Soit $J : \mathbb{R}^n \longrightarrow \mathbb{R}$ la fonction objective

Méthode du gradient : On considère l'algorithme suivant

$$(I) \ldots \begin{cases} x_0 \in \mathbb{R}^n & \text{l'étape initiale} \\ x_{k+1} = A(x_k), \ k \geq 0 & \text{les itérations} \end{cases}$$

On va minimiser la fonction objective $J(x)$, pour cela, on construit la suite (x_k) de telle sorte que $J(x_k) \geq J(x_{k+1})$, $\forall k \geq 0$

et d'après le théorème dans le cours, s'il existe une direction de descente "d" alors il existe $\delta > 0$ tel que $f(x+\delta d) < f(x)$

c-à-d, on construit la suite (x_k) où $x_{k+1} = x_k + \rho_k d_k$

tels que d_k est une direction de descente et $\rho_k > 0$ est le pas de descente.

On choisit la direction la plus forte descente c-à-d $d_k = -\nabla J(x_k)$

donc l'algorithme (I), sera :

$$\text{II} \begin{cases} x_0 \in \mathbb{R}^n \\ x_{k+1} = x_k - \rho_k \nabla J(x_k) \quad k \geq 0, \ \rho_k > 0 \end{cases}$$

c-à-d, on peut réécrire l'algorithme II sous la forme

Initialisations
$$\begin{cases} x_0 \in \mathbb{R}^n \text{ donnée} \\ \rho > 0 \text{ donnée} \\ k = 0;\ 0 < \varepsilon \text{ assez petit} \end{cases}$$

Itérations
$$\begin{cases} x_{k+1} = x_k - \rho \nabla J(x_k) \\ \|x_{k+1} - x_k\| \leq \varepsilon \text{ ou } \|\nabla J(x_k)\| \leq \varepsilon \\ k = k+1 \end{cases}$$

(car le min J vérifie la condition d'optimalité suffisante $\nabla J(x^*) = 0$)

Théorème Soit $J \in C^1(\mathbb{R}^n)$ coercive et strictement convexe

s'il existe $M > 0$ tel que

$$\|\nabla J(x) - \nabla J(y)\| \leq M \|x - y\| \Leftrightarrow \text{le gradien lipschitzien}$$

Alors la méthode du gradient converge $\forall\, \rho_k \in [\beta_1, \beta_2]$

tel que $0 < \beta_1 < \beta_2 < \frac{2}{M}$.

Proposition Si f est elliptique alors elle est coercive

et strictement convexe.

Exemple $f(x,y) = x^2 + y^2$, $\rho = 0,9 ; 0,5 ; 1,5$; $x_0 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

La méthode du gradien à pas variable :

On peut faire varier le pas à chaque itération

On obtient l'algorithme suivant

Initialisations

$x_0 \in \mathbb{R}^n$

$\rho_0 > 0$

$k = 0$

$\varepsilon > 0$ assez petit

→ Itération

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

$\|\nabla J(x_k)\| \leq \varepsilon$

$k = k+1$

```
clc; clear all;
x = input('le vecteur initial');
r = input('le pas');
k = 0; epsilon = 1;
while epsilon > 10^-2
    k = k + 1;
    r = r - 0,1;
    x = x - r * grd(x);
    gd = grd(x);
    epsilon = norm(gd);
end
min = x
iter = k
```

```
function f = grd(x)
    f(1) = 2 * x(1)
    f(2) = 2 * x(2)
end
```

Exemple : $f(x) = 4x^2 + e^x$

$\nabla f(x) = f'(x) = 8x + e^x$

$f \in C^1(\mathbb{R})$

On considère $x_0 = 1,5$

$\delta = 0,02$, $\varepsilon = 10^{-3}$

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

$x_1 = 1,5 - 0,02 \nabla f(1,5) = 1,17$

$x_2 = 1,17 - 0,02 \nabla f(1,17) = 0,91$

$\vdots$

$x_{30} = -0,10 - 0,02 \nabla f(-0,10) = -0,10$

$|x_{30} - x_{29}| = 8,5 \times 10^{-4} < \varepsilon$.

examiner la méthode du gradient à pas constant et à pas variable pour

$\delta := \delta - 0,05$ et $\delta = 1$

les cas suivant

1) $f(x,y) = \frac{1}{2}x^2 + \frac{7}{2}y^2$

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

3) $f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle Bx\rangle + d$

$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 4 & 2 \\ 0 & 0 & 2 \end{pmatrix}$, $B = (2, -1, 1)^T$

$d = -5$.

