Chapitre 3

Algorithmes

3.1 Méthode du gradient

Définition 3.1 *Soit f une fonction définie de $\mathbb{R}^n$ dans $\mathbb{R}$, $x_k \in \mathbb{R}^n$ et $d_k \in \mathbb{R}^n$. On dit que d_k est une direction de descente au point x_k si il existe $\beta > 0$ telle que*

$$f\left(x_k + \alpha d_k\right) < f\left(x_k + \alpha d_k\right), \forall \alpha \in \left]0, \beta\right[.$$

Proposition 3.1 *Soit f $\mathbb{R}^n \to \mathbb{R}$ une fonction différentiable au point $x_k \in \mathbb{R}^n$. Si*

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

alors d_k est une direction de descente.

Principe de la Méthode

On veut minimiser une fonction f. Pour cela on se donne un point de départ arbitraire $x_0 \in \mathbb{R}^n$, pour construire l'itéré suivant x_1 il faut penser qu'on veut se rapprocher du minimum de f, on veut donc que $f(x_1) < f(x_0)$. On cherche alors x_1 sous la forme : $x_1 = x_0 + \alpha_0 d_0$ où $d_k \in \mathbb{R}^n$ et $\alpha_0 > 0$. En pratique, on cherche α_0 et d_0 pour que $f(x_0 + \alpha_0 d_0) < f(x_0)$. On ne peut pas toujours trouver d_0. Quand d_0 existe on dit que c'est une direction de descente et α_0 est le pas de descente. La direction et le pas de descente peuvent être fixes ou changer à chaque itération. Le schéma général d'une méthode de descente est le suivant :

$$\begin{cases} x_0 \in \mathbb{R}^n \text{ donné}, \\ x_{k+1} = x_k + \alpha_k d_k,\ \alpha_k > 0, d_k \in \mathbb{R}^n, \end{cases}$$

où α_k et d_k sont choisis de telle sorte que :

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

Une idée pour trouver une direction descente est de faire un développement de Taylor à l'ordre 2 de la fonction f entre deux itérés x_k et $x_{k+1} = x_k + \alpha_k d_k$. La fonction f est de classe C^1.

$$f\left(x_{k+1}\right) = f\left(x_k + \alpha_k d_k\right) = f\left(x_k\right) + \alpha_k \nabla f\left(x_k\right)^T d_k.$$

Si
$$\nabla f\left(x_k\right)^T d_k < 0,$$
alors
$$f(x_{k+1}) < f(x_k).$$
Donc d_k est une direction descente.

On a :
$$\nabla f\left(x_k\right)^T .d_k = \left\|\nabla f\left(x_k\right)\right\| \left\|d_k\right\| \cos\left(\nabla f\left(x_k\right), d_k\right),$$
si $\cos\left(\nabla f\left(x_k\right), d_k\right) = -1$, on obtient la meilleure localement. Ce qui donne $d_k = -\nabla f\left(x_k\right).$

D'où
$$x_{k+1} = x_k - \alpha_k \nabla f\left(x_k\right).$$

Choix u pas α_k

Méthode du gradient à pas constant (fixe)

On prend $\alpha_k = \alpha$ conatant et
$$x_{k+1} = x_k - \alpha \nabla f\left(x_k\right), \ \alpha > 0.$$
Si α est trés grand on a un risque de divergence.

Si α est trés petit, la convergence est trés lente.

Si α est à gradient lipchitizien de constante L, on choisir $\alpha = \frac{1}{L}$.
$$\left\|\nabla f\left(x\right) - \nabla f\left(y\right)\right\| \leq L \left\|x - y\right\|.$$
Si f est fortement convexe de rapport a, on choisit $\alpha \in \left]0, \frac{2a}{L^2}\right[$.

Méthode du gradient à pas variable

On choisie α_k tel que :
$$\alpha_k > 0, \sum_{k\geq 0} \alpha_k = \infty \text{ et } \sum_{k\geq 0} \alpha_k^2 < \infty.$$

Exemple 3.1 *Soit* $\alpha_k = \dfrac{1}{k+1}$, *alors*
$$\sum_{k\geq 0} \frac{1}{k+1} \text{ diverge et } \sum_{k\geq 0} \frac{1}{\left(k+1\right)^2} \text{ converge.}$$

3.1.1 Méthode du gradient à pas optimal

L'algorithme de gradient à pas optimal est donné pas :

Etape1 : Choisir $x_0 \in \mathbb{R}^n$ et mettre $k = 0$.

Etape2 : Si $\nabla f(x_k) < 0$, mettre $x_k = x^*$. Terminer. Si non aller en (3).

Etape3 : Calculer la direction $d_k = \nabla f(x_k)$.

Etape4 : Calculerα_k, $\alpha_k = \arg\min_{\alpha>0}(f(x_k) - \alpha\nabla f(x_k))$.

Etape5 : Mettre $x_{k+1} = x_k - \alpha_k \nabla f(x_k)$, poser $k = k+1$ et aller à l'étape 2.

Test d'arrêt On fait les test suivant :

$$1.\ \ \|x_{k+1} - x_k\| \leq \varepsilon, \quad \varepsilon = 10^{-6}\|x_0\|,$$

et

$$2.\ \ \|\nabla f(x_{k+1})\| \leq \varepsilon, \quad \varepsilon = 10^{-6}\|\nabla f(x_0)\|.$$

Exemple 3.2 *Soit f la fonction définie pour tout couple (x, y) de $\mathbb{R}^2$ par :*

$$f(x, y) = x^2 + y^2 - 7x - 4y - xy + 35.$$

En démarrant du point initial $x_0 = (x_0, y_0) = (1, 1)$ et en appliquant l'algorithme du gradient , trouver la solution optimale x^* du problème suivant :

$$\begin{cases} \min f(x, y) \\ (x, y) \in \mathbb{R}^2 \end{cases}.$$

Solution : (1) : $x_1 = x_0 + \alpha_0 d_0$ *avec* $\alpha_0 = \arg\min_{\alpha>0}(f(x_0) - \alpha\nabla f(x_0))$ *et* $d_0 = \nabla f(x_0)$.
On calcul le gradient

$$\nabla f(x, y) = \begin{pmatrix} \dfrac{\partial f}{\partial x}(x, y) \\ \\ \dfrac{\partial f}{\partial x}(x, y) \end{pmatrix} = \begin{pmatrix} 2x - y - 7 \\ \\ 2y - x - 4 \end{pmatrix},$$

et

$$\nabla f(x_0) = \nabla f(1, 1) = (-6, -3), \ \ d_0 = -\nabla f(x_0) = (6, 3).$$

On calcul aussi α_0,

$$\alpha_0 = \arg\min_{\alpha>0}(f(x_0) - \alpha\nabla f(x_0)),$$

alors

$$f((x_0) - \alpha\nabla f(x_0)) = f((1, 1) - \alpha(6, 3)) = f(1 + 6\alpha, 1 + 3\alpha),$$

$$\begin{aligned}
& f\left(1+6\alpha,1+3\alpha\right) \\
= \; & \left(1+6\alpha\right)^2+\left(1+3\alpha\right)^2-7\left(1+6\alpha\right)-4\left(1+3\alpha\right)-\left(1+6\alpha\right)\left(1+3\alpha\right)+35 \\
= \; & 27\alpha^2-45\alpha+25,
\end{aligned}$$

$$f\left(1+6\alpha,1+3\alpha\right)^{'}=54\alpha-45=0\Rightarrow\alpha_0=\frac{45}{54}=\frac{5}{6},$$

donc

$$x_1=x_0+\alpha_0 d_0=\begin{pmatrix}1\\1\end{pmatrix}+\frac{5}{6}\begin{pmatrix}6\\3\end{pmatrix}=\begin{pmatrix}6\\\frac{7}{2}\end{pmatrix}.$$

