{\partial x_i}$.

Exemple 1.1.1

Soit $f(x) = f(x_1, x_2) = 2x_1 + x_2$ alors

$$\frac{f(x + te_1) - f(x)}{t} = \frac{2(x_1 + t) + x_2 - (2x_1 + x_2)}{t} = \frac{2t}{t} = 2$$

Donc

$$\lim_{t \to 0} \frac{f(x + te_1) - f(x)}{t} = 2$$

De même on a:

$$\frac{f(x + te_2) - f(x)}{t} = \frac{2x_1 + (x_2 + t) - 2x_1 - x_2}{t} = \frac{t}{t} = 1$$

Donc

$$\lim_{t \to 0} \frac{f(x + te_1) - f(x)}{t} = 1$$

Définition 1.1.2 Soit Ω un ouvert de R^n et $f: \Omega \to$ R. Le vecteur dont la $i^{ème}$composante est $\frac{\partial f(x)}{\partial x_i}$ est appelé gradient de f en x.

On le note $gradf(x)$, ou $\nabla f(x)$.

Exemple 1.1.2

Soit $f(x) = f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$

$\frac{\partial f(x)}{\partial x_1} = \frac{\partial}{\partial x_1}(100(x_2 - x_1^2)^2 + (1 - x_1)^2)$

$$\frac{\partial f(x)}{\partial x_1} = (200(x_2 - x_1^2)(-2x_1) - 2(1 - x_1))$$

$\frac{\partial f(x)}{\partial x_1} = -400(x_2 - x_1^2)x_1 - 2(1 - x_1)$ (1)

$$\frac{\partial f(x)}{\partial x_2} = \frac{\partial}{\partial x_2}(100(x_2 - x_1^2)^2 + (1 - x_1)^2)$$

$\frac{\partial f(x)}{\partial x_2} = 200(x_2 - x_1^2)$ (2)

De (1) et (2) on a :

$$\nabla f(x) = gradf(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}\right) = (-400(x_2 - x_1^2)x_1 - 2(1 - x_1), 200(x_2 - x_1^2))$$

Souvent, lorsque on travaille avec les matrices on écrit le gradient sous la forme d'un vecteur :

$$\nabla f(x) = gradf(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} -400(x_2 - x_1^2)x_1 - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix}$$

Définition 1.1.3 Soit Ω un ouvert de R^n et $f\colon \Omega \to$ R. La matrice $H = (h_{ij})_{n\times n}$, où les $h_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$ est appelée matrice hessienne. On la note H(x) $= \nabla^2 f(x)$.

Exemple 1.1.3

a) Si on prend $n = 2$ alors

$$\nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} \quad (3)$$

Déterminons $\nabla^2 f$, où $f(x) = f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_1}\right) = \frac{\partial}{\partial x_1}(-400(x_2 - x_1^2)x_1 - 2(1 - x_1))$$

$$\frac{\partial^2 f}{\partial x_1^2} = (800x_1^2 - 400\ (x_2 - x_1^2) + 2)$$

$$\frac{\partial^2 f}{\partial x_1^2} = 1200x_1^2 - 400x_2 + 2 \quad (4)$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_2}\right) = \frac{\partial}{\partial x_1}(200(x_2 - x_1^2))$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = -400x_1 \ (5)$$

$$\frac{\partial^2 f}{\partial x_2 \partial x_1} = \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_1}\right) = \frac{\partial}{\partial x_2}(-400(x_2 - x_1^2)x_1 - 2(1 - x_1))$$

$$\frac{\partial^2 f}{\partial x_2 \partial x_1} = \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_1}\right) = -400x_1 \ (6)$$

$$\frac{\partial^2 f}{\partial x_2^2} = \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_2}\right) = \frac{\partial}{\partial x_2}(200(x_2 - x_1^2))$$

$$\frac{\partial^2 f}{\partial x_2^2} = \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_2}\right) = 200$$

$$\frac{\partial^2 f}{\partial x_2^2} = 200(7)$$

Donc

$$\nabla^2 f = \begin{bmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}$$

1.1.2 Conditions nécessaires pour un minimum

Définition 1.3.1 Soit Ω un sous ensemble de R^n , $x^* \in \Omega$ et $f: \Omega \to$ R..

1) On dit que x^*est un point de minimum global de f sur Ω si

$$f(x^*) \leq f(x), \forall x \in \Omega$$

2) On dit que x^*est un point de minimum local de f sur Ω s'il existe un voisinage Vde x^*dans Ω tel que

$$f(x^*) \leq f(x), \forall x \in V$$

3) On dit que x^*est un point de maximum global de f sur Ω si

$$f(x^*) \geq f(x), \forall x \in \Omega$$

4) On dit que x^*est un point de maximum local de f sur Ω s'il existe un voisinage Vde x^*dans Ω tel que

$$f(x^*) \geq f(x), \forall x \in V$$

5) On dit que x^*est un point d'extrémum global de f sur Ω si x^*est soit un point de minimum global de f sur Ω, soit un point de maximum global de f sur Ω.
6) On dit que x^*est un point d'extrémum local de f sur Ω si x^*est soit un point de minimum local de f sur Ω, soit un point de maximum local de f sur Ω.

Dans toute la suite du cours, on parlera uniquement de minimisation de f. Pour la maximisation, faire la minimisation de $-f$ car $\max(f) = -\min(-f)$.

Théorème 1.3.1 (Condition nécessaire d'extrémum local)

Soit Ω un ouvert de R^n et $f: \Omega \rightarrow$ R. si la fonction possède un extrémum local en un point $x^* \in \Omega$ et si elle est dérivable en ce point, alors

$$\nabla f(x^*) = 0$$

Dans ce cas, x^* est appelé point critique de f.

Exemple 1.3.1

Soit $f(x) = f(x_1, x_2) = x_1^3 - x_2^3 + 3x_1x_2$. Déterminons ses points critiques

$$\frac{\partial f(x)}{\partial x_1} = 3x_1^2 + 3x_2$$

$$\frac{\partial f(x)}{\partial x_2} = -3x_2^2 + 3x_1$$

$\nabla f(x^*) = 0$ est équivalente au système d équations suivants

$$\begin{cases} 3x_1^2 + 3x_2 = 0 \\ -3x_2^2 + 3x_1 = 0 \end{cases} \Leftrightarrow \begin{cases} x_1^2 + x_2 = 0 \\ -x_2^2 + x_1 = 0 \end{cases} \Leftrightarrow \begin{cases} x_2x_1^2 + x_2^2 = 0 \\ -x_2^2 + x_1 = 0 \end{cases}$$

Ajoutons membre les deux équations du dernier système, nous obtenons :

$x_2x_1^2 + x_1 = 0 \Leftrightarrow x_1(x_2x_1 + 1) = 0 \Leftrightarrow x_1 = 0$ ou $x_2x_1 + 1 = 0$

$x_1 = 0 \Leftrightarrow x_2 = 0$

$(x_1, x_2) = (0,0)$ est donc un point critique de f.