4) $f(x) = \frac{1}{2}\langle Ax, x\rangle + \langle Bx\rangle$

$A = \begin{pmatrix} 10 & -2 \\ -2 & 10 \end{pmatrix}$

$B = (3, 3)^T$.

Méthode du gradien à pas optimal : Cette methode propose un choix du pas qui rend la fonction objective minimale le long de la direction de desente choisie, c-à-d, on cherche le pas ρ_k de la suite $x_{k+1} = x_k + \rho_k d_k$

qui réalise le minimum sur $\mathbb{R}^+$ de la fonction suivante :

$$\mathbb{R}^+ \longrightarrow \mathbb{R} \; ; \quad \phi_k(\rho) = J(x_k + \rho\, d_k)$$

sous la condition de la continûment differentiable de f

Exemple : On considère le problème de minimisation sans contrainte suivant

$$\begin{cases} \min \frac{1}{2}x^2 + \frac{7}{2}y^2 \\ (x,y) \in \mathbb{R}^2 \end{cases}$$

On suppose que $X^k = (x_k, y_k)_{k \geq 0}$ alors, d'après l'algorithme du gradient $X^{k+1} = X^k - \rho \nabla f(X^k)$, où $f(x,y) = \frac{1}{2}x^2 + \frac{7}{2}y^2$

$\nabla f(x,y) = (x, 7y)$ alors $\nabla f(X^k) = \nabla f(x^k, y^k) = (x^k, 7y^k)$

$$X^{k+1} = X^k - \rho(x^k, 7y^k) = (x^k, y^k) - \rho(x^k, 7y^k)$$

$$\Leftrightarrow \quad (x^{k+1}, y^{k+1}) = (x^k - \rho x^k, y^k - 7\rho y^k)$$

$$= (x_k(1-\rho), y^k(1-7\rho))$$

On calcule maintenant le pas optimal, on pose

$$\phi(\rho_p) = f[x_k(1-\rho_p), y^k(1-7\rho_p)]$$

$$= \frac{x_k^2}{2}(1-\rho_p)^2 + \frac{7}{2}y^{k^2}(1-7\rho_p)^2.$$

$$\phi'(\rho_p) = x_k^2(\rho_p - 1) + 7y_k^2(7\rho_p - 1).$$

Le min de ϕ est caractérisé par $\phi'(\rho_p) = 0$

c-à-d $\phi'(\rho_p) = 0 \Leftrightarrow x_k^2(\rho_p - 1) + 49y_k^2(7\rho_p - 1) = 0$

$$\Leftrightarrow \rho_p = \frac{x_k^2 + 7^2 y_k^2}{x_k^2 + 7^3 y_k^2}.$$

le schema de l'algorithme prend la forme :

$$(x_{k+1}, y_{k+1}) = (x_k, y_k) - \frac{x_k^2 + 7^2 y_k^2}{x_k^2 + 7^3 y_k^3} (x^k, 7 y_k).$$

· Règle d'Armijo :

α_k st vérifie la règle d'Armijo si On a :

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \alpha_k \beta_1 \nabla^T f(x_k) d_k \quad 0 < \beta_1 < 1 \ ,,$$

$\beta_1 = 0,9$ généralement

Remarques

1) si On Pose $\phi(\alpha_k) = f(x_k + \alpha_k d_k) \quad (\phi : \mathbb{R}^+ \longrightarrow \mathbb{R})$

On a $\phi(\alpha_k)$ st developpable au voisinage de "0" d'ordre 1" et On a :

$$\phi(\alpha_k) = \phi(0) + \alpha_k \phi'(0) \quad \text{ou} \quad \begin{cases} \phi(0) = f(x_k) \\ \phi'(0) = \langle d_k, \nabla J(x_k) \rangle \end{cases}$$

$$= \phi(0) + \alpha_k \nabla^T J(x_k) d_k .$$

2) si α_k Verifie la regle d'Armijo, On a

$$f(x_k + \alpha_k d_k) < f(x_k) .$$

Algorithme de la règle d'Armijo

① Initialisation

② Choisir $\hat{\alpha} \in]0, +\infty[\quad (\hat{\alpha} = 1$ généralement$)$

$0 < \beta_1 < 1 \quad (\beta_1 = 0,9 \quad$ " $)$

③ Calculer $f(x_k + \hat{\alpha} d_k)$

$f(x_k)$

$\nabla^T f(x_k) d_k < 0$

④ Etape Principale

Si $f(x_k + \hat{\alpha} d_k) \leq f(x_k) + \hat{\alpha} \beta_1 \nabla^T f(x_k) d_k$

