

II - Optimisation non linéaire sans contraintes:

Partie A:

1 - Introduction:

Nous allons étudier le problème d'optimisation sans contraintes où la minimisation de la fonction $f: \mathbb{R}^n \rightarrow \mathbb{R}$ sur tout l'espace \mathbb{R}^n .

Nous considérons le problème formulé de la façon suivante :

$$(P) \begin{cases} \min f(x) \\ x \in \mathbb{R}^n \end{cases}$$

On cherche donc à résoudre ce problème (P)

Il s'agit donc de déterminer un point x^* de \mathbb{R}^n tel que :

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

Remarque: La recherche d'un maximum pour la fonction f peut se ramener à la recherche d'un minimum.

Définition: (Domaine d'une fonction, fonction propre)

soit $f: \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction à valeurs réelles.
Le domaine de f , noté $\text{dom}(f)$ est l'ensemble défini par :

$$\text{dom}(f) = \{x \in \mathbb{R}^n : f(x) < +\infty\}.$$

si $\text{dom}(f) \neq \emptyset$, alors f est dite propre.

2 - Solution optimal :

Définition:

1 - On dit que la fonction f du problème (P) possède un minimum global en $x^* \in \mathbb{R}^n$ si et seulement si :

$$\forall x \in \mathbb{R}^n : f(x) \geq f(x^*).$$

2. On dit que le point x^* de \mathbb{R}^n est un minimum local du problème (p) si et seulement s'il existe un voisinage $V_\varepsilon(x^*)$ tel que :

$$\forall x \in V_\varepsilon(x^*), f(x) \geq f(x^*).$$

Remarques :

- Un minimum global est clairement un minimum local.
- si on dit simplement minimum on comprend minimum global.

Remarque :

- On s'intéressera à la recherche des points réalisant des minima car la recherche des maxima peut se ramener à celle des minima, comme le montre la proposition suivante :

Proposition :

Soient D un ensemble non vide de \mathbb{R}^n et f une fonction définie de D dans \mathbb{R} . Si $x^* \in D$ réalise un minimum (global ou local) de f sur D alors x^* réalise un maximum (global ou local) de $-f$ sur D . Plus précisément :

$$\max \{ f(x); x \in D \} = - \min \{ -f(x); x \in D \}.$$

Remarques :

- Les questions qui nous intéressent dans les problèmes d'optimisation sont les suivantes :

1. Existence et unicité de la solution du problème (p).
2. Caractérisation des solutions (les points réalisant le minimum).
3. Calcul, expression explicite du minimum et algorithmes permettant de l'approcher.

(suite chap II)

3 - Résultats d'existence et d'unicité :

Définition : (Fonction coercive)

On dit qu'une fonction définie de \mathbb{R}^n dans \mathbb{R} est coercive si $\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty$.

Exemples :

1) $f: \mathbb{R} \rightarrow \mathbb{R}$

$x \mapsto f(x) = x^2$ est une fonction coercive.

2) $g: \mathbb{R} \rightarrow \mathbb{R}$

$x \mapsto x^3$

g n'est pas une fonction coercive, car $\lim_{x \rightarrow -\infty} g(x) = -\infty$

3) $h: \mathbb{R}^2 \rightarrow \mathbb{R}$

$(x, y) \mapsto h(x, y) = x^2 - y^2$

h n'est pas une fonction coercive.

En effet, en choisissant la suite $x_n = (0, n)$, on a

$\lim_{n \rightarrow \infty} \|x_n\| = \infty$ mais $\lim_{n \rightarrow \infty} h(x_n) = -\infty$.

Théorème 3-1 : (conditions d'existence)

Soit $f: \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction propre, continue et coercive. Alors problème (P) admet au moins une solution.

Théorème 3-2 : (condition d'unicité)

Soit $f: \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction strictement convexe. Alors le problème (P) admet au plus une solution.