$x_2x_1 + 1 = 0 \Leftrightarrow x_1 = -\frac{1}{x_2} \Leftrightarrow \frac{x_2^3+1}{x_2} = 0 \Leftrightarrow x_2^3 + 1 = 0 \Leftrightarrow x_2 = -1 \Leftrightarrow x_1 = 1$

$(x_1, x_2) = (1, -1)$ est donc un autre point critique de f.

Les points critiques de f sont donc :

$(0,0)$ et $(1, -1)$

Exemple 1.3.1

Soit $f(x) = f(x_1, x_2) = x_1^2 - x_2^2$. Déterminons ses points critiques

$$\frac{\partial f(x)}{\partial x_1} = 2x_1$$

$$\frac{\partial f(x)}{\partial x_2} = -2x_2$$

$\nabla f(x^*) = 0$ est équivalente au système d équations suivants

$$\begin{cases} x_1 & = 0 \\ x_2 & = 0 \end{cases}$$

$(x_1, x_2) = (0,0)$ est donc un point critique de f.

1.2 Rappels sur le calcul Matriciel

1.2.1 Déterminants

Définition 1.2.1 Soit $A = (a_{ij})_{2\times 2}$ une matrice de type 2×2. Le nombre $\det A = a_{11}a_{22} - a_{21}a_{12}$ est appelé déterminant de A.

Exemple 1.2.1

Soit $A = \begin{bmatrix} -1 & 3 \\ 2 & 4 \end{bmatrix}$ alors $\det A = (-1)(4) - (2)(3) = -10$

Définition 1.2.2 Soit $A = (a_{ij})_{n\times n}$ une matrice carrée et M_{ij} la matrice obtenue de A en supprimant la $i^{ème}$ ligne et la $j^{ème}$ colonne de A. On appelle déterminant de A, le nombre

$$\det A = (-1)^{i+1} a_{i1} \det M_{i1} + (-1)^{i+2} a_{i2} \det M_{i2} + \cdots + (-1)^{i+n} a_{in} \det M_{in} \qquad (2.1)$$

$$i = 1,2,\cdots m$$

ou

$$\det A = (-1)^{j+1} a_{1j} \det M_{1j} + (-1)^{j+2} a_{2j} \det M_{2j} + \cdots + (-1)^{j+n} a_{nj} \det M_{nj} \qquad (2.2)$$

$$j = 1,2,\cdots n$$

Exemple 1.2.2

Calculons déterminant de $A = \begin{bmatrix} -1 & 2 & 3 \\ 1 & -4 & 5 \\ -2 & 3 & 6 \end{bmatrix}$.

Prenons $i = 1$, alors

$$\det A = (-1)^{1+1} a_{11} \det M_{11} + (-1)^{1+2} a_{12} \det M_{12} + (-1)^{1+3} a_{13} \det M_{13}$$

Où

$$\det M_{11} = \det\begin{bmatrix} -4 & 5 \\ 3 & 6 \end{bmatrix} = -24-15 = -39,\ \det M_{12} = \det\begin{bmatrix} 1 & 5 \\ -2 & 6 \end{bmatrix} = 6+10 = 16$$

$$\det M_{13} = \det\begin{bmatrix} 1 & -4 \\ -2 & 3 \end{bmatrix} = (1)(3)-(-2)(-4) = 3-8 = -5$$

$$\det A = (-1)^2(-1)(-39)+(-1)^3(2)(16)+(-1)^4(3)(-5) = 39-32-15 = -8$$

Remarque : On trouve le même résultat si on prend $i=2$ ou $i=3$.

Prenons $i=2$, alors

$$\det A = (-1)^{2+1}a_{21}\det M_{21} + (-1)^{2+2}a_{22}\det M_{22} + (-1)^{2+3}a_{23}\det M_{23}$$

Où

$$\det M_{21} = \det\begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix} = 2\times 6 - 3\times 3 = 3,\ \det M_{22} = \det\begin{bmatrix} -1 & 3 \\ -2 & 6 \end{bmatrix} = (-1)\times 6 - (-2)\times 3 = 0$$

$$\det M_{23} = \det\begin{bmatrix} -1 & 2 \\ -2 & 3 \end{bmatrix} = (-1)(3)-(-2)(2) = -3+4 = 1$$

$$\det A = (-1)^3(1)(3)+(-1)^4(-4)(0)+(-1)^5(5)(1) = -3+0-5 = -8$$

Définition 1.2.3 a) Le nombre $(-1)^{i+j}\det M_{ij}$ est appelé cofacteur de l'élément a_{ij} de la matrice $A=(a_{ij})_{n\times n}$.

b) Le nombre $\det M_{ij}$ mineur de l'élément a_{ij} de la matrice $A=(a_{ij})_{n\times n}$.

c) L'équation (1.1) donne le développement du déterminant de $A=(a_{ij})_{n\times n}$ par rapport aux éléments de la $i^{ème}$ ligne.

d) L'équation (1.2) donne le développement du déterminant de $A=(a_{ij})_{n\times n}$ par rapport aux éléments de la $j^{ème}$ colonne.

Exemple 1.2.3

Calculons déterminant de $A = \begin{bmatrix} -1 & 2 & 3 \\ 1 & -4 & 5 \\ -2 & 3 & 6 \end{bmatrix}$, en utilisant l'équation (2.2).

Prenons $j=1$, alors

$$\det A = (-1)^{1+1} a_{11} \det M_{11} + (-1)^{1+2} a_{21} \det M_{21} + (-1)^{1+3} a_{31} \det M_{31}$$

Où

$$\det M_{11} = \det\begin{bmatrix} -4 & 5 \\ 3 & 6 \end{bmatrix} = (-4)(6) - 3\times 5 = -39,\ \det M_{22} = \det\begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix} = 2\times 6 - 3\times 3 = 3$$

$$\det M_{31} = \det\begin{bmatrix} 2 & 3 \\ -4 & 5 \end{bmatrix} = 2\times 5 - 3\times(-4) = 22$$

$$\det A = (-1)^2(-1)(-39) + (-1)^3(1)(3) + (-1)^4(-2)(22) = 39 - 3 - 44 = -8$$

Remarque : On trouve le même résultat si on prend $j=2$ ou $j=3$.

1.2.2 Valeurs Propres et Vecteurs Propres

Définition 1.2.4 Un scalaire λ est appelé valeur propre d'une matrice $A=(a_{ij})_{n\times n}$ s'il existe un vecteur non nul X tel que $AX=\lambda X$. Le vecteur X est appelé vecteur propre de A associé à la valeur propre λ.

Définition 1.2.5

a) $p_n(\lambda) = \det(A-\lambda I)$ est appelé polynôme caractéristique de la matrice *A*.

b) L'équation $p_n(\lambda)=0$ est appelé l'équation caractéristique de la matrice *A*.

Théorème 1.2.1 Soit $A=(a_{ij})_{n\times n}$. λ est une valeur propre de *A* si et seulement si elle est solution de l'équation caractéristique: $P_n(\lambda) = \det(A-\lambda I) = 0$

Exemple 1.2.4

Trouver les valeurs propres de $A=\begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$.