$\hat{\alpha} = 2\hat{\alpha}$ On va ④ (généralement fois "2" pour chercher le plus grand $\hat{\alpha}$)

Sinon

$\hat{\alpha} = \dfrac{\hat{\alpha}}{2}$ (généralement, On cherche le plus petit $\hat{\alpha}$ vérifie la règle "le premier")

FIN

Exemple

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R}$$
$$(x,y) \mapsto f(x,y) = x^2 + y^4 .$$

On pose $(x_1, y_1) = (1,1)$ et $f(1,1) = 2$

et $d_1 = -\nabla f(1,1) = -\begin{pmatrix} 2x \\ 4y^3 \end{pmatrix}\Big|_{(1,1)} = -\begin{pmatrix} 2 \\ 4 \end{pmatrix}$

$\nabla f(x_k) = 2$ et $\|\nabla f(x_k)\|^2 = 20$.

On pose $\hat{\alpha} = 1$, $\beta_1 = 0{,}9$.

$$x_k + \hat{\alpha} d_k = (x_k, y_k) + \hat{\alpha} d_k$$

$$x_1 + \hat{\alpha} d_1 = (x_1, y_1) + \hat{\alpha} d_1 = (1,1) + \hat{\alpha}(-2,-4)$$
$$= (1 - 2\hat{\alpha}, 1 - 4\hat{\alpha})$$

$$f(x_1 + \alpha d_1) = f(1 - 2\hat{\alpha}, 1 - 4\hat{\alpha}) = (1 - 2\hat{\alpha})^2 + (1 - 4\alpha)^4$$
$$= 256\alpha^4 - 256\alpha^3 + 100\alpha^2 - 20\alpha + 2$$

On pose
$$\phi(\alpha) = f(x_1 + \alpha d_1) = 256\alpha^4 - 256\alpha^3 + 100\alpha^2 - 20\alpha + 2 .$$

On calcule $f(x_k) + \hat{\alpha}\beta_1 \nabla^T f(x_k) d_k$

$$f(x_1, y_1) + \hat{\alpha}\beta_1 \nabla f(x_1, y_1) d_1 = f(1,1) - \alpha\beta_1 \|\nabla f(1,1)\|^2$$
$$= f(1,1) - \hat{\alpha}(0{,}9) \times 20 = 2 - 18\hat{\alpha}$$

On pose $\theta(\alpha) = 2 - 18\alpha$.

$\hat{\alpha}$ vérifie la règle d'Armijo ssi

$$\phi(\hat{\alpha}) \leq \theta(\hat{\alpha})$$

On pose $\hat{\alpha} = 1$ $\quad \phi(1) = 82 \quad \theta(1) = -16$

On a $\quad \phi(\hat{\alpha}) > \theta(\hat{\alpha})$ n'est pas vérifié

On pose $\frac{\hat{\alpha}}{2} = 0{,}5$ $\quad \phi(0{,}5) = 1, \quad \theta(0{,}5) = -7$

$\phi(0{,}5) > \theta(0{,}5)$ n'est pas vérifié!

$\frac{\hat{\alpha}}{2} = 0,25$

$\phi(0,25) = 0,25 \quad \theta(0,25) = -25$.

n'est pas

$\frac{\hat{\alpha}}{2} = 0,125$ n'est pas

$\frac{\hat{\alpha}}{2} = 0,0625$ n'est pas

$\frac{\hat{\alpha}}{2} = 0,03125$ n'est pas

$\frac{\hat{\alpha}}{2} = 0,015625$

$\phi(0,015626) = 1,710952$

$\theta(0,015625) = 1,71875$

$\phi(\hat{\alpha}) < \theta(\hat{\alpha})$

On arrête et on prend $\hat{\alpha} = \alpha_1 = 0,015625$

donc $x_2 = x_1 - \alpha_1 \nabla f(x_1) = (1,1) - 0,015625\,(2,4) =$

et on pose $x_3 = x_2 - \alpha_2 \nabla f(x_2)$, et on examine α_2 pour la règle d'Armijo

Exercice

On considère le pb de minimisation suivant

$$\begin{cases} f(x) = \frac{1}{2}\langle Ax, x\rangle + \langle B, x\rangle \\ x \in \mathbb{R}^2 \end{cases}$$

tel que $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, $B = 10\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

* Calculer les deux premier terme en utilisant la méthode du gradient par le pas d'Armijo.

