Renaud Sirdey et Héléne Nillès

Optimisation de formes quadratiques

Éditions des Nik's News
www.niksnews.com/editions/

1998

L'œuvre appartient son auteur.
L'auteur est seul responsable du contenu de son œuvre.
L'auteur autorise les Éditions des Nik's News à :

- ajouter à son œuvre des informations les concernant ;
- diffuser gratuitement son œuvre ;
- choisir le ou les formats de diffusion de son œuvre.

Les Éditions des Nik's News s'engagent à ne plus publier une œuvre si son auteur le désire.

Soit une fonction $f : R^n \to R$ telle que :

$$f(x) = \frac{1}{2} < Ax, x > - < b, x > + c \tag{1}$$

avec $c \in R$, $b \in R^n$ et $A \in \mathcal{M}_{nn}(R)$ symétrique, définie positive. On note x^* la solution du problème :

$$f(x^*) \leq f(x), \ \forall x \in R^n \tag{2}$$

1 Etude des méthodes de gradient

1.1 Formes quadratiques

Dans un premier temps, on cherche à mettre la fonction

$$f(x_1, x_2) = \frac{3}{2}x_1^2 + x_2^2 + x_1x_2 - 7x_1 - 4x_2 \tag{3}$$

sous la forme (1). On recherche donc A, b et c.

On pose,

$$\begin{aligned} f(x_1, x_2) &= \frac{1}{2}\begin{pmatrix} x_1 & x_2 \end{pmatrix}\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} b_1 & b_2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + c \\ &= \frac{1}{2}a_{11}x_1^2 + a_{12}x_1x_2 + \frac{1}{2}a_{22}x_2^2 - b_1x_1 - b_2x_2 + c \end{aligned} \tag{4}$$

on en déduit, $a_{11} = 3$, $a_{12} = a_{21} = 1$ (on cherche une matrice symétrique), $a_{22} = 2$, $b_1 = 7$, $b_2 = 4$ et $c = 0$.

On a par conséquent,

$$f(x) = \frac{1}{2}x^T\begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}x - x^T\begin{pmatrix} 7 \\ 4 \end{pmatrix} \tag{5}$$

Pour vérifier que la matrice A est définie positive il suffit de s'assurer que ses valeurs propres sont toutes strictement positives.

Soit λ valeur propre de la matrice A, alors $det(A - \lambda I) = 0$,

$$\begin{aligned} det(A - \lambda I) &= (3 - \lambda)(2 - \lambda) - 1 \\ &= 5 - 5\lambda + \lambda^2 = 0 \end{aligned} \tag{6}$$

On a $\Delta = 5 > 0$ et deux solutions réelles,

$$\lambda_1 = \frac{5+\sqrt{5}}{2} \approx 3.618 \tag{7}$$

$$\lambda_2 = \frac{5-\sqrt{5}}{2} \approx 1.381 \tag{8}$$

Donc A est bien symétrique définie positive.

Calcul de x^*

On cherche x^* tel que

$$g(x^*) = 0, \text{ avec } g(x) = \begin{pmatrix} \frac{\delta f}{\delta x_1}(x) \\ \frac{\delta f}{\delta x_2}(x) \end{pmatrix} = \begin{pmatrix} 3x_1 + x_2 - 7 \\ 2x_2 + x_1 - 4 \end{pmatrix} \tag{9}$$

soit la solution du système

$$\begin{cases} 3x_1 + x_2 - 7 = 0 \\ 2x_2 + x_1 - 4 = 0 \end{cases} \tag{10}$$

Ce système n'a qu'une solution (qui est forcément un minimum car $f(x)$ est une fonction elliptique) telle que $x^* = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ (voir figure 1).