(2) : $x_2=x_1+\alpha_1 d_1, x_1=\left(6,\frac{7}{2}\right)$ *et* $d_1=-\nabla f\left(6,\frac{7}{2}\right)=-\left(\frac{-3}{2},3\right).$

$$\begin{aligned}
& f\left(6-\frac{3}{2}\alpha,\frac{7}{2}+3\alpha\right) \\
= \; & \left(6-\frac{3}{2}\right)^2+\left(\frac{7}{2}+3\alpha\right)^2-7\left(6-\frac{3}{2}\right)-4\left(\frac{7}{2}+3\alpha\right)-\left(6-\frac{3}{2}\right)\left(\frac{7}{2}+3\alpha\right)+35.
\end{aligned}$$

Alors

$$f\left(6-\frac{3}{2}\alpha,\frac{7}{2}+3\alpha\right)^{'}=\frac{63}{2}\alpha-\frac{45}{4}=0\Rightarrow\alpha_1=\frac{5}{14}.$$

donc

$$x_2=x_1+\alpha_1 d_1=\begin{pmatrix}6\\\frac{7}{2}\end{pmatrix}+\frac{5}{14}\begin{pmatrix}\frac{-3}{2}\\3\end{pmatrix}=\begin{pmatrix}5,44\\4,57\end{pmatrix}.$$

(3) : $x_3=x_2+\alpha_2 d_2, x_2=\left(\frac{153}{28},\frac{32}{7}\right)$ *et* $d_2=-\nabla f\left(\frac{153}{28},\frac{32}{7}\right)=-\left(\frac{9}{14},\frac{9}{28}\right).$

$$f\left(\frac{153}{28}+\frac{9}{14}\alpha,\frac{32}{7}+\frac{9}{28}\right)^{'}=0\Rightarrow\alpha_2=\frac{5}{2}.$$

$$x_3=\begin{pmatrix}\frac{153}{28}\\\frac{32}{7}\end{pmatrix}+\frac{5}{2}\begin{pmatrix}\frac{9}{14}\\\frac{9}{28}\end{pmatrix}=\begin{pmatrix}6\\4,83\end{pmatrix}.$$

Remarque 3.1 *L'algorithme du gradient à pas optimal a les propriétés suivantes :*
(a) : *Deux directions succesives de descente sont orthgonales.*
En effet

$$\begin{aligned}
& \frac{d}{d\alpha}\left(\left(f\left(x_k\right)-\alpha_k\nabla f\left(x_k\right)\right)\right)_{\alpha_k=\alpha}=0 \\
\Rightarrow \; & -\nabla f\left(x_k\right)^T\nabla\left(\left(f\left(x_k\right)-\alpha_k\nabla f\left(x_k\right)\right)\right)=0 \\
\Rightarrow \; & \nabla f\left(x_k\right)^T\nabla f\left(x_{k+1}\right)=0.
\end{aligned}$$

(b) : *Cas où* f *est quadratique : i.e* $f\left(x\right)=\frac{1}{2}x^TAx-b^Tx.$

$$\nabla f\left(x\right)=Ax-b \text{ et } \nabla^2 f\left(x\right)=A.$$

Le pas optimal α_k se déduit à partir de la relation suivante :

$$\nabla f(x_k)^T \nabla f(x_{k+1}) = 0 \ldots\ldots\ldots\ldots\ldots (I)$$

$$\begin{aligned}
(I) \quad &\Rightarrow \quad (Ax_k - b)^T (Ax_{k+1} - b) = 0 \\
&\Rightarrow \quad (Ax_k - b)^T (A(x_k - \alpha_k A(Ax_k - b)) - b) = 0 \\
&\Rightarrow \quad (Ax_k - b)^T (Ax_k - b - \alpha_k (Ax_k - b)) = 0 \\
&\Rightarrow \quad \|(Ax_k - b)\|^2 - \alpha_k \|(Ax_k - b)\|_A^2 = 0 \\
&\Rightarrow \quad \alpha_k = \frac{\|(Ax_k - b)\|^2}{\|(Ax_k - b)\|_A^2}.
\end{aligned}$$

On pose :

$$\begin{cases} B_k = Ax_k - b = \nabla f(x_k), \\ \alpha_k = \dfrac{\|B_k\|^2}{\|B_k\|_A^2}. \end{cases}$$

Algorithme (f est quadratique)

Etape 1 : Choiser $x_0 \in \mathbb{R}^n$,

Etape 2 : $x_{k+1} = x_k - \alpha_k B_k$,

$$\alpha_k = \frac{\|B_k\|^2}{\|B_k\|_A^2}, \ B_k = Ax_k - b = \nabla f(x_k).$$

Test d'arrêt On fait les test suivant :

$$\|B_k\| \leq \varepsilon, \ \varepsilon = 10^{-6} \|B_0\|.$$

Exemple 3.3 *Résoudre le problème suivant :*

$$\begin{cases} \min_{(x,y)\in\mathbb{R}^2} f(x,y) = 4x^2 + 4y^2 - 2xy - 6x - 6y, \\ (x_0, y_0) = (0,0). \end{cases}$$

Solution : *On a :*

$$\nabla f(x,y) = \begin{pmatrix} 8x - 2y - 6 \\ 8y - 2x - 6 \end{pmatrix} \ et \ \nabla^2 f(x,y) = \begin{pmatrix} 8 & -2 \\ -2 & 8 \end{pmatrix}.$$

Alors
$$f(x,y)=\frac{1}{2}(x,y)\begin{pmatrix}8 & -2\\ -2 & 8\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}-(6,6)\begin{pmatrix}x\\ y\end{pmatrix},$$
donc f est quadratique. On a :
$$x_{k+1}=x_k-\alpha_k B_k,$$
avec
$$\alpha_k=\frac{\|B_k\|^2}{\|B_k\|_A^2},\ B_k=Ax_k-b=\nabla f(x_k).$$
Pour $k=1,$ on a : $x_1=x_0-\alpha_0 B_0,\ B_0=\nabla f(0,0)=-(6,6)$ et $\alpha_0=\dfrac{\|B_0\|^2}{\|B_0\|_A^2},$
$$\begin{aligned}\|B_0\|^2 &= 36+36=72,\\ & \quad et\\ \|B_0\|_A^2 &= -(6,6)\begin{pmatrix}8 & -2\\ -2 & 8\end{pmatrix}\begin{pmatrix}-6\\ -6\end{pmatrix}=432,\end{aligned}$$
ce qui donne
$$\alpha_0=\frac{\|B_0\|^2}{\|B_0\|_A^2}=\frac{72}{432}=$$
donc
$$x_1=(0,0)+\frac{72}{432}(6,6)=(1,1),$$
et
$$\nabla f(x^*)=\nabla f(1,1)=0<10^{-6}\|B_0\|=7210^{-6}.$$
D'où $x^*=(1,1),\ f(x^*)=f(1,1)=-6.$

3.2 Méthode du gradient conjugué

Soit le problème d'optimisation san containtes suivant :
$$(Pscq):\begin{cases}\min f(x)=\frac{1}{2}x^TAx-b^Tx,\\ \qquad x\in\mathbb{R}^n,\end{cases}$$
où A est une matrice symétrique définie positive sur $\mathbb{R}^n$.