On détermine premièrement la fonction $\phi(\alpha) = f(x - \alpha \nabla f(x))$

$$f(x,y) = \frac{2}{2}x^2 + \frac{y^2}{2} + xy + 10x + 10y$$

$$= x^2 + \frac{1}{2}y^2 + xy + 10x + 10y$$

$$\nabla f(x,y) = \begin{pmatrix} 2x + y + 10 \\ y + x + 10 \end{pmatrix}.$$

$$\phi(\alpha) = f\left((x,y)^T - \alpha\begin{pmatrix} 2x+y+10 \\ y+x+10\end{pmatrix}\right)$$

$$\phi(\alpha) = f\begin{pmatrix} \underbrace{x - 2\alpha x - \alpha y - 10\alpha}_{1} \\ \underbrace{y - \alpha y - \alpha x - 10\alpha}_{2} \end{pmatrix}$$

$$\phi(\alpha) = (1)^2 + \frac{1}{2}(2)^2 + (1)(2) + 10(1) + 10(2)$$

Pour simplifier les choses on prend

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

et $$\phi(\alpha) = f\begin{pmatrix} (1-2\alpha)x \\ (1-2\alpha)y \end{pmatrix} = (1-2\alpha)^2(x^2+y^2)$$

$$\phi'(\alpha) = -4(1-2\alpha)(x^2+y^2)$$

$$\phi'(0) = -4(x^2+y^2)$$

$$\phi(0) = x^2 + y^2$$

si $\alpha_1 = 1$ $x_0 = (1,1)$; $\rho = 0{,}9$. ($\rho = \beta_1$)

$$\phi(0) - \alpha\rho\phi'(0)$$

$$= 2 + 0{,}9 \times 8 = 9{,}2 > \phi(1) = 1{,}28$$

On pose $\alpha = \frac{0{,}9}{2} = 0{,}45$

Méthode du gradient conjugué :

Déf Soit $A \in M_n(\mathbb{R})$ symétrique et définie positive

On dit que deux vecteurs $x, y \in \mathbb{R}^n$ sont A-conjugués

(ou conjugués par rapport à A) s'il vérifient $x^T A y = 0$. exemple $\begin{pmatrix} 2 \\ -1 \end{pmatrix}$

Exemple : $x = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ $x^T A y = 0 \iff 2y_1 + 4y_2 = 0 \iff 2y_1 = -4y_2 \iff y =$

Remarques

* Si $A=0$, chaque deux vecteur sont $-A$ conjugué. $x^TAy=0$ $\forall x,y \in \mathbb{R}^n$
* Si $A=I$. $x^TAy=0 \Leftrightarrow x^Ty=0$
$\Leftrightarrow$ x et y sont orthogonaux

Proposition Soit A une matrice symétrique définie positive, si $\{d_1, \ldots, d_k\}$ des vecteurs $-A$ conjugués, alors il sont linéairement indépendantes :

Preuve Supposons qu'ils existent $\alpha_i \in \mathbb{R}^*$, $\forall i=\overline{1,k}$ tels que

$$\alpha_1 d_1 + \alpha_2 d_2 + \cdots + \alpha_k d_k = 0 \quad \ldots *$$

en multipliant $(*)$ par A, on obtient

$$\alpha_1 d_1^T + \alpha_2 d_2^T + \cdots + \alpha_k d_k^T = 0 \quad \ldots (**) \quad \left(\text{car } (A+B)^T = A^T + B^T \text{ et } 0^T = 0\right)$$

$$\alpha_1 d_1^T A + \alpha_2 d_2^T A + \cdots + \alpha_k d_k^T A = 0 \quad \ldots (***)$$

et en multipliant $(***)$ par d_i, $\forall i=\overline{1,k}$, on obtient

$$\alpha_1 d_1^T A d_1 + \underbrace{\alpha_2 d_2^T A d_1}_{=0} + \cdots + \underbrace{\alpha_k d_k^T A d_1}_{=0} = 0 \Leftrightarrow \alpha_1 d_1^T A d_1 = 0$$