On cherche maintenant à mettre (1) sous la forme

$$f(x) = \frac{1}{2} < A(x - x^*), x - x^* > + c^* \tag{11}$$

Soit

$$\begin{aligned} f(x) &= \frac{1}{2} < A(x - x^*), x - x^* > + c^* \qquad (12) \\ &= \frac{1}{2}[< Ax, x > - < Ax, x^* > - < Ax^*, x, > + < Ax^*, x^* >] + c^* \end{aligned}$$

or $< Ax, x^* >= x^T A^T x^* = x^T A x^* =< x, Ax^* >=< Ax^*, x >$ car A est symétrique. (12) peut donc s'écrire

$$\frac{1}{2} < Ax, x > - < Ax^*, x > + \frac{1}{2} < Ax^*, x^* > \tag{13}$$

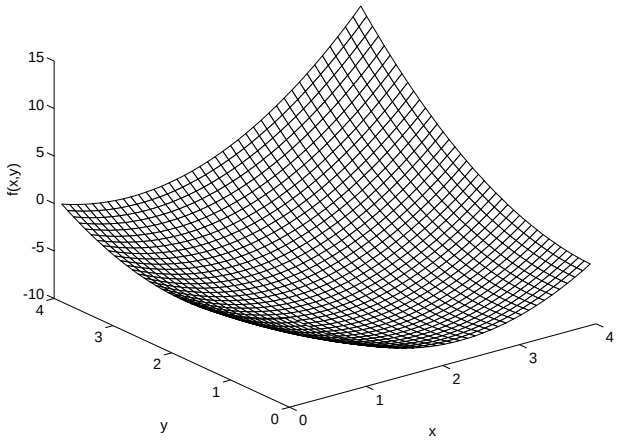
de plus, $Ax^* = b$ car x^* est tel que $Ax^* - b = 0$.

Finalement, on obtient l'expression

$$\frac{1}{2} < Ax, x > - < b, x > + \frac{1}{2} < Ax^*, x^* > \tag{14}$$

qui revient à (1) en choisissant $c^* = -\frac{1}{2} < Ax^*, x^* >$.

FIG. 1 – Représentation de la fonction $f(x_1, x_2)$.



1.2 Gradient à pas fixe

Condition sur ρ

Dans un premier temps, on considère le problème d'optimisation suivant

$$\min_{R} \frac{1}{2} ax^2 - bx + c, \text{ avec } a \in R^{*+} \text{ et } b, c \in R \tag{15}$$

On cherche alors à exprimer $x^{(k+1)}$ en fonction du résultat des itérations précédentes, soit :

$$\begin{aligned}
x^{(k+1)} &= x^{(k)} - \rho g(x^{(k)}) \qquad (16)\\
&= x^{(k)} - \rho a x^{(k)} + \rho b \\
&= x^{(k-1)} - \rho a x^{(k-1)} + \rho b - \rho a(x^{(k-1)} - \rho a x^{(k-1)} + \rho b) + \rho b \\
&= x^{(k-1)}(1 - \rho a)^2 + \rho b(2 - \rho a) \\
&= x^{(k-2)}(1 - \rho a)^2 + \rho a x^{(k-2)}(1 - \rho a)^2 + \rho b(1 - \rho a)^2 + \rho b(2 - \rho a) \\
&= x^{(k-2)}(1 - \rho a)^3 + \rho b(1 + (1 - \rho a) + (1 - \rho a)^2)
\end{aligned}$$

On en déduit la formule générale supposée vraie à l'étape n avec $n \leq k - 1$

$$x^{(k+1)} = x^{(k-n)}(1 - \rho a)^{(n+1)} + \rho b \sum_{i=0}^{n} (1 - \rho a)^i \tag{17}$$

et on montre que ceci est valable à l'étape $n+1$

$$\begin{aligned} x^{(k+1)} &= (x^{(k-(n+1))} - \rho g(x^{(k-(n+1))}))(1-\rho a)^{(n+1)} + \rho b \sum_{i=0}^{n} (1-\rho a)^i \quad (18) \\ &= (x^{(k-(n+1))} - \rho a x^{(k-(n+1))} + \rho b)(1-\rho a)^{(n+1)} + \rho b \sum_{i=0}^{n} (1-\rho a)^i \\ &= x^{(k-(n+1))}(1-\rho a)^{(n+2)} + \rho b \sum_{i=0}^{n+1} (1-\rho a)^i \end{aligned}$$

ce qui montre par réccurence la validité de l'expression (17).