Théorème 3-3 : (condition d'existence et d'unicité)

Soit $f: \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction de classe C^1 . On suppose

(3)

qu'il existe $\alpha > 0$ tel que :

$$\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n : \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \alpha \|x - y\|^2$$

Alors f est strictement convexe et coercive. En particulier le problème (P) admet une solution unique.

4 - Condition d'optimalité :

4-1 - Condition d'optimalité du 1^{er} ordre :

Théorème 4-1-1 : (condition nécessaire du 1^{er} ordre)

Soit $f: \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction gâteaux différentiable sur \mathbb{R}^n . Si x^* réalise un minimum (global ou local) de f sur \mathbb{R}^n , alors : $\nabla f(x^*) = 0 \dots (*)$

Définition : (point critique)

- Un point x^* vérifiant la relation (*) est appelé point critique ou point stationnaire.

Exemples :

1) On considère la fonction f définie sur \mathbb{R}^2 dans \mathbb{R} par : $f(x, y) = 2x + y^2 - y$
Le calcul du gradient de f donne :

$$\nabla f(x, y) = \begin{pmatrix} 2 \\ 2y - 1 \end{pmatrix}$$

Le système d'équation $\nabla f(x, y) = 0$ n'admet pas de solution.

En effet, $\nabla f(x, y) = \begin{pmatrix} 2 \\ 2y - 1 \end{pmatrix} = 0$

$$\Leftrightarrow \begin{cases} 2 = 0 \\ 2y - 1 = 0 \end{cases} \text{ (impossible.)}$$

Donc la fonction f n'admet aucun point critique.

D'où : f n'admet pas de minimum sur \mathbb{R}^n .

*) On considère la fonction $f: \mathbb{R} \rightarrow \mathbb{R}$

$$x \mapsto f(x) = x^3$$

$$\text{On a : } f'(x) = 3x^2$$

$$\text{donc : } f'(x) = 0 \Rightarrow x = 0$$

Donc $x = 0$ est un point critique pour f , mais ce n'est pas un minimum.

Théorème 4-1-2 : (condition nécessaire et suffisante du 1^{er} ordre dans le cas convexe)!

Soit $f: \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction gâteaux différentiable et convexe sur \mathbb{R}^n .

Un point x^* réalise un minimum global de f sur \mathbb{R}^n

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

Exemple :

On considère la fonction f définie de \mathbb{R}^2 dans \mathbb{R} par :

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

• f admet-elle un minimum ? Est-il unique ?

On a : f est une fonction propre, continue et coercive, alors elle admet au moins un minimum.

Le calcul du gradient de f donne :

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

Le ∇f est un opérateur strictement monotone.

En effet, pour $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ et $Y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ avec $X \neq Y$, on a :

$$\begin{aligned} \langle \nabla f(X) - \nabla f(Y), X - Y \rangle &= \left\langle \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} - \begin{pmatrix} 2y_1 \\ 2y_2 \end{pmatrix}, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle \\ &= (x_1 - y_1, x_2 - y_2) \begin{pmatrix} 2(x_1 - y_1) \\ 2(x_2 - y_2) \end{pmatrix} \quad (5) \end{aligned}$$

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle = 2(x_1 - y_1)^2 + 2(x_2 - y_2)^2 > 0.$$

Alors f est strictement convexe. Donc f admet au plus un minimum.

Par conséquent, f admet un unique minimum.

• points critiques ?

$$\nabla f(x) = \begin{pmatrix} 2x \\ 2y \end{pmatrix} = 0 \iff x = y = 0$$

Donc $x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ est le seul point critique.

Donc c'est l'unique minimum de f puisque la condition nécessaire et suffisante d'optimalité du premier ordre est satisfaite.

4-2. Conditions d'optimalité du 2nd ordre:

Théorème 4-2-1: (condition nécessaire du 2nd ordre)

On suppose que x^* est un minimum (local) de f et que f est deux fois gateaux différentiable sur \mathbb{R}^n . Alors :

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

(b) $\forall x \in \mathbb{R}^n, \langle \nabla^2 f(x^*) x, x \rangle \geq 0$

cad: que le Hessien de f est une matrice semi-définie positive.