-A Conjuguées

$$\alpha_1 d_1^T A d_2 + \alpha_2 d_2^T A d_2 + \cdots + \alpha_k d_k^T A d_2 = 0 \Leftrightarrow \alpha_2 d_2^T A d_2 = 0.$$

$$\vdots$$

$$\alpha_1 d_1^T A d_k + \alpha_2 d_2^T A d_k + \cdots + \alpha_k d_k^T A d_k = 0 \Leftrightarrow \alpha_k d_k^T A d_k = 0$$

$$\Rightarrow \alpha_i d_i^T A d_i = 0, \forall i=\overline{1,k} \text{ mais } d_i^T A d_i > 0$$

$$\Rightarrow \alpha_i = 0 \ \forall i=\overline{1,k} \Rightarrow \text{Contradiction}$$

Exemple On considère la matrice symétrique $A = \begin{pmatrix} 3 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{pmatrix}$

On va déterminer les vecteurs qui sont A-conjugués à partir du vecteur $d_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$

1 - On va montrer que A est definie positive, On calcul les determinant de sous matrice principale $\Delta_1 = 3 > 0$, $\Delta_2 = \begin{vmatrix} 3 & 0 \\ 0 & 4 \end{vmatrix} = 12 > 0$

et $\Delta_3 = \begin{vmatrix} 3 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{vmatrix} = 20 > 0$.

$\Rightarrow$ A est definie positive et de plus elle est symetrique ($A^T = A$).

On pose $d_2 = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix}$ et $d_3 = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$, on a :

$$d_1^T A d_2 = 0 \Leftrightarrow (1 \quad 0 \quad 0) \begin{pmatrix} 3 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = 0$$

$$\Leftrightarrow 3d_1 + d_3 = 0 \quad \text{---} \quad (1).$$

si on prend $d_1 = 1$ et $d_2 = 0$, de (1) on obtient $d_3 = -3$

$$\Rightarrow d_2 = (1, 0, -3)^T$$

Pour d_2, on a $d_1^T A d_3 = 0$ et $d_2^T A d_3 = 0$.

$$\begin{cases} d_1^T A d_3 = 0 \\ d_2^T A d_3 = 0 \end{cases} \Leftrightarrow \begin{cases} 3C_1 + C_3 = 0 \\ -6C_2 - 8C_3 = 0 \end{cases} \quad \text{--} \quad (2)$$

si on prend par exemple $c_1 = 1$ le systeme (2) devient

$$\begin{cases} 3 + c_3 = 0 \\ -6c_2 - 8c_3 = 0 \end{cases} \Leftrightarrow \begin{cases} c_3 = -3 \\ c_2 = 4 \end{cases} \Rightarrow d_3^T = (1 \quad 4 \quad -3).$$

<u>Remarque</u> On n'a pas l'unicité des vecteurs d_1, d_2, d_3 qui sont A - conjugués.

La methode : Supposons que f est quadratique strictement convexe

c-à-d $f(x) = q(x) = \frac{1}{2} x^T A x + b^T x + c$, $x \in \mathbb{R}^n$, $A \in M_n(\mathbb{R})$ symetrique definie positive.

$b \in \mathbb{R}^n$, $c \in \mathbb{R}$

cette méthode est de la forme $\begin{cases} x_0 \text{ donné} \\ x_{k+1} = x_k + \alpha_k d_k \quad . \ k \geq 1 \end{cases}$

où α_k est optimal et $d_1, \dots, d_k$ sont q-conjugues.

Calcul de α_k , α_k est optimal par rapport à d_k.

donc α_k est le min $\phi(\gamma)$ ou $\phi(\gamma) = q(x_k + \gamma d_k)$.

$\Rightarrow \phi'(\alpha_k) = 0$

$$\phi'(\alpha_k) = d_k^T \nabla q(x_k + \alpha_k d_k) = d_k^T \nabla q(x_{k+1}) = 0$$

$\nabla q(x) = Ax + b \quad \Rightarrow$

$$= d_k^T (A x_{k+1} + b) = 0$$

$$= d_k^T A x_{k+1} + d_k^T b = d_k^T A (x_k + \alpha_k d_k) + d_k^T b = 0$$

$$= d_k^T A x_k + \alpha_k d_k^T A d_k + d_k^T b = d_k^T (A x_k + b) + \alpha_k d_k^T A d_k = 0$$

$$\Leftrightarrow \alpha_k = - \frac{d_k^T (A x_k + b)}{d_k^T A d_k} \quad \text{----} \quad (1)$$