Déterminons le polynôme caractéristique de *A*

$$p_n(\lambda) = \det(A-\lambda I) = \det\begin{bmatrix} (1-\lambda) & 1 & -1 \\ 0 & (1-\lambda) & 0 \\ 1 & 0 & (1-\lambda) \end{bmatrix} = (\lambda-1)(\lambda^2 - 2\lambda + 2)$$

Cherchons les racines de l'équation caractéristique de *A*.

$$p_n(\lambda) = (\lambda - 1)(\lambda^2 - 2\lambda + 2) = 0$$

Les racines de cette équation sont : $1, 1+i$ et $1-i$. Donc $\lambda_1 = 1$, $\lambda_2 = 1+i$ et $\lambda_3 = 1-i$ sont les valeurs propres de *A*.

Calculons les vecteurs propres associés à ces valeurs propres

Pour $\lambda_1 = 1$, on a :

$$\begin{bmatrix} (1-1) & 1 & -1 \\ 0 & (1-1) & 0 \\ 1 & 0 & (1-1) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \text{avec} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \neq 0$$

$$\begin{cases} x_2 - x_3 = 0 \\ x_1 = 0 \end{cases} \Rightarrow X^{(1)} = \begin{bmatrix} 0 \\ x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \; x_2 \neq 0$$

Pour $\lambda_2 = 1-i$, on a :

$$\begin{bmatrix} (1-1+i) & 1 & -1 \\ 0 & (1-1+i) & 0 \\ 1 & 0 & (1-1+i) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \text{avec} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \neq 0$$

$$\begin{cases} ix_1 + x_2 - x_3 = 0 \\ ix_2 = 0 \\ x_1 + ix_3 = 0 \end{cases} \Rightarrow x_2 = 0, \; x_1 = -ix_3$$

Donc

$$X^{(2)} = \begin{bmatrix} -ix_3 \\ 0 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} -i \\ 0 \\ 1 \end{bmatrix}, \; x_3 \neq 0$$

Pour $\lambda_2 = 1+i$, on a :

$$\begin{bmatrix} (1-1-i) & 1 & -1 \\ 0 & (1-1-i) & 0 \\ 1 & 0 & (1-1-i) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \text{avec} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \neq 0$$

$$\begin{cases} -ix_1 + x_2 - x_3 = 0 \\ -ix_2 = 0 \\ x_1 - ix_3 = 0 \end{cases} \Rightarrow x_2 = 0\,,\ x_1 = ix_3$$

Donc

$$X^{(2)} = \begin{bmatrix} ix_3 \\ 0 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} i \\ 0 \\ 1 \end{bmatrix},\ x_3 \neq 0$$

Définition 1.2.5 Soit A une matrice de type $n \times n$ symétrique. Elle est dite définie positive si

$$x^T A x > 0, \text{ pour tout } x \in \mathrm{R}^n - \{0\}$$

Exemple 1.2.2

La matrice $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ est définie positive car

1) Elle symétrique.

2) $x^T A x > 0$ pour tout $x \in \mathrm{R}^3 - \{0\}$. En effet

$$x^T A x = [x_1 \quad x_2 \quad x_3] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1^2 + 5x_2^2 + 3x_3^2 > 0, x \in \mathrm{R}^3 - \{0\}$$

Définition 1.2.6 Soit A une matrice de type $n \times n$ et soit A_l, $l = 1, \dots, n$, la matrice extraite de A en éliminant les dernières $(n-l)$ lignes et les $(n-l)$ colonnes de A.On appelle déterminant principal de A qu'on note d_l, le déterminant de la matrice A_l, $l = 1, \dots, n$.

Exemple 1.2.3

Soit $A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$. Déterminons les déterminants principaux : d_l, $l = 1, 2,3$.

On a :

$A_1 = 2$, $A_2 = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$ et $A_3 = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$

$d_1 = 2$, $d_2 = detA_2 = det \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = 4 - 1 = 3$

Et $d_3 = det\, A_3 = det \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$

$$d_3 = det\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} = 2det\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} - (-1)\begin{bmatrix} -1 & -1 \\ 0 & 2 \end{bmatrix} + 0 = 2 \times (4-3) + (-2) = 4$$

Théorème 1.2.2 Soit A une matrice de type $n \times n$. Alors les affirmations suivantes sont équivalentes

1) La matrice A est définie positive
2) Les valeurs propres de A sont strictement positives
3) Les déterminants principaux $d_l, l = 1, \ldots, n$, de A sont strictement positifs

Exemple 1.2.4

La matrice $A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ de l'exemple 1.2.3 est définie positive car ses déterminants principaux : d_1 , d_2 et d_3 sont strictement positifs.

Déterminons les valeurs propres de A.

Déterminons le polynôme caractéristique de A : $P(\lambda) = det(A - \lambda I_3)$, où $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

$$P(\lambda) = det(A - \lambda I_3) = det\begin{bmatrix} 2-\lambda & -1 & 0 \\ -1 & 2-\lambda & -1 \\ 0 & -1 & 2-\lambda \end{bmatrix}$$
$$= (2-\lambda)det\begin{bmatrix} 2-\lambda & -1 \\ -1 & 2-\lambda \end{bmatrix}) - (-1)det\begin{bmatrix} -1 & -1 \\ 0 & 2-\lambda \end{bmatrix}$$

$P(\lambda) = det(A - \lambda I_3) = (2-\lambda)((2-\lambda)^2 - 1) - (2-\lambda)$

$P(\lambda) = det(A - \lambda I_3) = (2-\lambda)((2-\lambda)^2 - 1 - 1) = (2-\lambda)(\lambda^2 - 4\lambda + 2)$

$P(\lambda) = det(A - \lambda I_3) = (2-\lambda)\left(\lambda - (2+\sqrt{2})\right)\left(\lambda - (2-\sqrt{2})\right)$

Les racines de l'équation caractéristiques :
$P(\lambda) = (2-\lambda)\left(\lambda - (2+\sqrt{2})\right)\left(\lambda - (2-\sqrt{2})\right) = 0$

$\lambda_1 = 2$, $\lambda_2 = 2 + \sqrt{2}$, et $\lambda_3 = -1 = 2 - \sqrt{2}$

Ces racines représentent les valeurs propres de A. Elles sont toutes strictement positives. Donc d'après le théorème1.2.2 la matrice A est définie positive.

Exemple 1.2.6

Soit la matrice $A = \begin{bmatrix} 2 & -1 & 0 \\ 3 & -2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$

a) Déterminons les déterminant principaux : $d_l, l = 1, 2{,}3$ et les valeurs propres de A.

