Méthodes numériques : optimisation.
Préparation au partiel de mars 2015

Amic Frouvelle

4 mars 2015

1 Cas pathologique de la méthode de Newton

On considère la fonction $x \mapsto e^{x^4}$ de $\mathbb{R}$ dans $\mathbb{R}$. On cherche à minimiser f en approchant un zéro de f' par la méthode de Newton.

1. Montrer que les itérées de la méthode de Newton satisfont
$$x_{k+1} = x_k\Big(1 - \frac{1}{3 + 4x_k^4}\Big).$$
2. Montrer que quelque soit le choix de $x_0 \neq 0$, on obtient que la suite x_k converge linéairement vers 0 (l'unique minimum de la fonction). Quel est le taux de convergence ?
3. Pourquoi dit-on qu'il s'agit d'un cas pathologique de la méthode de Newton ? Quelle est la particularité de la fonction qui fait qu'on est dans ce cas ?
4. Expliquer l'allure des courbes ci-dessous, correspondant à diverses simulations de la suite x_k suivant des valeurs de x_0 différentes.

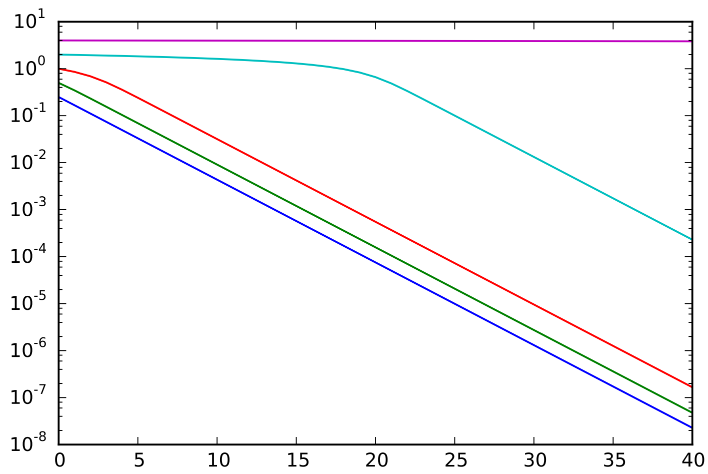

FIGURE 1 – Courbes représentant 40 itérations de la méthode pour $x_0 \in \{\frac{1}{4}, \frac{1}{2}, 1, 2, 4\}$.

2 Précision d'une méthode de recherche de minimum

On se donne une fonction $f : \mathbb{R} \to \mathbb{R}$ dont on connaît une approximation $\hat{f}$ satisfaisant

$$|f(x) - \hat{f}(x)| \leqslant \eta \|f\|_\infty.$$

On note $\hat{f}'_d(x)$ (resp. $\hat{f}'_c(x)$) les approximations de f' par différences finies à droite (resp. centrées) calculées à partir de $\hat{f}$ avec un paramètre de discrétisation ε. On rappelle que sous certaines hypothèses raisonnables, si on prend ε de l'ordre de grandeur de $\sqrt{\eta}$ (resp. $\eta^{\frac{1}{3}}$), alors l'erreur sur l'approximation est de l'ordre de grandeur de $\sqrt{\eta}$ (resp. $\eta^{\frac{1}{3}}$) :

$$|f'(x) - \hat{f}'_d(x)| \leqslant C_d \sqrt{\eta}\, \|f\|_\infty \quad \text{pour } \varepsilon = \sqrt{\eta},$$
$$|f'(x) - \hat{f}'_c(x)| \leqslant C_c\, \eta^{\frac{2}{3}}\, \|f\|_\infty \quad \text{pour } \varepsilon = \eta^{\frac{1}{3}},$$

avec C_d et C_c des constantes de l'ordre de un.

1. Rappeler pourquoi on ne peut pas prendre ε aussi petit qu'on veut pour augmenter la précision, et expliquer brièvement pourquoi et en quoi l'une des deux méthodes est meilleure.

On suppose que la fonction f est unimodale sur $[a, b]$, avec un minimum en $x_* \in]a, b[$, de classe C^2, et satisfaisant $f''(x) \geqslant h > 0$ sur $[a, b]$, avec h, $b - a$ et $\|f\|_\infty$ des constantes de l'ordre de un.

2. On effectue une approximation de x_* par méthode de dichotomie sur la dérivée, à l'aide de l'approximation $\hat{f}'_d$. Montrer qu'il existe une constante C_0, du même ordre de grandeur que un, telle que si $|\hat{f}'_d(x)| \geqslant C_0\sqrt{\eta}$, alors on peut savoir a priori que le signe de $f'(x)$ est le même que celui de $\hat{f}'_d(x)$. En conclure que l'on peut obtenir une approximation de $\hat{x}_*$ de x_* à une précision de l'ordre de grandeur de $\sqrt{\eta}$:
$$|\hat{x}_* - x_*| \leqslant C_1\sqrt{\eta},$$
pour une constante C_1 de l'ordre de grandeur de un.
3. Lorsque l'on utilise l'approximation $\hat{f}'_d$, comment adapter les réponses de la question précédente ?
4. On s'intéresse maintenant à approximer la valeur de $f(x_*)$ seulement, en appliquant la fonction $\hat{f}$. Que peut-on dire sur l'efficacité de l'une ou l'autre des méthodes ?
5. Que se passe-t-il si on ne suppose plus que $f''(x) \geqslant h > 0$ sur $[a, b]$? On peut prendre l'exemple du cas où on essaye de minimiser $x \mapsto (x - c)^4 + d$ sur $[c - 1, c + 1]$ par les mêmes méthodes.