Théorème (Zoutendijk)

considérons la suite d'itération $(x_k)_{k\geq 0}$ i.e $x_{k+1} = x_k + \alpha_k d_k$ $\forall k\geq 1$
où d_k sont des directions de descentes et α_k est le pas optimal (de Wolfe), si f est continûment différentiable et ∇f est lipschitzienne alors on a :
$$\sum_{k\geq 0} \frac{\left(\nabla^T f(x_k)\, d_k\right)^2}{\|d_k\|^2} < +\infty .$$

Construire les directions A-conjuguées :

Des directions A-conjuguées $d_0, \dots, d_k$ peuvent être générés à partir d'un ensemble de vecteurs linéairement indépendants $h_0, \dots, h_k$ en utilisant la procédure de Gram-Schmidt où $\langle \{d_i\}_{i=0,k} \rangle = \langle \{h_i\}_{i=0,k} \rangle$.
les d_i est construire comme suit :
$$\forall i \geq 0 \qquad d_{i+1} = h_{i+1} + \sum_{m=0}^{i} c_m^{(i+1)} d_m \quad \text{------} \quad (a) \quad \text{et } d_0 = h_0$$
ou $c_m^{(i+1)}$ sont choisis de manière à assurer la A-conjugaison des $d_0, \dots, d_{i+1}$.

Méthode du gradient conjugué dans le cas quadratique :

supposons que la fonction objective est quadratique c-à-d
$$f(x) = q(x) = \tfrac{1}{2} \langle Ax, x\rangle + \langle b, x\rangle .$$
En appliquant la procédure de G-S aux gradient $h_0 = -\nabla q(x_0)$ $h_1 = -\nabla q(x_1), \dots, h_{n-1} = -\nabla q(x_{n-1})$, et d'autre part on a :
$$\nabla q(x) = Ax + b \Rightarrow \nabla^2 q(x) = A .$$
La méthode se termine si $\nabla q(x_k) = 0$.

Construire les directions $d_0, \dots, d_k$:

On écrit ① sous la forme :

$$d_{k+1} = h_{k+1} + \beta_{k+1} d_k \quad k \geq 0 \quad \text{et } d_0 = -\nabla q(x_0)$$

car la direction d_k est obtenue comme combinaison linéaire du gradient et de direction précédente d_{k-1}

c-à-d $\quad d_{k+1} = -\nabla q(x_{k+1}) + \beta_{k+1} d_k \quad \forall k \geq 0$.

et les coefficients β_{k+1} étant choisis pour d_{k+1} soit conjuguée avec les directions précédentes d_k c-à-d $d_{k+1}^T A d_k = 0$.

$$\Rightarrow \left(-\nabla q(x_{k+1}) + \beta_{k+1} d_k\right)^T A d_k = 0$$

$$\Rightarrow -\nabla^T q(x_{k+1}) A d_k + \beta_{k+1} d_k^T A d_k = 0$$

$$\Rightarrow \beta_{k+1} = \frac{\nabla^T q(x_{k+1}) A d_k}{d_k^T A d_k}.$$

et de ①

$$\alpha_k = \frac{-d_k^T (Ax_k + b)}{d_k^T A d_k} = \frac{-d_k^T \nabla q(x_k)}{d_k^T A d_k}.$$

L'algorithme :

Initialisation

x_0 ; $\nabla q(x_0) = Ax_0 + b$; $k = 0$; ε assez petit . $d_0 = -\nabla q(x_0)$

Itérations $\quad \alpha_k = \dfrac{-d_k^T \nabla q(x_k)}{d_k^T A d_k}$;

$x_{k+1} = x_k + \alpha_k d_k$

$\beta_{k+1} = \dfrac{\nabla^T q(x_{k+1}) A d_k}{d_k^T A d_k}$; $\quad d_{k+1} = -\nabla q(x_{k+1}) + \beta_{k+1} d_k$.

$k = k+1$

critère d'arrêt $\|\nabla q(x_k)\| < \varepsilon$.

Théorème Dans le cas où la fonction objective est quadratique

On a $\beta_{k+1}^{HS} = \beta_{k+1}^{PRP} = \beta_{k+1}^{FR} = \beta_{k+1}^{CD} = \beta_{k+1}^{DY}$.

et dans le cas non quadratique, les quantités généralement différentes.