On a :

$A_1 = 2$, $A_2 = \begin{bmatrix} 2 & -1 \\ 3 & -2 \end{bmatrix}$ et $A_3 = \begin{bmatrix} 2 & -1 & 0 \\ 3 & -2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$

$d_1 = 2$, $d_2 = detA_2 = det\begin{bmatrix} 2 & -1 \\ 3 & -2 \end{bmatrix} = -4 + 3 = -1$

Et $d_3 = det\, A_3 = det\begin{bmatrix} 2 & -1 & 0 \\ 3 & -2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$

$$d_3 = det\begin{bmatrix} 2 & -1 & 0 \\ 3 & -2 & 1 \\ 0 & 0 & 1 \end{bmatrix} = 2det\begin{bmatrix} -2 & -1 \\ 0 & 1 \end{bmatrix} - (-1)\begin{bmatrix} 3 & 1 \\ 0 & 1 \end{bmatrix} + 0 = 2 \times (-2) + (3) = -1$$

Nous remarquons que la matrice A n'est pas définie positive car $d_2 = -1 < 0$ et $d_3 = -1 < 0$.

b) Déterminons les valeurs propres de A

Le polynôme caractéristique de A est $P(\lambda) = det(A - \lambda I_3)$, où $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

$$P(\lambda) = det(A - \lambda I_3) = det\begin{bmatrix} 2-\lambda & -1 & 0 \\ 3 & -2-\lambda & 1 \\ 0 & 0 & 1-\lambda \end{bmatrix} = (2-\lambda)det\begin{bmatrix} 2-\lambda & -1 \\ 3 & -2-\lambda \end{bmatrix}$$

$$P(\lambda) = det(A - \lambda I_3) = (1-\lambda)\big((2-\lambda)(-2-\lambda) - 1\big) + 3)$$

$$P(\lambda) = det(A - \lambda I_3) = (1-\lambda)\big(-(4-\lambda^2) + 3) = (1-\lambda)\big(\lambda^2 - 1\big)$$

$$P(\lambda) = det(A - \lambda I_3) = (1-\lambda)^2(\lambda + 1)$$

Les racines de l'équation caractéristiques :
$P(\lambda) = (1-\lambda)^2(\lambda + 1) = 0$

$\lambda_1 = 1$, $\lambda_2 = 1$, et $\lambda_3 = -1$

Ces racines représentent les valeurs propres de A. Elles ne sont pas toutes strictement positives. Donc d'après le théorème 1.2.2 la matrice A n'est pas définie positive.

Théorème 1.2.3 (conditions suffisantes de minimum local). Soit Ω un ouvert de R^n et $f : \Omega \rightarrow \mathrm{R}$. $x^* \in \Omega$ est point de minimum de f si

1) $\nabla f(x^*) = gradf(x^*) = 0$

2) $\mathrm{H}(x^*) = \nabla^2 f(x^*)$ est définie positive

Exemple 1.2.5

Soit $f(x) = f(x_1, x_2) = x_1^3 - x_2^3 + 3x_1x_2$. Le point $x^* = (1, -1)$ est il un point de minimum de f ?

De l'exemple 1.3.1, on :

$\frac{\partial f(x)}{\partial x_1} = 3x_1^2 + 3x_2$

$\frac{\partial f(x)}{\partial x_2} = -3x_2^2 + 3x_1$

$\nabla f(x) = (3x_1^2 + 3x_2, -3x_2^2 + 3x_1)$

$\nabla f(x^*) = \nabla f(1, -1) = (3 - 3, -3 + 3) = (0,0)$

$x^* = (1, -1)$ est donc un point critique.

Calculons la matrice hessienne de f au point x^* : $\mathrm{H}(x^*) = \nabla^2 f(x^*)$.

On a :

$\frac{\partial^2 f(x)}{\partial x_1^2} = 6x_1$, $\frac{\partial^2 f(x)}{\partial x_1 \partial x_2} = 3$, $\frac{\partial^2 f(x)}{\partial x_2 \partial x_1} = 3$ et $\frac{\partial^2 f(x)}{\partial x_2^2} = -6x_2$

Alors

$\mathrm{H}(x^*) = \nabla^2 f(x^*) = \nabla^2 f(1, -1) = \begin{bmatrix} 6 & 3 \\ 3 & 6 \end{bmatrix}$

Calculons les déterminants principaux de $\mathrm{H}(x^*)$

$d_1 = 6$ et $d_1 = det \begin{bmatrix} 6 & 3 \\ 3 & 6 \end{bmatrix} = 36 - 9 = 27$

Puisque

$d_1 > 0$ et $d_2 > 0$ alors $\mathrm{H}(x^*)$ est définie positive.

On vient de démontrer que

x^* est un point critique et $\mathrm{H}(x^*)$ est définie positive. Donc, D'après le théorème 1.2.3, $x^* = (1, -1)$ est un point de minimum de f.

Définition 1.2.6 Soit A une matrice de type $n \times n$ symétrique. Elle est dite définie négative si

$$x^T A x < 0, \text{ pour tout } x \in \mathrm{R}^n - \{0\}$$

Théorème 1.2.4 Soit A une matrice de type $n \times n$. Alors les affirmations suivantes sont équivalentes

1) La matrice A est définie négative

2) Les valeurs propres de A sont strictement négatives

3) Les déterminants principaux $d_l, l = 1, \dots, n$, de A sont tels que $(-1)^l d_l > 0, l = 1, \dots, n$

Théorème 1.2.5 Soit Ω un ouvert de R^n , $f: \Omega \to$ R . Et soit d_l^* , $l = 1, 2, \dots, n$ les déterminants principaux de la matrice hessienne $\nabla^2 f(x)$ au point $x^* \in \Omega$. Alors

1) f admet un minimum au point x^* si et seulement si $d_l^* > 0$, $l = 1, 2, \dots, n$.
2) f admet un maximum au point x^* si et seulement si $(-1)^l d_l^* > 0$, $l = 1, 2, \dots, n$.

Exemple 1.2.7

Soit $f(x) = f(x_1, x_2) = 2x_1x_2 + 2x_1 \ - x_1^2 - 2x_2^2$

1) Déterminons ses points critiques.
2) Déterminer leur nature.

Solution

a) Déterminons les points critiques de f .

$$\frac{\partial f(x)}{\partial x_1} = 2x_2 + 2 \ - 2x_1$$

$$\frac{\partial f(x)}{\partial x_2} = 2x_1 - 4x_2$$

$$\nabla f(x) = (2x_2 + 2 \ - 2x_1, 2x_1 - 4x_2)$$

$\nabla f(x)$ = (0,0) est équivalente au système d'équations suivant :

$$\begin{cases} 2x_2 + 2 \ - 2x_1 \ = 0 \\ \quad 2x_1 - 4x_2 = 0 \end{cases}$$

Ajoutons membre à membre ces deux équations nous obtenons :

$x_2 = 1$. Portons cette valeur dans la première équation de ce système, nous obtenons: $x_1 = 2$. $(x_1, x_2) = (2,1)$ est donc un point critique de f.

b) Déterminons la nature de ce point critique de f.

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_1}\right) = \frac{\partial}{\partial x_1}(2x_2 + 2 \ - 2x_1) = -2$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_2}\right) = \frac{\partial}{\partial x_1}(2x_1 - 4x_2)=2$$

$$\frac{\partial^2 f}{\partial x_2 \partial x_1} = \frac{\partial(2x_2 + 2 \quad - 2x_1)}{\partial x_2} = 2$$

$$\frac{\partial^2 f}{\partial x_2^2} = \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_2}\right) = \frac{\partial}{\partial x_2}(2x_1 - 4x_2) = -4$$

Donc

$$\nabla^2 f = \begin{bmatrix} -2 & 2 \\ 2 & -4 \end{bmatrix}$$

Calculons les déterminants principaux de $\mathrm{H}(x^*) = \nabla^2 f$

$d_1 = -2$ et $d_1 = det\begin{bmatrix} -2 & 2 \\ 2 & -4 \end{bmatrix} = 8 - 4 = 4$

Puisque

$(-1)^1 d_1 = 2 > 0$ et $2d_2 = 4 > 0$ alors $\mathrm{H}(x^*)$ est définie négative. Donc, D'après le théorème 1.2.3, $x^* = (1, -1)$ est un point de minimum de f.