D'où l'expression de $x^{(k)}$ en fonction de $x^{(0)}$

$$\begin{aligned} x^{(k)} &= x^{(0)}(1-\rho a)^k + \rho b \underbrace{\sum_{i=0}^{k-1} (1-\rho a)^i}_{\text{suite géométrique}} \quad (19) \\ &= x^{(0)}(1-\rho a)^k + \rho b \frac{1-(1-\rho a)^k}{\rho a} \end{aligned}$$

On s'intéresse à

$$\lim_{k \to +\infty} x^{(k)} \quad (20)$$

Si $\rho a \in]0,1]$, on a

$$\lim_{k \to +\infty} (1-\rho a)^k = 0 \quad (21)$$

Par conséquent

$$\lim_{k \to +\infty} x^{(k)} = \frac{b}{a} \quad (22)$$

Or $\frac{b}{a}$ est justement la solution de l'équation :

$$g(x) = ax - b = 0 \quad (23)$$

En se replacant dans le cadre de l'expression matricielle (1), on a :

$$\lim_{k \to +\infty} x^{(k)} = A^{-1} b \quad (24)$$

si les valeurs propres de la matrice ρA appartiennent toutes à l'intervalle $]0,1]$.

On a

$$\begin{aligned} f(x) &= \frac{1}{2} < P^T A x, P^T x > - < P^T b, P^T x > + c \quad (25) \\ &= \frac{1}{2} < P^T A P P^T x, P^T x > - < P^T b, P^T x > + c \\ &= \frac{1}{2} < D y, y > - < b^*, y > + c \end{aligned}$$

avec

- P : matrice orthogonale des vecteurs propres de A,
- D : matrice diagonale des valeurs propres de A,
- $y = P^T x$,
- $b^* = P^T b$.

Ce qui revient à la minimisation de n (dimension de y) fonctions indépendantes du type

$$\frac{1}{2}\lambda_i y_i^2 - b_i^* y_i \tag{26}$$

où λ_i est la $i^{ème}$ valeur propre de A et est strictement positive car A est définie positive. On se retrouve alors dans le cas du problème (15) et il est nécessaire que $\rho\lambda_i$ ($i^{ème}$ valeur propre de la matrice ρA) appartienne à l'intervalle $]0, 1]$.

Calcul des valeurs propres et des vecteurs propres de la matrice A

Les valeurs propres ont été calculées précédement, on a

$$\lambda_1 = \frac{5+\sqrt{5}}{2} \text{ et } \lambda_2 = \frac{5-\sqrt{5}}{2} \tag{27}$$

On cherche donc deux vecteurs propres orthogonaux (puisque A est symétrique) tels que

$$(A - \lambda_i I)Y_i = 0 \tag{28}$$

Soit, pour le vecteur Y_1, la solution du système

$$\begin{cases} (3-\lambda_1)Y_{11} + Y_{12} = 0 \\ Y_{11} + (2-\lambda_1)Y_{12} = 0 \end{cases} \tag{29}$$

on a alors $Y_{21} = (\lambda_1 - 3)Y_{11}$ et

$$Y_{11}\underbrace{(1-(2-\lambda_1)(3-\lambda_1))}_{=0} = 0 \tag{30}$$

on choisit donc Y_{11} arbitrairement dans R^* (car $Y \neq \vec{0}$), par exemple $Y_{11} = 1$. D'où

$$Y_1 = \begin{pmatrix} 1 \\ \frac{-1+\sqrt{5}}{2} \end{pmatrix} \tag{31}$$

on prend Y_2 tel que $< Y_1, Y_2 >= 0$ et $Y_2 \neq \vec{0}$, soit

$$Y_2 = \begin{pmatrix} \frac{1-\sqrt{5}}{2} \\ 1 \end{pmatrix} \tag{32}$$

Courbe d'isovaleurs

On s'intéresse maintenant à la courbe d'isovaleurs de la fonction (3) telle que

$$f(x_1, x_2) = 5 \tag{33}$$

le but est de montrer que cette courbe est une ellipse de centre x^* et d'axes les vecteurs propres de A.

Soit P la matrice de passage de la base canonique de R^n à la base de vecteurs propres de la matrice A,

$$P = \begin{pmatrix} Y_1 & Y_2 \end{pmatrix} \tag{34}$$

et soit :

- x : vecteur dans la base canonique de R^n centrée en O ;
- $\bar{x}$: vecteur équivalent à x dans la base canonique de R^n centrée en x^* ;
- $\hat{x}$: vecteur équivalent à x dans la base des vecteurs propres de A centrée en x^*.