Exemple:

- On considère la fonction f définie de \mathbb{R}^2 dans \mathbb{R} par :

$$f(x, y) = 3x^4 - 4x^2y + y^2$$

$$\nabla f(x, y) = \begin{pmatrix} 12x^3 - 8xy \\ -4x^2 + 2y \end{pmatrix} = 0 \iff \begin{cases} 12x^3 - 8xy = 0 \\ -4x^2 + 2y = 0 \end{cases}$$

$$\implies \begin{cases} x = 0 \\ y = 0 \end{cases}$$

Donc: $x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ est le seul point critique de f .
Est-il minimum ?

(Suite Chap 2) (suite Exemple)

- Calculons la matrice Hessienne de f au point x^*

$$\nabla^2 f(x, y) = \begin{pmatrix} 36x^2 - 8y & -8x \\ -8x & 2 \end{pmatrix}$$

$$\nabla^2 f(x^*) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$$

Cette condition n'est pas suffisante pour conclure que x^* est un minimum pour f .

Néanmoins, on peut vérifier que ce point critique n'est ni un minimum, ni un maximum en remarquant que :

$$f\left(\frac{1}{2}, \frac{1}{2}\right) = -0,0625 < f(0,0) = 0 < f(0,1) = 1.$$

Théorème 4.2.2: (Condition suffisante du 2nd ordre)

Soit $f: \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction 2 fois gâteaux différentiable.

Une condition suffisante pour que x^* soit un minimum local de f sur \mathbb{R}^n est :

(a) $\nabla f(x^*) = 0$. (stationnarité).

(b) $\forall x \in \mathbb{R}^n, x \neq 0, \langle \nabla^2 f(x^*)x, x \rangle > 0$

C.à.d que le Hessien de f est une matrice définie positive.

Exemple :

On considère la fonction f définie de \mathbb{R}^2 dans \mathbb{R} par :

$$f(x, y) = x^4 - 4xy + y^4.$$

Le point $x^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ est un point critique pour f .

En effet,

$$\nabla f(x, y) = \begin{pmatrix} 4x^3 - 4y \\ -4x + 4y^3 \end{pmatrix} \Rightarrow \nabla f(1, 1) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

passons maintenant à la condition de second ordre.

$$\nabla^2 f(x, y) = \begin{pmatrix} 12x^2 & -4 \\ -4 & 12y^2 \end{pmatrix}$$

$$\Rightarrow \nabla^2 f(x^*) = \begin{pmatrix} 12 & -4 \\ -4 & 12 \end{pmatrix}$$

Vérifions que c'est une matrice définie positive par la méthode des mineurs par exemple

$$\Delta_1 = 12 > 0$$

$$\Delta_2 = |\nabla^2 f(x^*)| = (12)^2 - (-4)^2 > 0$$

Donc la matrice Hessienne de f au point x^* est définie positive. Par conséquent, $x^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ est un minimum pour f .

Partie B :

Dans cette partie, on introduit une classe importante d'algorithmes de résolution des problèmes d'optimisation sans. Le concept central est celui de la direction de descente.

1. Notion d'algorithme :

Considérons le problème consistant à minimiser une fonction $f(x)$ où $x \in \mathbb{R}^n$, un algorithme itératif permettant de résoudre ce problème est un processus itératif générant une suite de vecteur (x_1, x_2, \dots, x_n) de \mathbb{R}^n en fonction d'une séquence d'instructions et d'une condition d'arrêt, c'est-à-dire qu'à partir d'un point initial x_0 , on engendre (construit) une suite (x_n) qui converge vers la solution du problème

$$(P) : \left\{ \min f(x) : x \in \mathbb{R}^n \right\}$$

2. Méthode du gradient :

L'algorithme du gradient désigne un algorithme d'optimisation différentiable, l'algorithme est itératif et procède donc par améliorations successives. Au point courant, un déplacement est effectué dans la direction opposée au gradient, de manière à faire décroître la fonction.