Méthode du gradient conjugué dans le cas général (non quadratique)

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction objective non nécessairement quadratique.

La méthode du gradient conjugué dans ce cas définie par

$$x_{k+1} = x_k + \alpha_k d_k .$$

où $\alpha_k \in \mathbb{R}$ étant déterminé par une recherche linéaire.

et les direction d_k sont définies par la récurrence suivante

$$d_k = \begin{cases} -\nabla q(x_0) \\ -\nabla q(x_k) + \beta_k d_{k-1} & k \geq 1 . \end{cases}$$

où β_k prend l'une des valeurs β_k^{PRP}, β_k^{FR} ; β_k^{CD}.

Exemple 1

On considère $f(x) = q(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$ où

$A = \begin{pmatrix} 5 & 1 \\ 1 & 2 \end{pmatrix}$ et $b = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$

Il est clair que f est continument différentiable et A symétrique et définie positif alors d'après la méthode du gradient conjugué on pose :

$x_0 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \Rightarrow d_0 = -\nabla q(x_0) = -Ax_0 + b = \begin{pmatrix} -5 \\ -3 \end{pmatrix} = d_0$

$$\alpha_0 = -\frac{d_0^T \nabla q(x_0)}{d_0^T A d_0} = -\frac{(-5\ \ -3)\begin{pmatrix} +5 \\ +3 \end{pmatrix}}{(-5\ \ -3)\begin{pmatrix} 5 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} -5 \\ -3 \end{pmatrix}} = \frac{34}{173} .$$

alors $x_{k+1} = x_k + \alpha_k d_0$.

$$x_1 = \overset{= x_0 + \alpha_0 d_0}{x_0 - \alpha_0 \nabla q(x_0)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \frac{34}{173}\begin{pmatrix} -5 \\ -3 \end{pmatrix} = \begin{pmatrix} 0,017 \\ 1,411 \end{pmatrix}$$

$$\beta_1 = \frac{\nabla^T q(x_1) A d_0}{d_0^T A d_0} = \frac{(Ax_1 - b)^T A d_0}{d_0^T A d_0} = \frac{\left[\begin{pmatrix} 5 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 0,017 \\ 1,411 \end{pmatrix} - \begin{pmatrix} 2 \\ 2 \end{pmatrix}\right]^T \begin{pmatrix} 5 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} -5 \\ -3 \end{pmatrix}}{\begin{pmatrix} -5 \\ -3 \end{pmatrix}^T \begin{pmatrix} 5 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} -5 \\ -3 \end{pmatrix}}$$

$$= \frac{\begin{pmatrix} -0,504 & 0,839 \end{pmatrix}\begin{pmatrix} -28 \\ -11 \end{pmatrix}}{\begin{pmatrix} -28 & -11 \end{pmatrix}\begin{pmatrix} -5 \\ -3 \end{pmatrix}} = \frac{4,883}{173} = 0,028$$

$$d_1 = -\nabla q(x_1) + \underbrace{0,028\begin{pmatrix} -5 \\ -3 \end{pmatrix}}_{\beta_1 d_1} = \begin{pmatrix} -0,504 \\ 0,839 \end{pmatrix} + 0,028\begin{pmatrix} -5 \\ -3 \end{pmatrix}$$

$$= \begin{pmatrix} -0,644 \\ 0,755 \end{pmatrix}$$

$$\alpha_1 = -\frac{d_1^T \nabla q(x_1)}{d_1^T A d_1}$$

$$x_2 = x_1 + \alpha_1 d_1$$

$$\beta_2 = \frac{\nabla q(x_2) A d_1}{d_1^T A d_1}$$

$$d_2 = -\nabla q(x_2) + \beta_2 d_1$$

et jusqu'à $\nabla q(x_k) = 0$

Exercice On Considère le pb de minimisation suivant

$$\begin{cases} \min \left\{ \frac{3}{2}x^2 + 2y^2 + \frac{3}{2}z^2 + xy + 2yz - 3x - z \right\} \\ (x, y, z) \in \mathbb{R}^3 \end{cases}$$

à partir du vecteur initial $x_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ calculer les deux premières itérations En utilisant la méthode du gradient Conjugué.