Théorème 3.1 *Le probleme $(Pscq)$ a une solution unique x^*, solution du système linéaire $Ax=b$, c'est à dire que x^*verifie :*
$$x^*=A^{-1}b.$$

Preuve. Exercice. ■

Définition 3.2 *Soit A une matrice symétrique définie positive sur $\mathbb{R}^n$ et d_1 et d_2 deux vecteurs de $\mathbb{R}^n$. On dit que d_1 et d_2 sont $A-$conjugués si*
$$d_1^TAd_2=0.$$

Exemple 3.4 *Considerons la matrice A suivante :*

$$A = \begin{pmatrix} 3 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{pmatrix}.$$

a) : Montrer que A est définie positive.

b) : Construire d_0, d_1, d_2 $A-$conjugués.

Solution : *a) :On calcule les déterminants des mineurs principaux de A.*

$$\begin{aligned} \triangle_1 &= 3 > 0,\ \triangle_2 = \begin{vmatrix} 3 & 0 \\ 0 & 4 \end{vmatrix} = 12 > 0 \\ \triangle_3 &= \det(A) = \begin{vmatrix} 3 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{vmatrix} = 20 > 0. \end{aligned}$$

Donc A est définie positive.

b) : On commence par : $d_0^T = (1,0,0)$ et on construit $d_1^T = (d_{11}, d_{12}, d_{13})$ et $d_2^T = (d_{21}, d_{22}, d_{23})$ tels que :

$$d_0^T A d_1 = 0,\ d_0^T A d_2 = 0 \text{ et } d_1^T A d_2 = 0.$$

Calcul de d_1 :

d_1 doit vérifier $d_0^T A d_1 = 0$

$$d_0^T A d_1 = 0 \Leftrightarrow (1,0,0) \begin{pmatrix} 3 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} d_{11} \\ d_{12} \\ d_{13} \end{pmatrix} = 3d_{11} + d_{13} = 0,$$

si on prend : $d_{11} = 1$et $d_{12} = 0$, alors l'équation précédente implique que $d_{13} = -3d_{11}$, donc $d_1 = (1, 0, -3)$. On peut vérifier facilement que :

$$d_0^T A d_1 = 0 \Leftrightarrow (1,0,0) \begin{pmatrix} 3 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -3 \end{pmatrix} = 0.$$

Calcul de d_2 :

d_2 doit vérifier $d_0^T A d_2 = 0$ et $d_1^T A d_2 = 0$. Alors

$$d_0^T A d_2 = 0 \Leftrightarrow (1,0,0) \begin{pmatrix} 3 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} d_{21} \\ d_{22} \\ d_{23} \end{pmatrix} = 3d_{21} + d_{23} = 0,$$

et

$$d_1^T A d_2 = 0 = 0 \Leftrightarrow (1,0,-3) \begin{pmatrix} 3 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} d_{21} \\ d_{22} \\ d_{23} \end{pmatrix} = -6d_{22} - 8d_{23} = 0.$$

Autrement dit d_2^T doit vérifier le système suivant :

$$\begin{cases} 3d_{21} + d_{23} = 0 \\ \\ -6d_{22} - 8d_{23} = 0. \end{cases}$$

On pend par exemple $d_{21} = 1$, le système précédent devient

$$\begin{cases} 3 + d_{23} = 0 \\ \\ -6d_{22} - 8d_{23} = 0 \end{cases} \Leftrightarrow \begin{cases} d_{23} = -3 \\ \\ -6d_{22} = -24 \end{cases} \Leftrightarrow \begin{cases} d_{23} = -3 \\ \\ d_{22} = 4 \end{cases}.$$

Par conséquent $d_2^T = (1, -3, 4)$.

Remarque 3.2 *On n'a pas unicité des vecteurs d_0, d_1, d_2 $A-$conjugués.*

Définition 3.3 *Soit A une matrice carreé symétrique. Les directions $d_0, d_1, ..., d_n$ sont dites $A-$conjugués si on a :*

$$d_i^T A d_j = 0, \; 0 \leq i, j \leq n.$$

Proposition 3.2 *Soit Q une matrice (n, n) symétrique et définie positive. Si les directions $d_0, d_1, ..., d_k$ avec $0 \leq k \leq n-1$, sont non nulles et A-conjuguées, alors elles sont linéairement indépendants.*

Preuve. Supposons que

$$\alpha_0 d_0 + \alpha_2 d_1 + ... + \alpha_n d_n = 0 \qquad (I).$$

On multiple l'égalité (I) par $d_i^T A$, $0 \leq i \leq n$. on obtient

$$\sum_{j=0}^{n} \alpha_i d_i^T A d_j = \alpha_i d_i^T A d_i = 0,$$

Or

$$\alpha_i d_i^T A d_i = 0 \Rightarrow \alpha_i = 0, \; 0 \leq i \leq n.$$

car A est définie positive et par conséquent $d_i^T A d_i > 0$. Donc le système $\{d_0, d_1, ..., d_n\}$ est libre. ■

Proposition 3.3 *Soit A une matrice symétrique définie positive. Si les vecteurs $d_0, d_1, ..., d_n$ sont deux à deux $A-$conjugués, alors ils forment une base de $\mathbb{R}^n$.*

Soit x^* la solution unique du système $Ax = b$, $(x^* = A^{-1}b)$, puisque $\{d_0, d_1, ..., d_{n-1}\}$ base de $\mathbb{R}^n$ deux à deux $A-$conjugués, on peut écrire :

$$x^* = \sum_{j=0}^{n-1} \alpha_j d_j \qquad (II).$$

En multpliant à gauche de (II) par $d_i^T A$, on obtient :

$$d_i^T A x^* = \sum_{j=0}^{n-1} \alpha_i d_i^T A d_j = \alpha_i d_i^T A d_i = \alpha_i \|d_i\|_A^2,$$

alors

$$\alpha_i = \frac{d_i^T A x^*}{\|d_i\|_A^2} = \frac{d_i^T b}{\|d_i\|_A^2}, \; i = 1, ..., n-1.$$

Problème du construction des directions $\mathbf{d}_i$?

On construit des directions d_i $\left(d_i^T A d_j = 0,\ \forall i \neq j\right)$ comme suit :

* : Fixons x_0 et choisissons $d_0 = \nabla f\left(x_0\right)$.