Cette description montre que l'algorithme fait partie de la famille des algorithmes à directions de descente.

La direction de descente choisie sera à chaque itération

$$d_k = -\nabla f(x_k).$$

Les points sont ainsi successivement générés par cette méthode de la manière suivante :

$$\begin{cases} x_0 \in \mathbb{R}^n \text{ donné} \\ x_{k+1} = x_k - \rho_k \nabla f(x_k), \rho_k \end{cases}$$

L'algorithme de gradient est donné par :

1. Initialisation $k=0$:

choix de $x_0, \rho_0 > 0$ et ε .

2. Itération :

$$x_{k+1} = x_k - \rho_k \cdot \nabla f(x_k)$$

3. Critère d'arrêt :

si $\|x_{k+1} - x_k\| < \varepsilon$ ou si $\|\nabla f(x_k)\| < \varepsilon$ Stop.

sinon, on pose $k = k+1$, et on retourne à 2.

avec ε est un réel positif (petit) donné qui représente la précision désirée (la valeur que nous devons fixer pour ε dépend du problème considéré).

2-1. Méthode du gradient à pas optimal :

Cette méthode consiste à faire les itérations suivantes :

$$\begin{cases} d_k = -\nabla f(x_k) \\ x_{k+1} = x_k + p_k d_k, \quad p_k > 0 \end{cases}$$

Où p_k est choisi par la règle de minimisation, il consiste à choisir, à chaque itération p_k comme étant la solution optimale du problème de minimisation monodimensionnelle de f le long de la demi-droite définie par le point x_k et la direction d_k .

Donc, p_k est choisi de manière à ce que :

$$f(x_k + p_k d_k) < f(x_k + p d_k), \quad \forall p > 0.$$

Dans ce cas, les directions de descente d_k générées vérifient : $d_{k+1}^t \cdot d_k = 0$, car si on introduit la fonction $g(p) = f(x_k + p d_k)$

On a :

$$g'(p) = \nabla f(x_k + p d_k)^t \cdot d_k,$$

et puisque g est dérivable on a nécessairement

$$g'(p_k) = 0 \quad \text{donc :}$$

$$\nabla f(x_k + p_k d_k)^t \cdot d_k = \nabla f(x_{k+1})^t \cdot d_k = -d_{k+1}^t \cdot d_k = 0$$

Cette opération de détermination du pas s'appelle la recherche linéaire.

Remarque :

- La méthode de gradient à pas optimal propose un choix du pas qui rend la fonction de coût minimale le long de la direction de descente choisie, plus précisément le de $x_{k+1} = x_k - p_k \nabla f(x_k)$.

(suite / chap 2, partie B)

Exemple :

soit la fonction quadratique suivante :
 $f(x) = \frac{1}{2} x^t A x - b^t x$, avec $A > 0$ (c-à-d. A est matrice définie positive)

On note $g(p) = f(x_k + p d_k)$, où le pas optimal p_k est caractérisé $g'(p_k) = 0$.

On a donc :
 $\nabla f(x_k + p_k d_k)^t \cdot d_k = (A(x_k + p_k d_k) - b)^t \cdot d_k = 0$
 $\nabla f(x_k + p_k A d_k)^t \cdot d_k = 0$

$$\Rightarrow p_k = \frac{(\nabla f(x))^t \cdot d_k}{d_k^t \cdot A d_k} > 0$$

Car d_k est une direction de descente et $d_k^t \cdot A d_k > 0$
La méthode du gradient à pas optimal peut s'écrire :

$$x_{k+1} = x_k + p_k d_k \quad \text{avec : } \begin{cases} d_k = b - A x_k \\ p_k = \frac{d_k^t \cdot d_k}{d_k^t \cdot A} \end{cases}$$

2 - Méthode de Newton :

- Cette méthode permettant de trouver une solution d'un système (en général non linéaire) de n équations avec n inconnues, autrement dit de trouver les zéros d'une fonction $F: \mathbb{R}^n \rightarrow \mathbb{R}$ i.e. résoudre l'équation non linéaire $F(x) = 0$ dans \mathbb{R}^n .