1.3 Résolution d'un système d'équations algébriques non linéaires

1.3.2 Newton-Raphson

Supposons que l'on a à résoudre un système d'équations algébriques non linéaires :

$$(\mathrm{I}) \qquad \begin{cases} f_1(x_1, x_2, \ldots, x_n) = 0 \\ f_2(x_1, x_2, \ldots, x_n) = 0 \\ \vdots \\ f_n(x_1, x_2, \ldots, x_n) = 0 \end{cases}$$

Sous forme vectorielle, le système (I) s'écrit :

$$F(X) = 0$$

où

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{et} \quad F(X) = \begin{bmatrix} f_1(X) \\ f_2(X) \\ \vdots \\ f_n(X) \end{bmatrix}.$$

La méthode de *Newton-Raphson* consiste à créer une suite de vecteurs $X^{(1)}, X^{(2)}, \cdots, X^{(k)}, \cdots$ tels que

$$X^{(k)} = X^{(k-1)} + DX^{(k-1)} \qquad k = 1,2,3,\cdots$$

Où $X^{(0)}$ est choisi arbitrairement et $DX^{(k-1)} = \begin{bmatrix} DX_1^{(k-1)} \\ DX_2^{(k-1)} \\ \vdots \\ DX_n^{(k-1)} \end{bmatrix}$ est la solution du système linéaire suivant:

$$J(X^{(k-1)})DX^{(k-1)} = -F(X^{(k-1)})$$

où $J(X^{(k-1)})$ est la matrice la matrice jacobienne :

$$J(X^{(k-1)}) = \begin{bmatrix} \frac{\partial f_1(X^{(k-1)})}{\partial x_1} & \frac{\partial f_1(X^{(k-1)})}{\partial x_2} & \cdots & \frac{\partial f_1(X^{(k-1)})}{\partial x_n} \\ \frac{\partial f_2(X^{(k-1)})}{\partial x_1} & \frac{\partial f_2(X^{(k-1)})}{\partial x_2} & \cdots & \frac{\partial f_2(X^{(k-1)})}{\partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial f_n(X^{(k-1)})}{\partial x_1} & \frac{\partial f_n(X^{(k-1)})}{\partial x_2} & \cdots & \frac{\partial f_n(X^{(k-1)})}{\partial x_n} \end{bmatrix}$$

Cette suite, si elle converge, tend vers un vecteur solution du système (I).

Exemple 1.3.1

En utilisant la méthode de Newton, calculer la première itération approximant la solution du système suivant:

$$\text{(I)} \qquad \begin{cases} x_1^2 + x_1 x_2 - 10 = 0 = f_1(x_1,x_2) \\ x_2 + 3x_1 x_2{}^2 - 57 = 0 = f_2(x_1,x_2) \end{cases}$$

Solution

La $k^{\text{ème}}$ itération de Newton $X^{(k)} = \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \end{bmatrix}$ est donnée par :

$$X^{(k)} = X^{(k-1)} + DX^{(k-1)},\; k = 1,2,3,\cdots \tag{1}$$

où

$DX^{(k-1)} = \begin{bmatrix} DX_1^{(k-1)} \\ DX_2^{(k-1)} \end{bmatrix}$ est la solution du système linéaire:

$$J(X^{(k-1)})DX^{(k-1)} = -F(X^{(k-1)}) \tag{2}$$

où

$$J(X^{(k-1)}) = \begin{bmatrix} \dfrac{\partial f_1(X^{(k-1)})}{\partial x_1} & \dfrac{\partial f_1(X^{(k-1)})}{\partial x_2} \\ \dfrac{\partial f_2(X^{(k-1)})}{\partial x_1} & \dfrac{\partial f_2(X^{(k-1)})}{\partial x_2} \end{bmatrix} \tag{3}$$

$$F(X^{(k-1)}) = \begin{bmatrix} f_1(X^{(k-1)}) \\ f_2(X^{(k-1)}) \end{bmatrix}$$

En résolvant le système (2) par la méthode de Cramer, nous obtenons :

$$\mathrm{DX}_1^{(k\text{-}1)} = \frac{-f_1(X^{(k-1)})\dfrac{\partial f_2(X^{(k-1)})}{\partial x_2} + f_2(X^{(k-1)})\dfrac{\partial f_1(X^{(k-1)})}{\partial x_2}}{\dfrac{\partial f_1(X^{(k-1)})}{\partial x_1}\dfrac{\partial f_2(X^{(k-1)})}{\partial x_2} - \dfrac{\partial f_1(X^{(k-1)})}{\partial x_2}\dfrac{\partial f_2(X^{(k-1)})}{\partial x_1}} \tag{4}$$

$$\mathrm{D}_2^{(k-1)} = \frac{-f_2(X^{(k-1)})\dfrac{\partial f_1(X^{(k-1)})}{\partial x_1} + f_1(X^{(k-1)})\dfrac{\partial f_2(X^{(k-1)})}{\partial x_1}}{\dfrac{\partial f_1(X^{(k-1)})}{\partial x_1}\dfrac{\partial f_2(X^{(k-1)})}{\partial x_2} - \dfrac{\partial f_1(X^{(k-1)})}{\partial x_2}\dfrac{\partial f_2(X^{(k-1)})}{\partial x_1}} \tag{5}$$

Pour k =1 et en partant de $X^{(0)} = \begin{bmatrix} x^{(0)} \\ y^{(0)} \end{bmatrix} = \begin{bmatrix} 1.5 \\ 3.5 \end{bmatrix}$, on a :

$$\frac{\partial f_1(X^{(0)})}{\partial x} = 2x_1^{(0)} + x_2^{(0))} \quad \Rightarrow \quad \frac{\partial f_1(X^{(0)})}{\partial x} = 2(1.5) + 3.5 = 6.5 \qquad (6)$$

$$\frac{\partial f_1(X^{(0)})}{\partial y} = x_1^{(0)} \quad \Rightarrow \quad \frac{\partial f_1(X^{(0)})}{\partial y} = 1.5 \qquad (7)$$

$$\frac{\partial f_2(X^{(k-1)})}{\partial x} = 3\left(x_2^{(0)}\right)^2 \quad \Rightarrow \quad \frac{\partial f_2(X^{(0)})}{\partial x} = 3(3.5)^2 = 36.75 \qquad (8)$$

$$\frac{\partial f_2(X^{(k-1)})}{\partial y} = 1 + 6x_1^{(0)}x_2^{(0)} \quad \Rightarrow \quad \frac{\partial f_2(X^{(0)})}{\partial y} = 1 + 6(1.5)(3.5) = 32.5 \qquad (9)$$

$$f_1(X^{(0)}) = f_1(1.5{,}3.5) = (1.5)^2 + (1.5)(3.5) - 10 = -2.5 \qquad (10)$$

$$f_2(X^{(0)}) = f_2(1.5{,}3.5) = (3.5) + 3(1.5)(3.5)^2 = 1.625 \qquad (11)$$

Portons les valeurs données par (6) – (11) dans (2), nous obtenons :

$$\begin{bmatrix} 6.5 & 1.5 \\ 36.75 & 32.5 \end{bmatrix} \begin{bmatrix} DX_1^{(0)} \\ DX_2^{(0)} \end{bmatrix} = -\begin{bmatrix} 2.5 \\ 1.625 \end{bmatrix}$$