** : On Note x_1 le vecteur ontenu par l'algorithne du gradient à pas optimal : $x_1 = x_0 - \alpha_0 d_0$, avec

$$\alpha_0 = \frac{\left\|\nabla f\left(x_0\right)\right\|^2}{\left\|\nabla f\left(x_0\right)\right\|_A^2} = \frac{d_0^T \nabla f\left(x_0\right)}{\left\|d_0\right\|_A^2}.$$

On Cherche d_1 comme combinaison linéaire $\nabla f\left(x_1\right)$ et d_0 :

$$d_1 = \nabla f\left(x_1\right) - B_0 d_0.$$

B_0 choisir de telle sorte que d_0 et d_1 sont $A-$conjugués.

$$d_1 = \nabla f\left(x_1\right) - B_0 d_0 \Leftrightarrow d_0^T A d_1 = d_0^T A \nabla f\left(x_1\right) - B_0 d_0^T A d_0 = 0,$$

ce qui implique

$$B_0 = \frac{d_0^T A \nabla f\left(x_1\right)}{d_0^T A d_0} = \frac{d_0^T A \nabla f\left(x_1\right)}{\left\|d_0\right\|_A^2}.$$

Maintenant : $x_2 = x_1 - \alpha_1 d_1$, avec $\alpha_1 = \frac{d_1^T \nabla f(x_1)}{\|d_1\|_A^2}$ et on cherche d_2 comme suit :

$$d_2 = \nabla f\left(x_2\right) - B_1 d_1,\ \text{qui soit } A - \text{conjugués avec } d_0 \text{ et } d_1,$$

avec

$$B_1 = \frac{d_1^T A \nabla f\left(x_2\right)}{\left\|d_1\right\|_A^2}.$$

Algorithme du gradient conjugué

Etape 1 : Choiser $x_0 \in \mathbb{R}^n$,

Etape 2 : $x_{k+1} = x_k - \alpha_k d_k,\ \alpha_k = \dfrac{d_k^T \nabla f\left(x_k\right)}{\left\|d_k\right\|_A^2}$,

$$d_k = \nabla f\left(x_{k+1}\right) - B_k d_k,\ B_k = \frac{d_k^T A \nabla f\left(x_{k+1}\right)}{\left\|d_k\right\|_A^2}.$$

Exemple 3.5 *Résoudre le problème suivant :*

$$\begin{cases} \min_{(x,y)\in\mathbb{R}^2} f\left(x,y\right) = \dfrac{3}{2}x^2 + \dfrac{1}{2}y^2 - xy - 2x, \\ \\ \left(x_0, y_0\right) = \left(0,0\right), \end{cases}$$

en utilisant l'algorithme du gradient conjugué.

Solution : *On a :*

$$f\left(x,y\right) = \frac{1}{2}\left(x,y\right)\begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} - \left(2,0\right)\begin{pmatrix} x \\ y \end{pmatrix}$$

et

$$\nabla f(x,y) = (3x + -y - 2, 2y - x), \nabla f(0,0) = (-2,0).$$

Calcul de x_1 :

$$x_1 = x_0 - \alpha_0 d_0,\ \alpha_0 = \frac{d_0^T \nabla f(x_0)}{\|d_0\|_A^2} = \frac{4}{12} = \frac{1}{3}, d_0 = \nabla f(0,0) = (-2,0),$$

donc

$$x_1 = (0,0) - \frac{1}{3}(-2,0) = \left(\frac{2}{3}, 0\right).$$

Calcul de x_2 :

$$\begin{aligned}
x_2 &= x_1 - \alpha_1 d_1, \alpha_1 = \frac{d_1^T \nabla f(x_1)}{\|d_1\|_A^2}, \nabla f(x_1) = \nabla f\left(\frac{2}{3}, 0\right) = \left(0, -\frac{2}{3}\right), \\
d_1 &= \nabla f(x_1) - \alpha_0 B_0,\ B_0 = \frac{d_0^T A \nabla f(x_1)}{\|d_0\|_A^2}, \\
d_0^T A \nabla f(x_1) &= (-2,0)\begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ -\frac{2}{3} \end{pmatrix} = \frac{-4}{3}, \\
\|d_0\|_A^2 &= (-2,0)\begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} -2 \\ 0 \end{pmatrix} = 12,
\end{aligned}$$

$$\begin{aligned}
d_1 &= \nabla f(x_1) - \alpha_0 B_0 = \left(0, -\frac{2}{3}\right) + \frac{1}{9}(-2,0) = \left(-\frac{2}{9}, -\frac{2}{3}\right), \\
d_1^T \nabla f(x_1) &= \left(-\frac{2}{9}, -\frac{2}{3}\right)\begin{pmatrix} 0 \\ -\frac{2}{3} \end{pmatrix} = \frac{4}{9}, \\
\|d_1\|_A^2 &= \left(-\frac{2}{9}, -\frac{2}{3}\right)\begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} -\frac{2}{9} \\ -\frac{2}{3} \end{pmatrix} = \frac{8}{27}, \\
\alpha_1 &= \frac{d_1^T \nabla f(x_1)}{\|d_1\|_A^2} = \frac{3}{2}.
\end{aligned}$$

donc

$$x_2 = x_1 - \alpha_1 d_1 = \left(\frac{2}{3}, 0\right) + \frac{3}{2}\left(\frac{2}{9}, \frac{2}{3}\right) = (1,1),$$

et

$$\nabla f(x_2) = \nabla f(1,1) = 0.$$

D'où $x^* = (1,1),\ f(x^*) = f(1,1) = -1.$

Remarque 3.3 *L'algorithme du gradient conjugué converge ver la solution en* $n-$*itérations.*

Théorème 3.2 *Les directions* $\{d_0, d_1, ..., d_{n-1}\}$ *engendrées par l'algorithme du gradient conjugué sont* $A-$*conjuguées.*

Preuve. La preuve se fait par récurence.

(i) : On montre que la $A-$conjuguées est vraie pour $k=1$, i,e. $d_0^T A d_1 = 0$.

En effet. On a :

$$B_0 = \frac{d_0^T A \nabla f(x_1)}{\|d_0\|_A^2},\ d_1 = \nabla f(x_1) - B_0 d_0,$$

donc

$$\begin{aligned} d_0^T A d_1 &= d_0^T A (\nabla f(x_1) - B_0 d_0) = d_0^T A \nabla f(x_1) - B_0 d_0^T A d_0 \\ &= d_0^T A \nabla f(x_1) - d_0^T A d_0 \frac{d_0^T A \nabla f(x_1)}{\|d_0\|_A^2} = d_0^T A \nabla f(x_1) - d_0^T A \nabla f(x_1) = 0. \end{aligned}$$

(ii) : On Suppose que $\{d_0, d_1, ..., d_k\}, k < n-1$ sont $A-$conjuguées et on montre que $\{d_0, d_1, ..., d_{k+1}\}, k < n-1$ sont $A-$conjuguées, c-à-d on montre que

$$d_j^T A d_{k+1} = 0,\ j \in \{0, ..., k\}. \qquad (I)$$

En effet $\{0, ..., k\} = \{0, ..., k-1\} \cup \{k\}$

(ii_1) : La relation (I) est vraie pour $j = k$.

En effet. On a :

$$B_k = \frac{d_k^T A \nabla f(x_{k+1})}{\|d_k\|_A^2},\ d_{k+1} = \nabla f(x_{k+1}) - B_k d_k,$$

donc

$$\begin{aligned} d_k^T A d_{k+1} &= d_k^T A (\nabla f(x_{k+1}) - B_k d_k) = d_k^T A \nabla f(x_{k+1}) - B_k d_k^T A d_k \\ &= d_k^T A \nabla f(x_{k+1}) - d_k^T A d_k \frac{d_k^T A \nabla f(x_{k+1})}{\|d_k\|_A^2} \\ &= d_k^T A \nabla f(x_{k+1}) - d_k^T A \nabla f(x_{k+1}) = 0. \end{aligned}$$

(ii_2) : La relation (I) est vraie pour $j \in \{0, ..., k-1\}$.