Solution On peut écrire le pb sous forme matricielle suivant

$$\begin{cases} \min \frac{1}{2} Ax^2 - bx \\ x \in \mathbb{R}^3 \end{cases} \qquad x = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \text{ et } A \begin{pmatrix} 3 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{pmatrix}, \; b = \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}$$

$$\nabla f(x_k) = Ax_k - b = \begin{pmatrix} 3 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} - \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 3x_1 + x_3 - 3 \\ 4x_2 + 2x_3 \\ x_1 + 2x_2 + 3x_3 - 1 \end{pmatrix}$$

$$\nabla f(x_0) = Ax_0 - b = -b = \begin{pmatrix} -3 \\ 0 \\ -1 \end{pmatrix}$$

$$d_0 = -\nabla f(x_0) = \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}$$

$$\alpha_0 = -\frac{\nabla f(x_0) d_0}{d_0^T A d_0} = \frac{10}{36} = 0{,}2778 = -\frac{\begin{pmatrix} -3 \\ 0 \\ -1 \end{pmatrix}^T \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}}{\begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}^T A \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}}$$

$$x_1 = x_0 + \alpha_0 d_0 = \begin{pmatrix} 0{,}8334 \\ 0 \\ 0{,}2778 \end{pmatrix}$$

$$\nabla f(x_1) = Ax_1 - b = \begin{pmatrix} -0{,}222 & 0{,}556 & 0{,}668 \end{pmatrix}^T$$

$$\beta_0 = \frac{\nabla f(x_1) A d_0}{d_0^T A d_0} = \frac{2{,}892}{36} = 0{,}08025.$$

$$d_1 = -\nabla f(x_1) + \beta_0 d_0 = \begin{pmatrix} 0{,}46275 \\ -0{,}5556 \\ -0{,}58655 \end{pmatrix}$$

$\vdots$

$$x_3 = x^* = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Méthode de Newton

le principe de la méthode de Newton consiste à minimiser successivement les approximation du second ordre de f, plus précisement si f est deux fois differentiable au sens de Gateaux, elle admet un developpement de Taylor à l'ordre "2" au voisinage de x_k, et on a :

$$f(x) = f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2} \langle \nabla^2 f(x_k)(x - x_k), (x - x_k) \rangle + o(\|x - x_k\|^2)$$

Posons : $q(x) = f(x_k) + \nabla^T f(x_k)(x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k).$

est une forme quadratique de la forme $\frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle + c$ tel que

$$A = \nabla^2 f(x_k),\ b = \nabla^T f(x_k) \text{ et } c = f(x_k)\ [\text{Remplace } x - x_k \to x]$$

donc $\nabla q(x) = Ax + b = \nabla^2 f(x_k)(x - x_k) + \nabla^T f(x_k)$

Soit x_{k+1} l'optimum de q, alors il verifie $\nabla q(x_{k+1}) = 0$. c-à-d

$$\nabla^2 f(x_k)(x_{k+1}-x_k) + \nabla f(x_k) = 0$$

Donc $$\nabla^2 f(x_k)(x_{k+1}-x_k) = -\nabla f(x_k)$$

Alors $$x_{k+1} = x_k - \left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k).$$

L'algorithme de la méthode :

Initialisation : $\varepsilon > 0$ (assez petit), x_0 (le point de départ), $k=1$

Itération

si $\|\nabla f(x_0)\| < \varepsilon$ stop

sinon
$$x_{k+1} = x_k - \left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k)$$

$k = k+1$.

Exemples $f(x,y) = \frac{1}{2}x^2 + x\cos y$ et $x_0 = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} 0 \\ \pi/2 \end{pmatrix}$

$$\nabla f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + \cos y \\ -x\sin y \end{pmatrix} \Rightarrow \nabla f\begin{pmatrix} 0 \\ \pi/2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\nabla^2 f(x,y) = \begin{pmatrix} 1 & -\sin y \\ -\sin y & -x\cos y \end{pmatrix} \Rightarrow \nabla^2 f(0, \pi/2) = \begin{pmatrix} 1 & -1 \\ -1 & 0 \end{pmatrix}$$

$$x_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = x_0 - \left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k)$$

Remarques

1) Cette méthode fonctionne très bien pour les pb de petite dimension ($n \leq 10$) lorsque on peut calculer facilement la matrice hessienne et son inverse, et on a des problèmes pour les grandes tailles.

2) l'inverse d'une matrice n'est pas toujours existe, alors x_{k+1} n'est pas toujours défini.

3) si l'inverse de la matrice hessienne existe, la direction $d_k = -\left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k)$ n'est pas toujours une direction de descente.