La résolution de ce système par la méthode de Cramer, c'est à dire en utilisant (4) et(5), nous donne :

$$\mathrm{DX}_1^{(0)} = 0.536 \quad \text{et} \quad \mathrm{DX}_2^{(0)} = -0.657$$

En tenant compte de ces valeurs, l'équation (1) pour k=1, nous donne :

$$x_1^{(1)} = x_1^{(0)} + DX_1^{(0)} = 1.5 + 0.536 = 2.036$$

$$x_2^{(1)} = x_2^{(0)} + DX_2^{(0)} = 3.5 - 0.657 = 2.843$$

$$X^{(1)} = \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \end{bmatrix} = \begin{bmatrix} 2.036 \\ 2.843 \end{bmatrix}$$

Chapitre II

Optimisation sans Contraintes

2.1 Définitions

Définition 2.1.1 Soit $f: \mathrm{R}^n \to \mathrm{R}$.

a) Le problème qui consiste à déterminer un élément x^* d'une partie Ω de R^n, s'il existe, pour lequel $f(x^*)$ est la plus petite (resp. la plus grande) valeur de f sur Ω est appelé problème d'optimisation.

On l'écrit sous la forme mathématique :

$$(P): \quad \min_{x\in\Omega} f(x) \quad \text{ou} \quad (P): \quad \max_{x\in\Omega} f(x)$$

b) Ω est appelé l'ensemble des contraintes.
c) La fonction f est appelé fonction objective.
d) Si $\Omega = \mathrm{R}^n$ on dit que (P) est un problème d'optimisation sans contraintes.

L'ensemble Ω est souvent donné sous la forme :Tapez une équation ici.

$\Omega = \{ x \in \mathrm{R}^n / g_i(x) = 0, i = 1, \dots, p \ et \ h_i(x) \leq 0 , i = 1, \dots, m\}$

Où

$g_i(x) = 0, i = 1, \dots, p$ sont appelées contraintes d'égalités.

$h_i(x) \leq 0 , i = 1, \dots, m$ sont appelées contraintes d'inégalités.

Exemple 2.1.1

$$(P): \quad \max_{x\in\mathrm{R}^2} f(x), \text{ où } f(x) = 4x_1 + 6x_2 - 2x_1^2 - 2x_1x_2 - 2x_2^2$$

est un problème d'optimisation sans contraintes. Par contre le problème suivant :

(P): $\max_{x\in\Omega\subset\mathrm{R}^2} f(x)$, où $f(x) = x_1 + x_2^4$ et Ω est l'ensemble des points (x_1, x_2) vérifiant:

$$\begin{cases} 3x_1 + x_2^2 \leq 9 \\ x_1 \geq 0 \\ x_2 \geq 0 \end{cases}$$

est appelé problème d'optimisation avec contraintes.

2.2 Méthode du gradient

Considérons le problème d'optimisation sans contraintes suivant:

$$(P): \quad \min_{x\in\Omega} f(x)$$

Le principe de cette méthode est de créer une suite de vecteurs $x^{(k)}$, $k = 1,2,3, \ldots$ telle que

$$x^{(k+1)} = x^{(k)} - \beta^{(k)} \nabla f(x^{(k)}) \, , k = 0,1,2,3, \ldots \tag{1.42}$$

Où $x^{(0)}$est choisi arbitrairement et $\beta^{(k)}$est un scalaire minimisant la fonction :

$$g(r) = f(x^{(k)} - r d^{(k)}) \tag{1.43}$$

$r = \beta^{(k)}$ est appelé profondeur de descente ou pas pour passer de l'étape k à l'étape $k + 1$.

$\nabla f(x^{(k)})$ est le gradient f.Cette suite, si elle converge, converge vers le point de minimum x^* de f.

Cette méthode est aussi connue sous le nom de méthode de la plus grande descente.

Remarque. Les problèmes de minimisation et de maximisation sont équivalents :

$$(P): \quad \min_{x \in \Omega} f(x) \qquad\qquad (P^*): \qquad \max_{x \in \Omega}(-f(x))$$

Car $\min_{x \in \Omega} f(x) = -\max_{x \in \Omega}(-f(x))$.

Donc un problème de maximisation peut être ramené à un problème de minimisation.

Exemple 2.2.2

En utilisant la méthode du gradient, calculer les deux premières itérations $x^{(1)}$ et $x^{(2)}$ approximant la solution x^* du problème d'optimisation sans contraintes suivant :

$$(P): \quad \max_{x \in \mathrm{R}^2} f(x), \text{ où } f(x) = 4x_1 + 6x_2 - 2x_1^2 - 2x_1x_2 - 2x_2^2$$

<u>Solution</u>

D'après la remarque précédente le problème de maximisation (P): $\max_{x \in \mathrm{R}^2} f(x)$

est équivalent au problème (P^*): $\min_{x \in \mathrm{R}^2}(-f(x))$

Appliquons la méthode gradient à la fonction $-f(x)$ au lieu de la fonction $f(x)$.

$= (1,1)$. Calcul de $x^{(1)}$.

Déterminons les dérivées partielles de f.

$$\frac{\partial f(x)}{\partial x_1} = 4 - 4x_1 - 2x_2$$

$$\frac{\partial f(x)}{\partial x_2} = 6 - 2x_1 - 4x_2$$

Donc $\nabla f(x) = (4 - 4x_1 - 2x_2\,, 6 - 2x_1 - 4x_2)$.

Choisissons $x^{(0)} = \left(x_1^{(0)}, x_2^{(0)}\right) = (1{,}1)$

Alors

$\nabla f\left(x^{(0)}\right) = (4 - 4(1) - 2(1)\,, 6 - 2(1) - 4(1)) = (-2\,, 0)$

$x^{(1)} = x^{(0)} + r^{(0)} \nabla f\left(x^{(0)}\right) = (1{,}1) + r^{(0)}(-2\,, 0) = \left(1 - 2r^{(0)}, 1\right)$

Où $r^{(0)}$ est la valeur minimisant la fonction :

$$g(r) = -f\left(x^{(0)} + r\nabla f\left(x^{(0)}\right)\right) = -f(1 - 2r, 1)$$

Ou maximisant $f(1 - 2r, 1)$.

Calculons $f(1 - 2r, 1)$

$f(1 - 2r, 1) = 4(1 - 2r) + 6(1) - 2(1 - 2r)^2 - 2(1 - 2r)(1) - 2(1)^2$

$f(1 - 2r, 1) = 4 - 8r + 6 - 2\left(1 + 4r^2 - 4r\right) - 2 + 4r - 2$

$f(1 - 2r, 1) = 4 - 8r + 6 - 2\left(1 + 4r^2 - 4r\right) - 2 + 4r - 2$

$\qquad\qquad = 2(1 - 2r) - 2\left(1 + 4r^2 - 4r\right) + 4$

Donc

$g(r) = -f(1 - 2r, 1) = 2\,(1 + 4r^2 - 4r) - 2(1 - 2r) - 4$

Le point de minimisation de $g(r)$ est $r^{(0)} = \frac{1}{4}$.