En effet. Soit $j \in \{0, ..., k-1\}$,

$$d_{k+1}^T A d_j = (\nabla f(x_{k+1}) - B_k d_k)^T A d_j = \nabla f(x_{k+1})^T A d_j - B_k d_k^T A d_j \qquad (II).$$

Puisque

$$d_k^T A d_j = 0,\ j \in \{0, ..., k-1\}.$$

Par conséquent (II) devient

$$d_{k+1}^T A d_j = \nabla f(x_{k+1})^T A d_j,\ j \in \{0, ..., k-1\}.$$

On a :

$$\nabla f(x_{j+1}) - \nabla f(x_j) = \alpha_j A d_j, j = 0, ..., k-1,$$

par conséquent

$$A d_j = \frac{1}{\alpha_j} (\nabla f(x_{j+1}) - \nabla f(x_j)),$$

La relation (II) donne

$$\begin{aligned}
d_{k+1}^{T}Ad_{j} &= \frac{1}{\alpha_{j}}d_{k+1}^{T}\left(\nabla f\left(x_{j+1}\right)-\nabla f\left(x_{j}\right)\right) \\
&= \frac{1}{\alpha_{j}}\nabla f\left(x_{j+1}\right)^{T}\left(\nabla f\left(x_{j+1}\right)-\nabla f\left(x_{j+1}\right)^{T}\nabla f\left(x_{j}\right)\right) \\
&= \frac{1}{\alpha_{j}}\nabla f\left(x_{j+1}\right)^{T}\nabla f\left(x_{j+1}\right)-\frac{1}{\alpha_{j}}\nabla f\left(x_{j+1}\right)^{T}\nabla f\left(x_{j}\right)=0.
\end{aligned}$$

Par conséquent la relation (II) est vraie pour $j\in\{0,...,k-1\}$.

(ii_1)) et ii_2) impliquent que la relation (II) est vraie pour $j\in\{0,...,k\}$. Par conséquent les directions $\{d_0,d_1,...,d_{n-1}\}$ sont $A-$conjugués. ■

Théorème 3.3 *Partant d'un point initial $x_0\in\mathbb{R}^N$ quelconque, l'algorithme du gradient conjugué précédent converge vers la solution optimale unique x^* du probleme $(Pscq)$ dans $n-$iterations, c'est à dire qu'on a :*

$$x_n^*=x^* \text{ et } Ax_n^*=Ax^*=b.$$

Preuve. Exercice. ■

3.3 Méthode de relaxation

Principe de la méthode : Minimiser la fonction f " composante par composante".

i.e Les directions de descentes successives sont les axes de $\mathbb{R}^n$.

On suppose que $x^{(k)}=\left(x_1^{(k)},x_2^{(k)},...,x_n^{(k)}\right)$ est connu et on cherche $x^{(k+1)}=\left(x_1^{(k+1)},x_2^{(k+1)},...,x_n^{(k+1)}\right)$ comme suit :

$$\begin{aligned}
\left(x_1^{(k+1)},x_2^{(k)},...,x_n^{(k)}\right) &\leftarrow \arg\min_{\varphi\in\mathbb{R}} f\left(\varphi,x_2^{(k)},...,x_n^{(k)}\right), \\
\left(x_1^{(k+1)},x_2^{(k+1)},...,x_n^{(k)}\right) &\leftarrow \arg\min_{\varphi\in\mathbb{R}} f\left(x_1^{(k+1)},\varphi,...,x_n^{(k)}\right), \\
&\vdots \\
\left(x_1^{(k+1)},x_2^{(k+1)},...,x_n^{(k+1)}\right) &\leftarrow \arg\min_{\varphi\in\mathbb{R}} f\left(x_1^{(k+1)},x_2^{(k+1)},...,x_{n-1}^{(k+1)},\varphi\right).
\end{aligned}$$

Algorithme de la méthode de relaxation

Etape 1 : Choiser $x_0\in\mathbb{R}^n$,

Etape 2 : Pour $i=1,...,n$, on calcule la solution $x_{1+1}^{(k+1)}$:

$$\left(x_1^{(k+1)},x_2^{(k+1)},...,x_{i-1}^{(k+1)},x_{i+1}^{(k)},...,x_n^{(k)}\right)\leftarrow \arg\min_{\varphi\in\mathbb{R}} f\left(x_1^{(k+)},x_2^{(k+1)},...,x_{i-1}^{(k+1)},\varphi,...,x_n^{(k)}\right).$$

Test d'arrêt On fait le test suivant :

$$1.\ \|\nabla f(x_{k+1})\| \ < \ \varepsilon,\ \varepsilon: \text{ donnée,}$$

$$2.\ \|x_{k+1} - x_k\| \ < \ \varepsilon,\ \varepsilon: \text{ donnée.}$$

Exemple 3.6 *On considère le problème d'optimisation suivant :*

$$\begin{cases} \min f(x_1, x_2) = x_1^2 + x_2^2 + x_1 x_2 \\ (x_1, x_2) \in \mathbb{R}^2. \end{cases}$$

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

En effet, on a :

$$\left(x_1^{(1)}, 1\right) \leftarrow \arg\min_{\varphi \in \mathbb{R}} f(\varphi, 1),$$

alors

$$f(\varphi, 1) = \varphi^2 + 1 + \varphi,\ \ f(\varphi, 1)' = 2\varphi + 1 = 0,$$

donc

$$x_1^{(1)} = \varphi = -\frac{1}{2}.$$

aussi

$$\left(-\frac{1}{2}, x_2^{(1)}\right) \leftarrow \arg\min_{\varphi \in \mathbb{R}} f\left(-\frac{1}{2}, \varphi\right),$$

alors

$$f\left(-\frac{1}{2}, \varphi\right) = \frac{1}{4} + \varphi^2 - \frac{1}{2}\varphi,\ \ f\left(-\frac{1}{2}, \varphi\right)' = 2\varphi - \frac{1}{2} = 0,$$

ce qui donne

$$x_2^{(1)} = \varphi = \frac{1}{4},$$

donc

$$x^{(1)} = \left(x_1^{(1)}, x_2^{(1)}\right) = \left(-\frac{1}{2}, \frac{1}{4}\right).$$