Principe de la méthode de Newton :

- L'idée de la méthode itérative de Newton est de minimiser à chaque itération l'approximation quadratique de f (de classe C^2) au point courant x_k et donnée par le développement de Taylor d'ordre 2 :

$$f(x) \cong q(x) = f(x_k) + (\nabla f(x_k), x - x_k) + \frac{1}{2} (x - x_k)^t \cdot H(x_k) \cdot (x - x_k)$$

On considère le problème : $\{ \min q(x), x \in \mathbb{R}^n \}$

Une condition nécessaire pour que le minimum de $q(x)$ soit atteint est $\nabla q(x) = 0$ tel que :

$$\nabla q(x) = \nabla f(x_k) + H(x_k)(x - x_k) = 0$$

Le vecteur généré à l'itération $(k+1)$ est le vecteur minimisant $q(x)$

$$\nabla q(x_{k+1}) = 0 \Rightarrow \nabla f(x_k) + H(x_k)(x_{k+1} - x_k) = 0$$

$$\Rightarrow x_{k+1} = x_k - (H(x_k))^{-1} \cdot \nabla f(x_k).$$

Donc, la direction de descente choisie à chaque itération est : $d_k = - (H(x_k))^{-1} \cdot \nabla f(x_k)$

Elle est bien une direction de descente car elle vérifie la relation :

$$(d_k; \nabla f(x_k)) < 0, \quad \forall k \geq 0$$

En effet, $d_k = - (H(x_k))^{-1} \cdot \nabla f(x_k)$

$$\Rightarrow \nabla f(x_k) = - H(x_k) \cdot d_k.$$

$$(d_k; \nabla f(x_k)) = (d_k, - H(x_k) \cdot d_k) = - d_k^t \cdot H(x_k) \cdot d_k < 0.$$

Car la matrice hessienne est définie positive.

d_k est appelée la direction de Newton et les points successivement générés par cette méthode de la manière :

$$\begin{cases} x_0 \in \mathbb{R}^n \text{ donné, } p = 1 \\ x_{k+1} = x_k - (H(x_k))^{-1} \cdot \nabla f(x_k) \end{cases}$$

(Suite du chapitre II): (partie B)

3. La méthode de relaxation:

Introduction:

Cette méthode permet de ramener un problème de minimisation dans \mathbb{R}^n à la résolution d'un problème de minimisation dans \mathbb{R} à chaque itération.

Principe de la méthode:

à l'origine, on cherche à résoudre dans \mathbb{R}^n :

$$(P): \begin{cases} \min f(x) \\ x \in \mathbb{R}^n \end{cases} \quad \text{avec: } x = (x_1, \dots, x_n)^t \\ f: \mathbb{R}^n \rightarrow \mathbb{R}.$$

Etant donné un itéré $x^k = (x_1^k, \dots, x_n^k)^t$, on fixe toutes les coordonnées de x sauf la première et on minimise sur la 1^{ère} coordonnée.

$$\begin{cases} \min f(x_1, x_2^k, \dots, x_n^k) = \min \phi(x) \\ x \in \mathbb{R} \end{cases}$$

On distingue ainsi la 1^{ère} coordonnée de x^{k+1} notée x_1^{k+1} , on peut utiliser la méthode de Newton dans \mathbb{R} .

On recommence ensuite en fixant la 1^{ère} coordonnée à x_1^{k+1} et les $(n-2)$ coordonnées pour faire varier la 2^{ème} coordonnée.

On minimise dans \mathbb{R} par rapport à la 1^{ère} coordonnée et ainsi de suite jusqu'à la n ème coordonnée.

L'algorithme obtenu est le suivant:

• Algorithme de relaxation:

1. $k=0$, choix de $x^0 \in \mathbb{R}^n$ et ε (précision)

2. Itération:

Pour i variant de 1 jusqu'à n .

