Méthodes numériques : optimisation.
Partiel du 9 mars 2015

Recommandations

Le partiel est probablement long, mais le barème en tiendra compte. Il n'est clairement pas nécessaire de tout avoir résolu pour avoir la note maximale. Les trois exercices sont indépendants.

Les questions de chaque exercice ne sont pas censées être bloquantes pour les suivantes. N'hésitez donc pas à admettre des résultats et passer à la suite.

Les dernières questions de chaque exercice, notées *, sont un peu plus difficiles ou demandent un peu plus d'initiative.

1 Cas pathologique de la méthode de la sécante

On considère la fonction $f : x \mapsto x^4$ de $\mathbb{R}$ dans $\mathbb{R}$. On cherche à minimiser f en approchant un zéro de f' par la méthode de la sécante. On se fixe des réels x_0 et x_1 avec $x_0 \neq x_1$.

1. Montrer que les itérées de la méthode de la sécante satisfont, pour $k \geqslant 1$ et lorsque l'on a $x_k \neq x_{k-1}$:
$$x_{k+1} = x_k\Big(1 - \frac{x_k^2}{x_k^2 + x_k x_{k-1} + x_{k-1}^2}\Big).$$
2. Si $x_k \neq 0$, on définit $y_k = \frac{x_{k+1}}{x_k}$. Montrer que pour $k \geqslant 1$, si $x_k \neq 0$, on peut définir y_k et y_{k-1} et on a $y_k = h(y_{k-1})$ pour la fonction $h : x \mapsto \frac{x+1}{x^2+x+1}$.
3. Étudier les variations de la fonction h sur $\mathbb{R}$ et en déduire que si $x_0 \neq 0$, $x_1 \neq 0$ et $x_0 \neq -x_1$, alors :
   - pour tout $k \geqslant 1$, y_k est bien défini et $y_k \in [-\frac{1}{3}, 1] \setminus \{0\}$.
   - pour tout $k \geqslant 2$, $y_k \in [\frac{2}{3}, 1[$.
   - pour tout $k \geqslant 3$, $y_k \in]\frac{2}{3}, \frac{15}{19}]$.
   - La suite $(x_k)_{k\in\mathbb{N}}$ converge linéairement vers 0, mais ne converge pas superlinéairement.
4. Pourquoi dit-on qu'il s'agit d'un cas pathologique de la méthode de la sécante ? Quelle est la particularité de la fonction qui fait qu'on est dans ce cas ?
5. * Quel est le taux de convergence linéaire de la suite $(x_k)_{k\in\mathbb{N}}$ lorsque $x_0 \neq 0$, $x_1 \neq 0$ et $x_0 \neq -x_1$? On pourra montrer que la fonction h est contractante sur $[\frac{2}{3}, 1]$.

2 Convergence de la méthode de descente de gradient à pas fixe en dimension un

1. Soit $(x_k)_{k\in\mathbb{N}}$ une suite de réels. Montrer que si c n'est pas une valeur d'adhérence de la suite, alors il existe $\delta > 0$ et un rang k_0 tels que pour tout $k \geqslant k_0$, $x_k \notin [c-\delta, c+\delta]$.
2. On suppose que $|x_{k+1} - x_k| \to 0$ lorsque $k \to \infty$ et que la suite $(x_k)_{k\in\mathbb{N}}$ admet deux valeurs d'adhérence $a \in \mathbb{R}$ et $b \in \mathbb{R}$ avec $a < b$. Montrer que tout réel $c \in [a, b]$ est valeur d'adhérence de la suite $(x_k)_{k\in\mathbb{N}}$.

Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction C^1 bornée inférieurement. On se donne $x_0 \in \mathbb{R}$ et on suppose que f' est L-lipschitzienne sur $S_0 = \{x \in \mathbb{R} | f(x) \leqslant f(x_0)\}$. On définit la suite $(x_k)_{k\in\mathbb{N}}$ correspondant à la méthode de descente de gradient à pas fixe, avec un pas $\alpha \in]0, \frac{2}{L}[$:

$$x_{k+1} = x_k - \alpha f'(x_k).$$

3. Montrer que $|x_{k+1} - x_k| \to 0$ lorsque $k \to \infty$.
4. Montrer que si la suite $(x_k)_{k\in\mathbb{N}}$ admet deux valeurs d'adhérence $a \in \mathbb{R}$ et $b \in \mathbb{R}$ avec $a < b$, alors f est constante sur $[a, b]$, et obtenir une contradiction.
5. En déduire que lorsque $k \to \infty$, on a soit $x_k \to +\infty$, soit $x_k \to -\infty$, soit $x_k \to x_* \in \mathbb{R}$, avec $f'(x_*) = 0$.
6. Le cas $x_k \to +\infty$ (ou $-\infty$) est-il possible ?
7. * Expliquer brièvement pourquoi le raisonnement des questions précédentes ne s'applique pas en dimension supérieure ou égale à deux.

3 Précision de la méthode de la section dorée

On se donne une fonction $f : [a, b] \to \mathbb{R}$ dont on connaît une approximation $\hat{f}$ avec une précision relative $\eta \ll 1$, c'est-à-dire satisfaisant pour tout $x \in [a, b]$,

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

où $\|f\|_\infty = \sup_{t \in [a,b]} |f(t)|$.

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 $0 < h \leqslant f''(x) \leqslant H$ sur $[a, b]$, avec h, H, et $\|f\|_\infty$ des constantes de l'ordre de grandeur de 1.

1. Montrer, en utilisant une estimation de second ordre pour la fonction f, que si x et y sont des réels de $[a, b]$ avec $x_* \leqslant x \leqslant y$, on a

   $$\hat{f}(y) - \hat{f}(x) \geqslant \frac{h}{2}|y - x|^2 - 2\eta \|f\|_\infty.$$

   On peut montrer exactement de la même manière (et on l'admettra) que si $x \leqslant y \leqslant x_*$, alors

   $$\hat{f}(x) - \hat{f}(y) \geqslant \frac{h}{2}|y - x|^2 - 2\eta \|f\|_\infty.$$

2. En déduire qu'il existe une constante C_0 de l'ordre de grandeur de 1 telle que si x et y sont des réels de $[a, b]$ avec $x < y$ et $|y - x| > C_0\sqrt{\eta}$, alors
   — soit $x_* \in [x, y]$,
   — sinon, si $y < x_*$, alors $\hat{f}(x) > \hat{f}(y)$, et si $x_* < x$, alors $\hat{f}(x) < \hat{f}(y)$.

On effectue une approximation de x_* par la méthode de la section dorée sur la fonction $\hat{f}$. On rappelle que les points d'évaluation de la méthode sont donnés des quadruplets (a_n, c_n, d_n, b_n), avec $a_0 = a$, $b_0 = b$ et

$$c_n = a_n + (1 - \alpha)(b_n - a_n), \quad d_n = a_n + \alpha(b_n - a_n),$$

où $\alpha = \frac{1}{2}(\sqrt{5} - 1) \approx 0,618$. On rappelle également que $\alpha^2 = 1 - \alpha$.

3. Montrer que $d_n - c_n = \alpha^3(b_n - a_n)$. Déduire de la question précédente que tant que l'on a $b_n - a_n > \frac{C_0}{\alpha^3}\sqrt{\eta}$, alors on a toujours $x_* \in [a_{n+1}, b_{n+1}]$, puis que les suites a_n, b_n, c_n et d_n convergent vers un réel $\hat{x}_*$ tel que