Maintenant on calcule $x^{(2)} = \left(x_1^{(2)}, x_1^{(2)}\right)$:

$$\left(x_1^{(2)}, \frac{1}{4}\right) \leftarrow \arg\min_{\varphi \in \mathbb{R}} f\left(\varphi, \frac{1}{4}\right),$$

on a :

$$f\left(\varphi, \frac{1}{4}\right) = \varphi^2 + \frac{1}{16} + \frac{1}{4}\varphi,$$

et

$$f\left(\varphi, \frac{1}{4}\right)' = 2\varphi + \frac{1}{4} = 0,$$

alors

$$\varphi = -\frac{1}{8} = x_1^{(2)}.$$

Pour $x_1^{(2)}$*, on a :*

$$\begin{aligned}\left(-\frac{1}{8}, x_2^{(2)}\right) &\leftarrow \arg\min_{\varphi \in \mathbb{R}} f\left(-\frac{1}{8}, \varphi\right), \\ f\left(-\frac{1}{8}, \varphi\right) &= \frac{1}{64} + \varphi^2 - \frac{1}{8}\varphi, \\ f\left(-\frac{1}{8}, \varphi\right)' &= 2\varphi^2 - \frac{1}{8} = 0,\end{aligned}$$

alors

$$\varphi = \frac{1}{16} = x_2^{(2)}.$$

Donc

$$x^{(2)} = \left(x_1^{(2)}, x_1^{(2)}\right) = \left(-\frac{1}{8}, \frac{1}{16}\right).$$

$$\begin{aligned}\nabla f(x_1, x_2) &= \begin{pmatrix} 2x_1 + x_2 \\ x_1 + 2x_2 \end{pmatrix}, \nabla f\left(-\frac{1}{8}, \frac{1}{16}\right) = \begin{pmatrix} -\frac{1}{4} + \frac{1}{16} \\ -\frac{1}{8} + \frac{1}{8} \end{pmatrix} = \begin{pmatrix} \frac{3}{16} \\ 0 \end{pmatrix}, \\ \left\|\nabla f\left(x^{(2)}\right)\right\| &< \varepsilon = 10^{-6}\left\|\nabla f\left(x^0\right)\right\|.\end{aligned}$$

D'où $x^* = x^{(2)}$, $f(x^*) = 0.011719$.

3.4 Méthode de Newton

Considérons le problème d'optimisation sans contraintes :

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

où f une fonction définie de $\mathbb{R}^n$ dans $\mathbb{R}$ de classe C^2.

Le principe de la méthode de Newton consiste à minimiser successivement les approximations du second ordre de la foction f, plus précisément si

$$f(x) = f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0), \; x_0 \in \mathbb{R}^n.$$

La minimisation de la fonction f donne le point x_1 solution du système : $\nabla f(x) = 0$. Alors

$$\nabla f(x_1) = \nabla f(x_0) + \nabla^2 f(x_0)(x_1 - x_0) = 0,$$

ce qui donne

$$\nabla^2 f(x_0)(x_1 - x_0) = -\nabla f(x_0),$$

ce qui impliqe

$$x_1 = x_0 - \left[\nabla^2 f(x_0)\right]^{-1} \nabla f(x_0), \quad \left(\nabla^2 f(x_k) : \text{ inversible}\right),$$

à étape $k+1$, on a

$$x_{k+1} = x_k - \left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k), \quad \left(\nabla^2 f(x_k) : \text{ inversible}\right).$$

Algorithme de la méthode de Newton

Etape 1 : Choiser $x_0 \in \mathbb{R}^n$,

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

Test d'arrêt On fait le test suivant :

$$\|\nabla f(x_{k+1})\| < \varepsilon, \ \varepsilon : \text{ donnée.}$$

Si f est quadratique, ona :

$$x_{k+1} = x_k - A^{-1}(Ax_k - b) = x_k - A^{-1}Ax_k - A^{-1}b = A^{-1}b.$$

Dans ce cas la méthode de Newton converge en une seule itération.

Exemple 3.7 *Considérons la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par :*

$$f(x,y) = x^2 + 2y^2, \ x_0 = (2,2).$$

On a :

$$\nabla f(x,y) = \begin{pmatrix} 2x \\ 4y \end{pmatrix}, \nabla f(x_0) = \nabla f(2,2) = \begin{pmatrix} 4 \\ 8 \end{pmatrix},$$

et

$$\nabla^2 f(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}, \quad \left[\nabla^2 f(x_0)\right]^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{4} \end{pmatrix},$$

alors

$$\begin{aligned} x_1 &= x_0 - \left[\nabla^2 f(x_0)\right]^{-1} \nabla f(x_0) \\ &= \begin{pmatrix} 2 \\ 2 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{4} \end{pmatrix} \begin{pmatrix} 4 \\ 8 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned}$$

et

$$\nabla f(0,0) = 0.$$

D'où $x^* = (0,0)$, $f(x^*) = f(0,0) = 0$.

Théorème 3.4 *Si f est une forme quadratique définie positive, alors la méthode de Newton convergence vers la solution optimale en une seule itération $\forall x_0 \in \mathbb{R}^n$.*
Preuve. *Soit $f(x) = \frac{1}{2}x^T Ax - b^T x$, A est une matrice symétrique définie positive sur $\mathbb{R}^n$. On a :*

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

Alors

$$\begin{aligned} x_1 &= x_0 - \left[\nabla^2 f(x_0)\right]^{-1} \nabla f(x_0) \\ x_0 - A^{-1}(Ax_o - b) &= x_0 - A^{-1}Ax_0 - A^{-1}b \\ &= A^{-1}b. \end{aligned}$$

■

3.5 Exercices corrigés

Exercice 3.1 *On considére la fonctionla fonction $f : \mathbb{R}^2 \to \mathbb{R}$, définie par :*

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

et le point initial,

$$x_0 = (x_0, y_0) = (2, 3).$$

En appliquant l'algorithme du gradient à pas optimal résoudre le problème suivant :

$$\begin{cases} \min f(x, y) \\ (x, y) \in \mathbb{R}^2 \end{cases}.$$

Solution : 1 : On a : $x_{k+1} = x_k - \alpha_k \nabla f(x_k), \alpha_k = \arg\min_{\alpha>0} (f(x_k) - \alpha \nabla f(x_k))$,
et

$$\nabla f(x, y) = \begin{pmatrix} \frac{\partial f}{\partial x}(x, y) \\ \\ \frac{\partial f}{\partial x}(x, y) \end{pmatrix} = \begin{pmatrix} 8x - 4y \\ \\ 4y - 4x \end{pmatrix},$$

$$\nabla f(x_0) = \nabla f(2, 3) = (4, 4),\ d_0 = -\nabla f(x_0) = -(4, 4).$$

Pour $k = 0$, on a,

$$\begin{aligned} f((x_0) - \alpha \nabla f(x_0)) &= f((1, 1) - \alpha(6, 3)) = f(2 - 4\alpha, 3 - 4\alpha) \\ &= 4(2 - 4\alpha)^2 - 4(2 - 4\alpha)(3 - 4\alpha) + 2(3 - 4\alpha)^2, \end{aligned}$$

alors
$$f\left(2-4\alpha,3-4\alpha\right)'=0\Rightarrow\alpha=\frac{1}{2}=\alpha_0,$$
donc
$$x_1=x_0+\alpha_0d_0=\left(2,3\right)-\frac{1}{2}\left(4,4\right)=\left(0,1\right).$$
Maintenant en calculant le gradient de f au point x_1, on trouve :
$$\nabla f\left(x_1\right)=\nabla f\left(0,1\right)=\left(4,4\right),\ d_0=-\nabla f\left(x_0\right)=-\left(-4,4\right).$$
Pour $k=1$, on a,
$$\begin{aligned} f\left(\left(x_1\right)-\alpha\nabla f\left(x_1\right)\right) &= f\left(\left(0,1\right)+\alpha\left(4,-4\right)\right)=f\left(4\alpha,1-4\alpha\right) \\ &= 4\left(4\alpha\right)^2-4\left(4\alpha\right)\left(1-4\alpha\right)+2\left(1-4\alpha\right)^2, \end{aligned}$$
alors
$$f\left(4\alpha,1+4\alpha\right)'=0\Rightarrow\alpha=\frac{1}{10}=\alpha_1,$$
donc
$$x_2=x_1+\alpha_1d_1=\left(0,1\right)-\frac{1}{10}\left(4,-4\right)=\left(\frac{2}{5},\frac{3}{5}\right).$$
De la même manière on trouve : $x_3=\left(0,\frac{1}{5}\right)$.