On a

$$\begin{aligned} \bar{x} &= x - x^* \\ \hat{x} &= P^{-1}\bar{x} = P^{-1}(x - x^*) \end{aligned} \tag{35}$$

et

$$\begin{aligned} \bar{x} &= P\hat{x} \\ x &= \bar{x} + x^* = P\hat{x} + x^* \end{aligned} \tag{36}$$

En partant de la forme (11), et en exprimant $f(x)$ à l'aide du vecteur $\hat{x}$ on obtient

$$\begin{aligned} f(x) &= \frac{1}{2}(P\hat{x} + x^* - x^*)^T A(P\hat{x} + x^* - x^*) + c^* \\ &= \frac{1}{2}\hat{x}^T P^T A P \hat{x} + c^* \end{aligned} \tag{37}$$

or $P^{-1} = P^T$ (car P est une matrice orthogonale) et $P^{-1}AP = P^TAP = D$ (matrice diagonale des valeurs propres), on a donc

$$\begin{aligned} f(x) &= \frac{1}{2}\hat{x}^T D \hat{x} + c^* \\ &= \frac{1}{2}\begin{pmatrix} \hat{x}_1 & \hat{x}_2 \end{pmatrix}\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}\begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} + c^* \\ &= \frac{1}{2}\lambda_1\hat{x}_1^2 + \frac{1}{2}\lambda_2\hat{x}_2^2 + c^* \end{aligned} \tag{38}$$

On s'intéresse à la courbe (33), on obtient

$$\begin{aligned} \frac{1}{2}\lambda_1\hat{x}_1^2 + \frac{1}{2}\lambda_2\hat{x}_2^2 + c^* &= 5 \\ \frac{1}{2}\lambda_1\hat{x}_1^2 + \frac{1}{2}\lambda_2\hat{x}_2^2 &= 5 - c^* \end{aligned} \tag{39}$$

et on arrive à

$$\frac{\lambda_1}{2(5-c^*)}\hat{x}_1^2 + \frac{\lambda_2}{2(5-c^*)}\hat{x}_2^2 = 1 \tag{40}$$

ce qui est bien l'équation d'une ellipse (voir figure 2) de demi grand axe

$$a = \sqrt{\frac{4(5-c^*)}{5+\sqrt{5}}} \approx 2.78 \tag{41}$$

et de demi petit axe

$$b = \sqrt{\frac{4(5-c^*)}{5-\sqrt{5}}} \approx 4.50 \tag{42}$$

Plus généralement, la courbe d'isovaleurs telle que $f(x_1, x2) = \beta$ est une ellipse de demi grand axe

$$a = \sqrt{\frac{4(\beta-c^*)}{5+\sqrt{5}}} \tag{43}$$

et de demi petit axe

$$b = \sqrt{\frac{4(\beta-c^*)}{5-\sqrt{5}}} \tag{44}$$

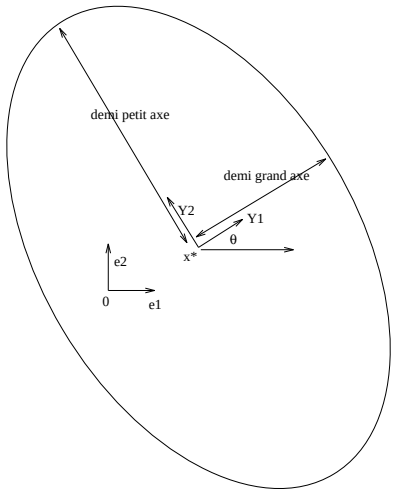
Simulation

L'application de la méthode du gradient à pas fixe à la fonction (3) donne

$$\begin{aligned} x^{(0)} &= \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \\ x^{(1)} &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} - 0.2 \begin{pmatrix} 1-7=6 \\ 2-4=-2 \end{pmatrix} = \begin{pmatrix} 1.2 \\ 1.4 \end{pmatrix} \\ x^{(2)} &= \begin{pmatrix} 1.6 \\ 1.4 \end{pmatrix}, \end{aligned} \tag{45}$$

Courbes de niveaux et directions de descente, voir tableau 1 et figure 3.