3 Étude de cas en dimension 2

On se propose d'étudier un cas pour lequel les hypothèses du premier théorème de convergence de la méthode de descente de gradient à pas fixe s'appliquent, mais où la suite (X_k), même en étant sur un compact, ne converge pas vers un point critique et a une infinité de valeurs d'adhérence.

Pour cela on va étudier le cas de la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ suivante (définie en coordonnées polaires, on écrira toujours $X = \binom{r\cos\theta}{r\sin\theta}$, avec $r \geqslant 0$ et $\theta \in \mathbb{R}$) :

$$f(X) = F(r, \theta) = \begin{cases} (r-1)^2\Big(2 - \cos(4\ln(r-1) + \theta)\Big) & \text{si } r > 1 \\ 0 & \text{si } r \leqslant 1. \end{cases} \tag{1}$$

3.1 Étude de la fonction

1. Montrer que la fonction f est bien définie par la formule (1), au sens où le terme de droite ne dépend pas du choix de θ qui est fait pour écrire $X = \binom{r\cos\theta}{r\sin\theta}$.

2. On pose $\theta_0 \in [0, \frac{\pi}{2}]$ l'angle tel que $\cos\theta_0 = \frac{2}{\sqrt{5}}$ et $\sin\theta_0 = \frac{1}{\sqrt{5}}$. Montrer que f est de classe C^1 sur $\mathbb{R}^2$ et que le gradient est bien défini par la formule suivante :

$$\nabla f(X) = \begin{cases} g(r,\theta)\,\mathbf{u}(\theta) + h(r,\theta)\,\mathbf{v}(\theta) & \text{si } r > 1 \\ \binom{0}{0} & \text{si } r \leqslant 1, \end{cases} \tag{2}$$

où

$$\mathbf{u}(\theta) = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \quad \mathbf{v}(\theta) = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix},$$

$$g(r,\theta) = \partial_r F(r,\theta) = 2(r-1)\Big(2 + \sqrt{5}\sin(4\ln(r-1) + \theta - \theta_0)\Big)$$

$$h(r,\theta) = \frac{1}{r}\partial_\theta F(r,\theta) = \frac{(r-1)^2}{r}\sin(4\ln(r-1) + \theta).$$

3. Déduire de cette formule que les seuls points critiques de f sont les points $X \in \mathbb{R}^2$ tels que $r \leqslant 1$, et que ce sont des points de minimum global.
4. Déduire également que l'on a, pour tout $X \in \mathbb{R}^2$ tel que $r > 1$, $\|\nabla f(X)\| \leqslant 10(r-1)$.

3.2 Méthode de descente de gradient à pas fixe

On admet (facultatif : le montrer) que ∇f est L-lipschitz sur $\mathbb{R}^2$ pour une constante $L \geqslant 10$. On prend $0 < \alpha < \frac{1}{L} (\leqslant \frac{1}{10})$, on définit la suite X_k par

$$X_0 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad X_{k+1} = X_k - \alpha\nabla f(X_k),$$

et on note $r_k = \|X_k\|$.

1. Montrer que $S_0 = \{X \in \mathbb{R}^2, f(X) \leqslant f(X_0)\}$ est inclus dans la boule fermée de centre 0 et de rayon 2. En déduire que la suite $f(X_k)$ est décroissante, de limite 0 et que $\|\nabla f(X_k)\|$ tend vers 0 lorsque $k \to \infty$.
2. Montrer que l'on a pour tout $k \in \mathbb{N}$,

$$r_k > 1 \quad \text{et} \quad (1 - 10\alpha)(r_k - 1) \leqslant r_{k+1} - 1 \leqslant (1 + 10\alpha)(r_k - 1),$$

et en déduire, avec la question précédente, que $r_k \to 1$ lorsque $k \to \infty$.

3. On définit θ_k de sorte que $X_k = \binom{r_k\cos\theta_k}{r_k\sin\theta_k} = r_k\mathbf{u}(\theta_k)$. En calculant $X_{k+1} \cdot X_k$, montrer que

$$\cos(\theta_{k+1} - \theta_k) = \frac{r_k - \alpha g(r_k,\theta_k)}{\sqrt{(r_k - \alpha g(r_k,\theta_k))^2 + \alpha^2 h(r_k,\theta_k)^2}}.$$

En calculant $\sin^2(\theta_{k+1} - \theta_k)$ et en utilisant l'inégalité $|t| \leqslant \frac{\pi}{2}|\sin t|$ pour $t \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, en déduire qu'à partir d'un certain rang, quitte à redéfinir θ_{k+1} en lui ajoutant un nombre entier de fois 2π, on peut obtenir

$$|\theta_{k+1} - \theta_k| \leqslant 2\alpha(r_k - 1)^2.$$

4. On note $t_k = 4\ln(r_k - 1) + \theta_k$. Montrer qu'il existe un intervalle $[a, b]$ tel qu'à partir d'un certain rang, si $t_k \in [a, b]$, alors $t_{k+1} \geqslant t_k$.
5. Montrer enfin que si l'on choisit α suffisamment petit, alors à partir d'un certain rang on obtient $|t_{k+1} - t_k| < b - a$. En déduire que pour ce choix de α, il existe un $n \in \mathbb{N}$ tel que pour tout $k \in \mathbb{N}$ on ait $t_k \geqslant a - 2n\pi$.
6. En conclure que pour ce choix de α, la suite θ_k tend vers $+\infty$, et que la suite X_k admet pour valeurs d'adhérences le cercle unité $C = \{X \in \mathbb{R}^2, \|X\| = 1\}$ en entier.