Exercice 3.2 *Soit la forme quadratique $f:\mathbb{R}^n\to\mathbb{R}$ définie par :*
$$f\left(x\right)=\frac{1}{2}x^TAx-b^Tx,\ \forall x\in\mathbb{R}^n,b\in\mathbb{R}^n,$$
A : *Matrice symétrique et définie positive.*
1 : *Montrer que si les directions $d_1,d_2,...,d_n$ sont $A-$conjuguées, alors ils sont linéairement indépendantes.*
2 : *Pour $n=2$, on a : $f\left(x,y\right)=\frac{3}{2}x^2+\frac{1}{2}y^2-xy-2x$.*
2.1 : *Calculer la matrice Hessienne (A) de $f\left(x,y\right)$.*
2.2 : *Montrer que A est définie positive.*
2.3 : *Construire d_0,d_1,d_2 $A-$conjugués.*
3 : *En appliquant l'algorithme du gradient à pas optimal (effectuer trois iterations) déterminer le minimum du problème suivant :*
$$(I)\ \begin{cases}\min f\left(x,y\right)=x^2+2y^2-2xy+2y+2 \\ \left(x,y\right)\in\mathbb{R}^2\end{cases},\ \textit{avec}\ \left(x,y\right)=\left(0,0\right).$$

Solution : 1 : Supposons que
$$\beta_1d_1+\beta_2d_2+...+\beta_nd_n=0, \tag{3.1}$$

Multiplions l'égalité (1) par $d_j^T A$, $1 \leq j \leq n$, on trouve :

$$\sum_{i=1}^{n} \beta_j d_j^T A d_i = \beta_j d_j^T A d_j = 0.$$

Or

$$\beta_j d_j^T A d_j = 0 \Rightarrow \beta_j = 0,\ j = 1, .., n,$$

car A est définie positive et par conséquent $d_j^T A d_j > 0$. Donc $d_1, d_2, ..., d_n$ sont linéairement indépendantes.

2.1 : $H_f = A = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}$.

2.2 : $\Delta_1 = 3 > 0$, $\Delta_2 = 2 > 0 \Rightarrow A : D.P.$

2.3 : on pose $d_0 = (1, 0)$, $d_1 = (d_{11}, d_{12}) =?$, $d_2 = (d_{21}, d_{22}) =?$.

On a : $d_1^T A d_0 = 0 \Rightarrow (d_{11}, d_{12}) \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 3d_{11} - d_{12} = 0.$

$\Rightarrow \begin{cases} d_{11} = \frac{1}{3} \\ d_{12} = 1 \end{cases}$, $d_1 = \left(\frac{1}{3}, 1\right)$.

$d_2^T A d_1 = (d_{21}, d_{22}) \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{3} \\ 1 \end{pmatrix} = d_{22} = 0 \Rightarrow \begin{cases} d_{21} = 2 \\ d_{22} = 0 \end{cases}$, $d_2 = (2, 0)$.

$(3.1.a)$: $\nabla f(x_0) = (0, -2)$, $\alpha_0 = \arg\min_{\alpha > 0} (f x_0 - \alpha \nabla f(x_0)) = \frac{1}{4}$,

$x_1 = x_0 - \alpha_0 \nabla f(x_0) = \begin{pmatrix} 0 \\ -\frac{1}{2} \end{pmatrix}$.

$(3.1.b)$: $\nabla f(x_1) = \begin{pmatrix} 0 \\ -\frac{3}{2} \end{pmatrix}$, $\alpha_1 = \arg\min_{\alpha > 0} (f(x_1 - \alpha \nabla f(x_1))) = \frac{1}{2}$,

$x_2 = x_1 - \alpha_1 \nabla f(x_1) = \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$.

$(3.1.c)$: $\nabla f(x_2) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $\alpha_2 = \arg\min_{\alpha > 0} (f(x_2 - \alpha \nabla f(x_2))) = \frac{1}{4}$.

$x_3 = x_2 - \alpha_2 \nabla f(x_2) = \begin{pmatrix} -\frac{1}{2} \\ -\frac{3}{4} = 0.75 \end{pmatrix}$.

Exercice 3.3 *Soit $a \in \mathbb{R}$. On Considére la fonction $f_a : \mathbb{R}^2 \to \mathbb{R}$ définie par :*

$$f_a(x_1, x_2) = x_1^2 + x_2^2 + a x_1 x_2 - 2x_1 - 2x_2.$$

1 : *Etudier la convexité de la fonction f_a selon les valeurs de a.*

2 : *Déterminer en fonction des valeurs du paramètre a les points critiques de la fonction f_a et préciser leur nature.*

4. *On suppose que $a = 1$. En démarrant du point initial $x_0 = (0, 1)$ et en appliquant l'algorithme du gradient conjugué, trouver la solution optimale x^* du problème suivant :*

$$(I) \begin{cases} \min f_a(x_1, x_2) \\ (x_1, x_2) \in \mathbb{R}^2 \end{cases}.$$

Solution : 1 : On a : $\nabla^2 f(x) = \begin{pmatrix} 2 & a \\ a & 2 \end{pmatrix} \Rightarrow \begin{cases} \Delta_1 = 2 \\ \Delta_2 = 4 - a^2 \end{cases}$

donc f est convexe si $a \in [-2, 2]$.

2 : Les points critiques : $\nabla f(x) = \begin{pmatrix} 2x + ay - 2 \\ 2y + ax - 2 \end{pmatrix} = 0 \Rightarrow \begin{cases} 2x + ay - 2 = 0 \\ 2y + ax - 2 = 0 \end{cases} \Rightarrow$ $\begin{cases} x = y \\ x = \frac{a}{2+a} \end{cases}$,

la nature du points critiques : $\nabla^2 f(x) = \begin{pmatrix} 2 & a \\ a & 2 \end{pmatrix}$.

si $a \in]-2, 2[\Rightarrow \begin{cases} \Delta_1 > 0 \\ \Delta_2 > 0 \end{cases} \Rightarrow \nabla^2 f$ est D.P sur $\mathbb{R}^2 \Rightarrow \left(\frac{a}{2+a}, \frac{a}{2+a}\right)$ est un minimum local strict de f

et comme f est convexe alors le point $\left(\frac{a}{2+a}, \frac{a}{2+a}\right)$ est un minimum globale de f sur $\mathbb{R}^2$.

4 : Pour $a = 1$, on a : $f_1(x_1, x_2) = x_1^2 + x_2^2 + x_1 x_2 - 2x_1 - 2x_2, x_0 = (0, 1)$,

$\nabla f(x) = \begin{pmatrix} 2x + y - 2 \\ 2y + x - 2 \end{pmatrix}, \nabla f(x_0) = \begin{pmatrix} -1 \\ 0 \end{pmatrix} = d_0,\ \alpha_0 = \frac{d_0^T \nabla f(x_0)}{\|d_0\|_A^2} = \frac{1}{2}$,

4.1 : $x_1 = x_0 - \alpha_0 d_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \left(\frac{1}{2}, 1\right), \nabla f(x_1) = \begin{pmatrix} 0 \\ -\frac{1}{2} \end{pmatrix}$.