   $$|\hat{x}_* - x_*| \leqslant C_1\sqrt{\eta},$$

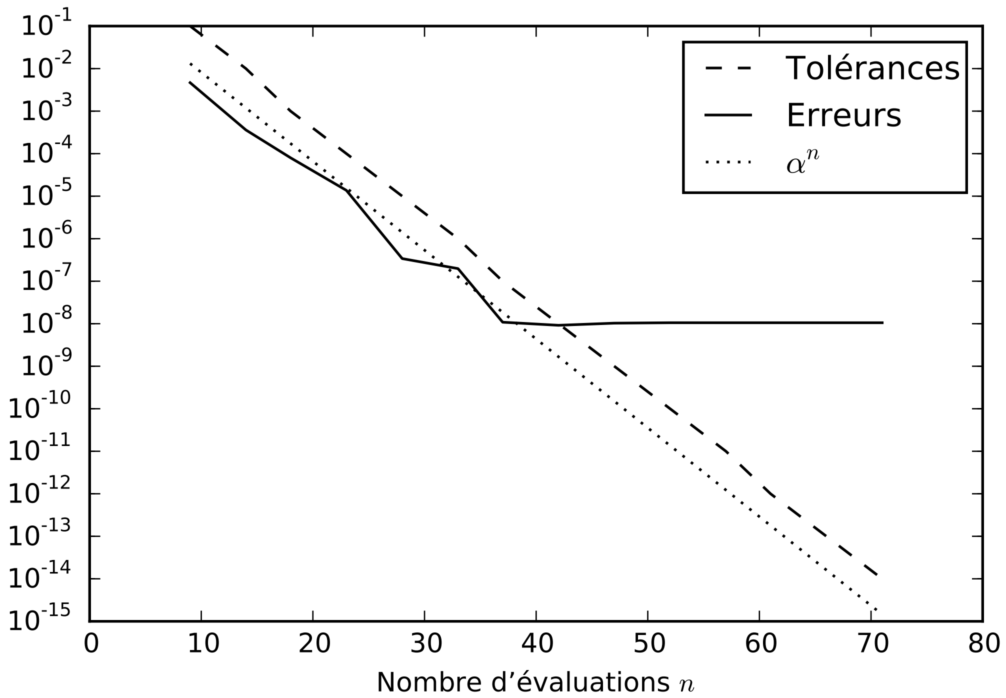
   avec $C_1 = \frac{C_0}{\alpha^3}$.

4. La constante C_1 étant encore une constante de l'ordre de grandeur de un, on dit que la précision de l'approximation de x_* est de l'ordre de grandeur de $\sqrt{\eta}$. Quel est le critère d'arrêt que l'on peut donner à l'algorithme pour obtenir cette précision ? Donner une estimation du coût total de la méthode dans ce cas, en terme du nombre d'évaluations de la fonction $\hat{f}$.

5. Si on s'intéresse plutôt à approximer la valeur de $f(x_*)$ seulement, en appliquant la fonction $\hat{f}$ en $\hat{x}_*$, quel est l'ordre de grandeur de la précision obtenue sur $|\hat{f}(\hat{x}_*) - f(x_*)|$? On pourra utiliser un développement de Taylor à l'ordre 2 pour f au point x_*.

(suite de l'exercice au verso)

On s'intéresse maintenant à un exemple pratique pour comprendre ce qui se passe sur une machine. On considère la fonction sin sur $[-3,-1]$, qui satisfait bien les hypothèses précédentes et admet un minimum en $-\frac{\pi}{2}$. Voilà ce que l'on obtient lorsque l'on trace par exemple les points d'ordonnée $|-\frac{\pi}{2}-\hat{x}_*|$, et d'abscisses n, pour différentes tolérances données (de 10^{-1} à 10^{-15}), où à chaque fois $\hat{x}_*$ est le point final donné par l'algorithme, et n le nombre d'évaluations de la fonction sin lors de l'exécution de l'algorithme.



On suppose que la fonction `sin` de la machine donne autant de précision que possible sur les nombres à virgule flottante. On va se baser sur le fait que tous les réels de $[-1-2^{-53},-1+2^{-54}]$ sont codés par le même nombre à virgule flottante en machine : -1.

6. Montrer que si $x \in [-\frac{\pi}{2}-\delta, -\frac{\pi}{2}+\delta]$, alors $\sin x \in [-1,-1+\delta^2]$ (on pourra utiliser l'inégalité $\cos t \geqslant 1-\frac{t^2}{2}$, valable pour tout $t \in \mathbb{R}$). En déduire que si $\delta = 2^{-27} \approx 7\cdot 10^{-9}$, et que si `x` est une variable dont la valeur est dans $[-\frac{\pi}{2}-\delta, -\frac{\pi}{2}+\delta]$, si on donne l'instruction `sin(x)` à la machine, la valeur renvoyée sera exactement `-1`. Expliquer l'allure du graphique.
7. * Que se passe-t-il si on ne suppose plus que $f''(x) \geqslant h > 0$ sur $[a,b]$ (tout en conservant l'hypothèse que f est unimodale) ? On peut prendre l'exemple du cas où on essaye de minimiser $x \mapsto x^4+1$ sur $[-1,1]$ par la même méthode.