Donc

$x^{(1)} = \left(1 - 2r^{(0)}, 1\right) = \left(\frac{1}{2}, 1\right)$

Calcul de $x^{(2)}$.

$$x^{(2)} = x^{(1)} + r^{(1)}\nabla f\left(x^{(1)}\right)$$

$$\nabla f\left(x^{(1)}\right) = \nabla f(\frac{1}{2},1) = \left(4-4(\frac{1}{2})-2(1)\,,6-2(\frac{1}{2})-4(1)\right) = (0\,,1)$$

$$x^{(2)} = x^{(1)} + r^{(0)}\nabla f\left(x^{(1)}\right) = \left(\frac{1}{2},1\right) + r^{(1)}(0\,,1) = \left(\frac{1}{2},1+r^{(1)}\right)$$

Où $r^{(1)}$ est la valeur minimisant la fonction :

$$g(r) = -f\left(x^{(1)} + r\nabla f\left(x^{(1)}\right)\right) = -f\left(\left(\frac{1}{2},1\right)+r(0\,,1)\right) = -f\left(\frac{1}{2},1+r\right)$$

$$f\left(\frac{1}{2},1+r\right) = 4\left(\frac{1}{2}\right)+6(1+r)-2\left(\frac{1}{4}\right)-2\left(\frac{1}{2}\right)(1+r)-2(1+r)^2$$

$$f\left(\frac{1}{2},1+r\right) = -2(1+r)^2+5(1+r)+\frac{3}{2}$$

$$g(r) = -f\left(\frac{1}{2},1+r\right) = 2(1+r)^2-5(1+r)-\frac{3}{2}$$

Le point de minimisation de $g(r)$ est $r^{(1)} = \frac{1}{4}$

Donc

$$x^{(2)} = \left(\frac{1}{2},1+r^{(1)}\right) = \left(\frac{1}{2},\frac{5}{4}\right)$$

2.3 Méthode de Newton

Considérons le problème d'optimisation sans contraintes suivant:

$$(P): \quad \min_{x\in\Omega} f(x)$$

Le principe de la méthode de Newton est de créer une suite de vecteurs $x^{(k)}$, $k = 1,2,3,...$ telle que

$$x^{(k+1)} = x^{(k)} - \left[H^{(k)}\right]^{-1}\nabla f(x^{(k)})\,, k = 0,1,2,3,... \tag{1.44}$$

Où $x^{(0)}$est choisi arbitrairement,

$H^{(k)} = \nabla^2 f(x^{(k)})$ est la matrice hessienne au point $x^{(k)}$.

$\nabla f(x^{(k)})$ est le gradient f point $x^{(k)}$.

Cette suite, si elle converge, converge vers le point de minimum x^* de f.

Remarque. Les éléments de R^n, comme $x^{(k)}$, ∇f,…, sont représentés par des vecteurs colonnes.

Exemple 2.3.1

En utilisant la méthode de Newton, calculer la deux première itérations $x^{(1)}$ approximant la solution x^* du problème d'optimisation sans contraintes suivant :

$$(P): \quad \min_{x\in \mathrm{R}^2}(f(x)),\text{où } f(x) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$$

<u>**Solution**</u>

Calcul de $x^{(1)}$.

Déterminons les dérivées partielles de f.

$$\frac{\partial f(x)}{\partial x_1} = 1 + 4x_1 + 2x_2$$

$$\frac{\partial f(x)}{\partial x_2} = -1 + 2x_1 + 2x_2$$

Donc $\nabla f(x) = (1 + 4x_1 + 2x_2 \, , -1 + 2x_1 + 2x_2 \,)$.

Choisissons $x^{(0)} = \left(x_1^{(0)}, x_2^{(0)}\right) = (0,0)$, alors $\nabla f\left(x^{(0)}\right) = (1 \, , -1)$.

$$H(x) = \begin{bmatrix} \dfrac{\partial^2 f(x)}{\partial x_1^2} & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f(x)}{\partial x_2^2} \end{bmatrix}$$

$$\frac{\partial^2 f(x)}{\partial x_1^2} = \frac{\partial}{\partial x_1}\left(\frac{\partial f(x)}{\partial x_1}\right) = \frac{\partial}{\partial x_1}(1 + 4x_1 + 2x_2) = 4$$

$$\frac{\partial^2 f(x)}{\partial x_1 \partial x_2} = \frac{\partial}{\partial x_1}\left(\frac{\partial f(x)}{\partial x_2}\right) = \frac{\partial}{\partial x_1}(-1 + 2x_1 + 2x_2) = 2$$

$$\frac{\partial^2 f(x)}{\partial x_2^2} = \frac{\partial}{\partial x_2}\left(\frac{\partial f(x)}{\partial x_2}\right) = \frac{\partial}{\partial x_2}(-1 + 2x_1 + 2x_2) = 2$$

$$\frac{\partial^2 f(x)}{\partial x_2 \partial x_1} = \frac{\partial}{\partial x_2}\left(\frac{\partial f(x)}{\partial 1}\right) = \frac{\partial}{\partial x_1}\left((1 + 2x_1 + 2x_2)\right) = 2$$

$$H(x) = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$$

$$[H(x)]^{-1} = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}$$

$$[H(1,1)]^{-1}\nabla f(x^{(k)}) = \frac{1}{4}\begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 4 \\ -6 \end{bmatrix}$$

$$x^{(1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \frac{1}{4}\begin{bmatrix} 4 \\ -6 \end{bmatrix} = \begin{bmatrix} -1 \\ \frac{3}{2} \end{bmatrix}$$

$$\nabla f\left(x^{(1)}\right) = \left(1 + 4(-1) + 2(\frac{3}{2}), -1 + 2(-1) + 2(\frac{3}{2})\right)$$

$$\nabla f\left(x^{(1)}\right) = (0, 0)$$

La méthode de Newton a convergé après une itération.