4.2 : $x_2 = x_1 - \alpha_1 d_1,\ \beta_0 = \frac{d_0^T \nabla f(x_1)}{\|d_0\|_A^2} = \frac{1}{4}$,

$d_1 = \begin{pmatrix} 0 \\ -\frac{1}{2} \end{pmatrix} - \frac{1}{4}\begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{4} \\ \frac{-1}{2} \end{pmatrix}, \alpha_1 = \frac{d_1^T \nabla f(x_1)}{\|d_1\|_A^2} = \frac{2}{3}$

donc $x_2 = \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix} + \frac{2}{3}\begin{pmatrix} \frac{1}{4} \\ \frac{-1}{2} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} \\ \frac{2}{3} \end{pmatrix} = x^*, \nabla f(x^*) = 0.$

Exercice 3.4 *Soit f une fonction définie de $\mathbb{R}^2$ dans $\mathbb{R}$ par :*

$$f(x_1, x_2) = x_1^2 + x_2^2 - 3x_1 - x_1 x_2 + 3,\ \ x_0 = (0, 0).$$

1 : *En appliquant les conditions d'optimalité déterminer les points critiques de f et préciser leur nature.*

2 : *Résoudre le problème* $\min_{(x_1, x_2) \in \mathbb{R}^2} f(x_1, x_2)$, *en appliquant :*

2.1. *L'algorithme du gradient à pas optimal (effectuer trois iterations)*

2.2. *La méthode de relaxation (effectuer trois iterations).*

Solution : 1 : On a :

$$\nabla f(x) = \begin{pmatrix} 2x_1 - x_2 - 3 \\ 2x_2 - x_1 \end{pmatrix} = 0 \Rightarrow \begin{pmatrix} x_1 = 2 \\ x_2 = 1 \end{pmatrix} = x_0$$

donc $x_0 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ le seul point critique.

$1 ** : \nabla^2 f(x) = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$, puisque $\Delta_1 > 0, \Delta_2 > 0$, on a $\nabla^2 f$ est $D.P.$

$\begin{cases} \nabla f(x_0) = 0 \\ \nabla^2 f(x_0) \;\; D.P \end{cases} \Rightarrow \begin{pmatrix} x_1 = 2 \\ x_2 = 1 \end{pmatrix} = x_0$ est minimum local stricte.

$(2.1.a)$: On a : $\nabla f(x_0) = (-3, 0), \alpha_0 = \arg\min_{\alpha>0}(f x_0 - \alpha \nabla f(x_0)) = \frac{1}{2}$,

$x_1 = x_0 - \alpha_0 \nabla f(x_0) = \begin{pmatrix} \frac{3}{2} = 1.5 \\ 0 \end{pmatrix}$.

$(2.1.b)$: $\nabla f(x_1) = \begin{pmatrix} 0 \\ -\frac{3}{2} \end{pmatrix}, \alpha_1 = \arg\min_{\alpha>0}(f(x_1 - \alpha \nabla f(x_1))) = \frac{1}{2}$,

$x_2 = x_1 - \alpha_1 \nabla f(x_1) = \begin{pmatrix} \frac{3}{2} = 1,5 \\ \frac{3}{4} = 0.75 \end{pmatrix}$.

$(2.1.c)$: $\nabla f(x_2) = \begin{pmatrix} \frac{-3}{4} \\ 0 \end{pmatrix}$, $\alpha_2 = \arg\min_{\alpha>0}(f(x_2 - \alpha \nabla f(x_2))) = \frac{1}{2}$.

$x_3 = x_2 - \alpha_2 \nabla f(x_2) = \begin{pmatrix} \frac{15}{8} = 1.875 \\ \frac{3}{4} = 0.75 \end{pmatrix}$.

$(2.2.a)$: $\begin{cases} \left(x_1^{(1)}, 0\right) \leftarrow \arg\min_{\varphi\in\mathbb{R}} f(\varphi, 0), \varphi = \frac{3}{2} = x_1^{(1)}. \\ \left(\frac{3}{2}, x_2^{(1)}\right) \leftarrow \arg\min_{\varphi\in\mathbb{R}} f\left(\frac{3}{2}, \varphi\right), \varphi = \frac{3}{4} = x_1^2, \end{cases}, \left(x_1^{(1)}, x_1^{(2)}\right) = \left(\frac{3}{2}, \frac{3}{4}\right)$,

$(2.2.b)$: $\begin{cases} \left(x_1^{(2)}, \frac{3}{4}\right) \leftarrow \arg\min_{\varphi\in\mathbb{R}} f\left(\varphi, \frac{3}{4}\right), \varphi = \frac{15}{8} = x_2^1, \\ \left(\frac{15}{8}, x_2^{(2)}\right) \leftarrow \arg\min_{\varphi\in\mathbb{R}} f\left(\frac{15}{8}, \varphi\right), \varphi = \frac{15}{16} = x_2^2, \end{cases}, \left(x_1^{(2)}, x_2^{(2)}\right) = \left(\frac{15}{8}, \frac{15}{16}\right)$.

$(2.2.c)$: $\begin{cases} \left(x_1^{(3)}, \frac{15}{16}\right) \leftarrow \arg\min_{\varphi\in\mathbb{R}} f\left(\varphi, \frac{15}{16}\right), \varphi = \frac{63}{32} = x_3^1, \\ \left(\frac{63}{32}, x_2^{(3)}\right) \leftarrow \arg\min_{\varphi\in\mathbb{R}} f\left(\frac{63}{32}, \varphi\right), \varphi = \frac{63}{64} = x_3^2, \end{cases}, \left(x_1^{(3)}, x_2^{(3)}\right) = \left(\frac{63}{32}, \frac{63}{64}\right)$.

Exercice 3.5 *On considére la fonctionla fonction $f : \mathbb{R}^2 \to \mathbb{R}$, définie par :*

$$f(x_1, x_2) = (x-2)^4 + (x-2)^2 y^2 + (x+1)^2.$$

En appliquant la méthode de Newton résoudre le problème suivant :

$$(I) \begin{cases} \min f(x, y), \\ (x, y) \in \mathbb{R}^2, \\ x^0 = (x_0, y_0) = (1, 1), \end{cases}$$

Solution : 1 : On a :

$$\nabla f(x, y) = \begin{pmatrix} 4(x-2)^3 + 2(x-2)y^2 + 2(x+1) \\ \\ 2(x-2)^2 y \end{pmatrix},$$

et
$$\nabla^2 f(x,y) = \begin{pmatrix} 12(x-2)+2y^2+2 & 4(x-2)y \\ 4(x-2)y & 2(x-2)^2 \end{pmatrix}.$$
Alors pour $k=1$, on a :
$$x_1 = x_0 - \left[\nabla^2 f(x_0)\right]^{-1} \nabla f(x_0) = \left(1, \frac{1}{2}\right) \text{ et } f(x_1) = 1,5.$$
Pour $k=2$, on a :
$$x_2 = x_1 - \left[\nabla^2 f(x_1)\right]^{-1} \nabla f(x_1) = (1,39,-0,696) \text{ et } f(x_2) = 16.642,$$
pour $k=3$, on a :
$$x_3 = x_2 - \left[\nabla^2 f(x_2)\right]^{-1} \nabla f(x_2) = (1,746,-0,949) \text{ et } f(x_3) = 15.642.$$