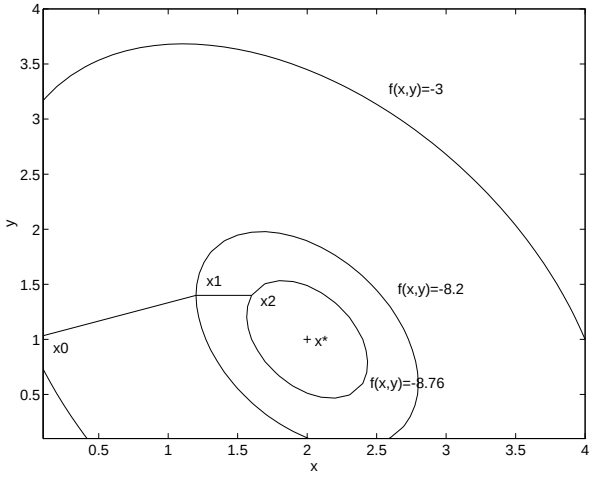
FIG. **2** – Courbes des isovaleurs telle que $f(x_1, x_2) = 5$, on a $\theta \approx 31$ deg, $a \approx 2.78$ (demi grand axe) et $b \approx 4.50$ (demi petit axe). Les vecteurs propres sont normalisés.



i	$f(x^{(i)})$	a	b
0	-3	1.82	2.94
1	-8.2	0.66	1.08
2	-8.76	0.36	0.59

TAB. 1 – Propriétés des courbes de niveaux obtenues en simulant la méthode du gradient à pas fixe (3 itérations).

FIG. 3 – Courbes de niveaux et directions de descente obtenues en simulant la méthode du gradient à pas fixe (3 itérations, $\rho = 0.2$).



1.3 Méthode du gradient à pas optimal

On note g le gradient de la fonction (1), soit

$$g(x) = \nabla f(x) \text{ pour } x \in R^n \tag{46}$$

Dans un premier temps, on cherche à montrer que

$$g(x - \rho g(x)) = g(x) - \rho A g(x) \tag{47}$$

On a $g(x) = Ax - b$, donc

$$\begin{aligned} g(x - \rho g(x)) &= A(x - \rho(Ax - b)) - b \\ &= A(x - \rho Ax + \rho b) - b \\ &= Ax - \rho A^2 x + \rho Ab - b \\ &= Ax - b - \rho A^2 x + \rho Ab \\ &= Ax - b - \rho A(Ax - b) \\ &= g(x) - \rho A g(x) \end{aligned} \tag{48}$$

on en déduit la valeur du pas optimal qui doit vérifier

$$\begin{aligned} < g(x^{(k)} - \rho g(x^{(k)})), g(x^{(k)} > &= 0 \\ < g(x^{(k)}) - \rho A g(x^{(k)}), g(x^{(k)}) > &= 0 \\ < g(x^{(k)}), g(x^{(k)}) > -\rho < A g(x^{(k)}), g(x^{(k)}) > &= 0 \\ \|g(x^{(k)})\|^2 - \rho \|g(x^{(k)})\|_A^2 &= 0 \end{aligned} \tag{49}$$

i	$f(x^{(i)})$	a	b
0	-3	1.82	2.94
1	-8.71	0.39	0.64
2	-8.98	0.08	0.14

TAB. 2 – Propriétés des courbes de niveaux obtenues en simulant la méthode du gradient à pas optimal (3 itérations).

on rappelle qu'il est possible de définir une norme par rapport à A car cette matrice est définie positive.

D'où

$$\rho = \frac{\|g(x^{(k)})\|^2}{\|g(x^{(k)})\|_A^2} \tag{50}$$

Simulation

L'application de la méthode du gradient à pas optimal à la fonction (3) donne

$$\begin{aligned} x^{(0)} &= \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \\ x^{(1)} &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} - 0.2857 \begin{pmatrix} -6 \\ -2 \end{pmatrix} = \begin{pmatrix} 1.71 \\ 1.57 \end{pmatrix} \\ x^{(2)} &= \begin{pmatrix} 1.90 \\ 1.00 \end{pmatrix} \end{aligned} \tag{51}$$

Courbes de niveaux et directions de descente, voir tableau 2 et figure 4.