On calcule la solution x_i^{k+1} de

$$\begin{cases} \min f(x_1^{k+1}, \dots, x_{i-1}^{k+1}, x, x_{i+1}^k, \dots, x_i^k) \\ x \in \mathbb{R} \end{cases}$$

3. Critère d'arrêt :

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

sinon $k = k + 1$ et retourner à l'étape (2).

ou bien :

si $\|\nabla f(x^{k+1})\| < \varepsilon$. (stop)

sinon $k = k + 1$ et retourner à (2).

Exemple :

$$f: \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$(x, y) \mapsto f(x, y) = 3x^2 + 3y^2.$$

$$x^0 = \begin{pmatrix} x_1^0 \\ x_2^0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{et } \varepsilon = 10^{-5}.$$

- Calculons $x^1 = \begin{pmatrix} x_1^1 \\ x_2^1 \end{pmatrix}$

pour $i = 1$:

$$\bar{x}^1 = \begin{pmatrix} x \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ x_2^0 \end{pmatrix}$$

x est la solution de :

$$\begin{cases} \min f(\bar{x}^1) \\ x \in \mathbb{R} \end{cases} \iff \begin{cases} \min \phi(x) = 3x^2 + 3 \\ x \in \mathbb{R} \end{cases}$$

$$\phi'(x) = 6x \quad \text{et } \phi'(x) = 0 \Rightarrow x = 0$$

$$\phi''(x) = 6 > 0$$

x est un minimum. Donc $x_1^* = 0$.

pour $i = 2$:

$$\bar{x}^2 = \begin{pmatrix} x_1^* \\ x \end{pmatrix} = \begin{pmatrix} 0 \\ x \end{pmatrix}$$

(Suite : partie B : Chapitre II)

Exemple : (La méthode de Newton).

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

$$= \frac{1}{2} (x, y) \underbrace{\begin{pmatrix} 4 & -2 \\ -2 & 2 \end{pmatrix}}_A \begin{pmatrix} x \\ y \end{pmatrix} - \underbrace{(-2, 2)}_{b^t} \begin{pmatrix} x \\ y \end{pmatrix}$$

N-B :

- l'écriture de f sous la forme : $x^t A x - b^t x$.

$$x^0 = \begin{pmatrix} 10 \\ 5 \end{pmatrix}, \quad \epsilon = 10^{-5}.$$

$$\nabla^2 f(x) = \begin{pmatrix} 4 & -2 \\ -2 & 2 \end{pmatrix}, \quad \forall x \in \mathbb{R}^2.$$

$$\nabla f(x) = \begin{pmatrix} 4x - 2y + 2 \\ 2y - 2x - 2 \end{pmatrix}$$

$$[\nabla^2 f(x)]^{-1} = \frac{1}{4} \cdot \begin{pmatrix} 2 & 2 \\ 2 & 4 \end{pmatrix}$$

$$x^1 = x^0 - [\nabla^2 f(x)]^{-1} \cdot \nabla f(x)$$

$$= \begin{pmatrix} 10 \\ 5 \end{pmatrix} - \frac{1}{4} \begin{pmatrix} 2 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 32 \\ -12 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\|\nabla f(0, 1)\| = \left\| \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\| = 0 < \epsilon.$$

Le minimum de f est $x^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ et $f(x^*) = -1$.

Remarques:

1/. L'inconvénient de cette méthode est sa sensibilité au choix du problème de départ x^0 .

En pratique On essaie de se rapprocher de x^* par une méthode de type gradient, puis continuer avec la méthode de Newton.

2/. Un autre inconvénient est le calcul de l'inverse du Hessien à chaque itération des méthodes dite quasi-Newton ont été développés. Les dernières gardent la rapidité de la méthode de Newton tout en évitant le calcul de l'inverse.

3/. L'avantage de cette méthode est sa rapidité de fait qu'elle a une vitesse de convergence quadratique.

4/. Lorsqu'elle est appliquée à une fonction quadratique, elle converge au bout d'une seule itération quelque soit le point de départ x^0 choisi.