Remarque. Ce résultat est valable pour toutes fonctions quadratiques :

$$f(x) = \frac{1}{2}x^T A x + b^T x + c$$

Le gradient et la hessienne de ce type de fonctions sont données respectivement par

$$\nabla f(x) = Ax + b$$

$$H(x) = A$$

2.4 Méthode du Gradient conjugué

Le principe de la méthode du gradient conjugué est de créer une suite de vecteurs $x^{(k)}$, $k = 1,2,3, \ldots$ telle que

$$x^{(k+1)} = x^{(k)} + r^{(k)} d^{(k)}, k = 0,1,2,3, \ldots$$

Où $x^{(0)}$est choisi arbitrairement

$r^{(k)}$ est la valeur minimisant la fonction :

$$g(r) = f\left(x^{(k)} + r d^{(k)} \right)$$

Il est donné par

$$r^{(k)} = \frac{\nabla J\left(x^{(k)} \right)^T d^{(k)}}{\left(A d^{(k)} \right)^T d^{(k)}},\ k = 0,1,2,3, \dots \qquad (1.45)$$

$$d^{(0)} = -\nabla J\left(x^{(0)} \right) \qquad (1.46)$$

$$d^{(k)} = -\nabla J\left(x^{(k)} \right) + \gamma^{(k)} d^{(k-1)},\ \gamma^{(0)} = 0,\ \ k = 1,2, \dots \qquad (1.47)$$

$$\gamma^{(k)} = \frac{\nabla J\left(x^{(k)} \right)^T \nabla J\left(x^{(k)} \right)}{\nabla J\left(x^{(k-1)} \right)^T \nabla J\left(x^{(k-1)} \right)}\ ,\quad k = 1,2, \dots \qquad (1.48)$$

Exemple 2.4.1

En utilisant la méthode du gradient conjugué, calculer les deux premières itérations
$x^{(1)}\ et\ \ x^{(2)}$ approximant la solution x^* du problème d'optimisation sans contraintes suivant :

$$(P):\qquad \min_{x \in \mathrm{R}^2} (f(x)), \text{où } f(x) = x_1^2 + x_1 x_2 + \tfrac{1}{2} x_2^2$$

__Solution__

Calcul de $x^{(1)}$.

Déterminons les dérivées partielles de f.

$$\frac{\partial f(x)}{\partial x_1} = 2x_1 + x_2$$

$$\frac{\partial f(x)}{\partial x_2} = x_1 + x_2$$

Donc $\nabla f(x) = (2x_1 + x_2\,, x_1 + x_2)$.

Choisissons $x^{(0)} = \left(x_1^{(0)}, x_2^{(0)} \right) = (1,1)$, alors $\nabla f\left(x^{(0)} \right) = (3\,,2)$.

$$d^{(0)} = -\nabla J\left(x^{(0)} \right) = (-3, -2)$$

$$x^{(1)} = x^{(0)} + r^{(0)} d^{(0)} = (1\,,1) - r^{(0)}(3\,,2)$$

$$x^{(1)} = \left(1 - 3r^{(0)}\,, 1 - 2r^{(0)} \right)$$

Où $r^{(0)}$ est la valeur minimisant la fonction :

$g(r) = f(1-3r, 1-2r)$=$(1-3r)^2$+$(1-3r)(1-2r)$+$\frac{1}{2}(1-2r)^2$

$g^{'}(r) = -6(1-3r) - 3(1-2r) - 2(1-3r)$+$-2(1-2r)$=0

$g^{'}(r) = -6(1-3r) - 3(1-2r) - 2(1-3r) - 2(1-2r)$=0

$g^{'}(r) = -6 + 18r - 3 + 6r - 2 + 6r - 2 + 4r = 0$

$34r - 13 = 0$

$r^{(0)} = \frac{13}{34}$

Donc

$x^{(1)} = \left(1 - 3r^{(0)}, 1 - 2r^{(0)}\right)$=$\left(1 - 3\frac{13}{34}, 1 - 2\frac{13}{34}\right)$

$x^{(1)} = \frac{1}{34}(-5, 8)$

Calcul de $x^{(2)}$.

$\nabla f\left(x^{(1)}\right)$=$\nabla f\left(\frac{1}{34}(-5,8)\right)$=$\left(\frac{2(-5)}{34} + \frac{8}{34}, -\frac{5}{34} + \frac{8}{34}\right) = \frac{1}{34}(-2,3)$

$d^{(1)} = -\nabla J\left(x^{(1)}\right) + \gamma^{(1)} d^{(0)}$

Où

$$\gamma^{(1)} = \frac{\nabla J\left(x^{(1)}\right)^T \nabla J\left(x^{(1)}\right)}{\nabla J\left(x^{(0)}\right)^T \nabla J\left(x^{(0)}\right)} = \frac{\frac{1}{34^2}(4+9)}{(9+4)} = \frac{1}{34^2}$$

Donc

$d^{(1)} = -\nabla J\left(x^{(1)}\right) + \gamma^{(1)} d^{(0)}$=$-\frac{1}{34}(-2,3) + \frac{1}{34^2}(-3,-2) = \frac{1}{34^2}(65,-104)$

$x^{(2)} = x^{(1)} + r^{(1)} d^{(1)} = \frac{1}{34}(-5,8) + \frac{r^{(1)}}{34^2}(65,-104)$

$x^{(2)} = \left(\frac{-5}{34} + \frac{65r^{(1)}}{34^2}, \frac{8}{34} - \frac{104r^{(1)}}{34^2}\right)$

Où $r^{(1)}$ est la valeur minimisant la fonction :

$$g(r) = f\left(\frac{-5}{34} + \frac{65r^{(1)}}{34^2}, \frac{8}{34} - \frac{104r^{(1)}}{34^2}\right)$$

$$g(r)=\left(\frac{-5}{34} + \frac{65r^{(1)}}{34^2}\right)^2+\left(\frac{-5}{34} + \frac{65r^{(1)}}{34^2}\right)\left(\frac{8}{34} - \frac{104r^{(1)}}{34^2}\right)+\frac{1}{2}\left(\frac{8}{34} - \frac{104r^{(1)}}{34^2}\right)^2$$

$$g'(r)=2\,\frac{65}{34^2}\left(\frac{-5}{34} + \frac{65r^{(1)}}{34^2}\right)+\frac{65}{34^2}\left(\frac{8}{34} - \frac{104r^{(1)}}{34^2}\right) - \frac{104}{34^2}\left(\frac{-5}{34} + \frac{65r^{(1)}}{34^2}\right) - \frac{104}{34^2}\left(\frac{8}{34} - \frac{104r^{(1)}}{34^2}\right)$$

$$g'(r)=r^{(1)}\left[2\,\frac{65^2}{34^4} + \frac{65}{34^2}\left(\frac{-104}{34^2}\right) + \frac{104^2}{34^2} - \frac{104\times 65}{34^2})\right] + \left[2\,\frac{65}{34^2}\left(\frac{-5}{34}\right) + \left(\frac{8}{34}\right)\frac{65}{34^2} - \frac{104}{34^2}\left(\frac{8}{34}\right) + \left(\frac{-104}{34^2}\right)\left(\frac{-5}{34}\right)\right]$$

$$g'\left(r^{(1)}\right)=\frac{5746}{34}r^{(1)} - 442 =0$$

$$r^{(1)} = 34\,\frac{442}{5746} \approx 2.6154$$

$$x^{(2)} = \left(\frac{-5}{34} + \frac{65 \times 2.6254}{34^2}, \frac{8}{34} - \frac{104 \times 2.6154}{34^2}\right)$$

$$x^{(2)} = \left(\frac{-170 + 170.001}{34^2}, \frac{272 - 272.0066}{34^2}\right)$$

$$x^{(2)} \approx (0,0)$$

$$\nabla f(x) = (2x_1 + x_2\,, x_1 + x_2)$$

$\nabla f\left(x^{(2)}\right)$=$\nabla f(0,0)$=$(2 \times 0 + 0,0 + 0) = (0,0)$

Donc $x^{(2)} \approx (0,0)$ est le point de minimum de $f(x) = x_1^2 + x_1x_2 + \frac{1}{2}x_2^2$