2 Mise en oeuvre

2.1 Test d'arrêt

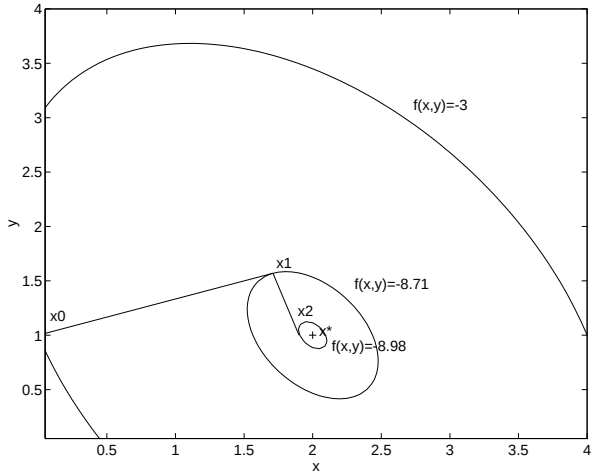
On propose le test d'arrêt suivant :

$$\|x^{(k+1)} - x^{(k)}\| \leq \epsilon \tag{52}$$

2.2 Macro du gradient à pas fixe

```
function [xk,k]=mygrad(A,b,x0,rho,n,epsilon)
xk=x0;
```

FIG. 4 – Courbes de niveaux et directions de descente obtenues en simulant la méthode du gradient à pas optimal (3 itérations).




```
xk_old=x0;
for k=1:n
    xk=xk-rho*(A*xk-b);
    if norm(xk-xk_old)<epsilon
        return;
    end;
    xk_old=xk;
end;
```

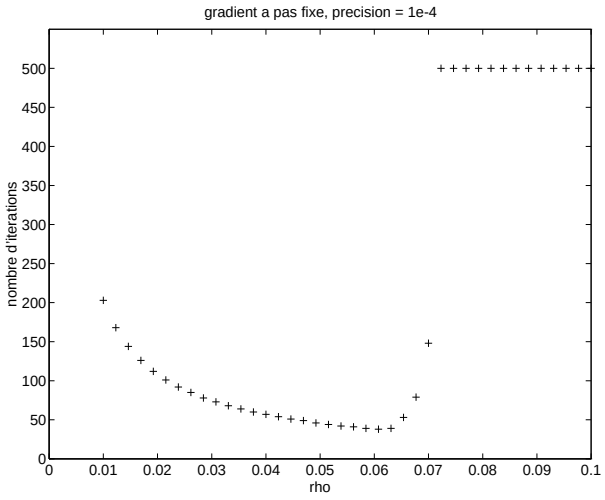
2.3 Macro du gradient à pas optimal

```
function [xk,k]=mygrad_o(A,b,x0,n,epsilon)
xk=x0;
xk_old=x0;
for k=1:n
    gxk=A*xk-b;
    rho = (gxk´*gxk)/(gxk´*A*gxk);
    xk=xk-rho*(A*xk-b);
    if norm(xk-xk_old)<epsilon
        return;
    end;
```

```
    xk_old=xk;
end;
```

2.4 Test du gradient à pas fixe pour plusieurs valeurs du pas

FIG. 5 – Evolution du nombre d'itérations en fonction de la valeur du pas. Lorsque le nombre d'itérations est égale à 500, la méthode a divergée.



Voir figure 5. On remarque qu'il existe un ρ optimal minimisant le nombre d'itérations et approximativement égale à 0.63.

2.5 Nombre d'itérations nécessaires vis à vis de la précision demandée

Voir figure 6.

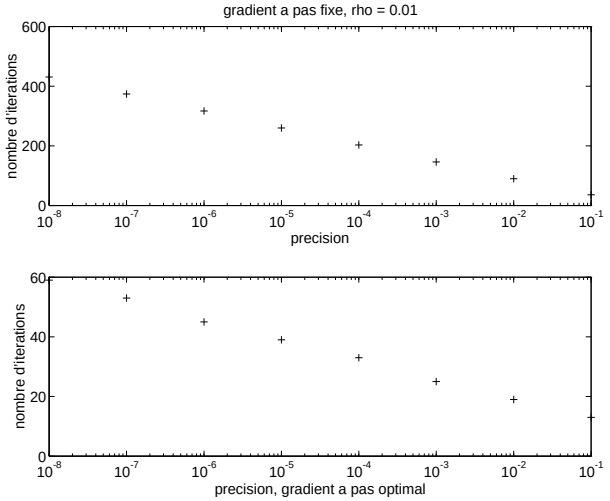



FIG. **6** – Nombre d'itérations nécessaires pour atteindre la précision demandée. L'axe des abscisses est logarithmique. On remarque que les performances du gradient optimal sont bien meilleures que celles du gradient à pas fixe.