Méthodes numériques : optimisation.
Partiel du 10 mars 2016

Recommandations

Le partiel est probablement long, mais le barème en tiendra compte. Les trois exercices sont indépendants. Pour les questions demandant d'écrire du code en python, on supposera que les librairies `numpy` et `matplotlib` ont déjà été chargées, par exemple dans IPython avec la directive `%pylab inline`.

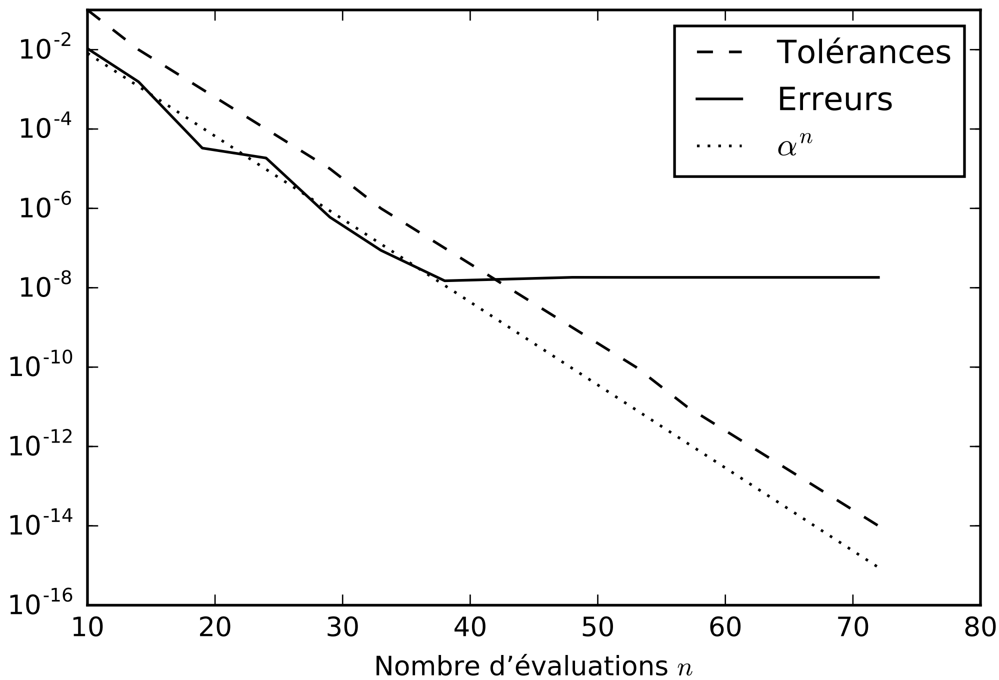
Les questions de chaque exercice ne sont pas censées être bloquantes pour les suivantes. N'hésitez donc pas à admettre des résultats et passer à la suite.

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

1 Méthode de la section dorée

1. Pourquoi la méthode de la section dorée appliquée à la fonction $x \mapsto \sqrt{1+x^2}$ entre les points $a_0 = -1$ et $b_0 = 2$ fournit-elle des suites $(a_k)_{k\in\mathbb{N}}$, $(b_k)_{k\in\mathbb{N}}$, $(c_k)_{k\in\mathbb{N}}$, et $(d_k)_{k\in\mathbb{N}}$ qui convergent vers 0 ? Quelle est la vitesse de convergence ? Que se passe-t-il dans le cas où l'on prend $a_0 = 1$ et $b_0 = 2$?
2. Écrire une fonction en python appelée `SectionDoree` qui prend en argument une fonction f, deux points $a < b$, une tolérance ε, et qui effectue la méthode de la section dorée sur f entre a et b en s'arrêtant dès que les deux extrémités de l'intervalle sont à moins de ε l'un de l'autre. La fonction renverra le point correspondant au milieu de l'intervalle final.
3. Coder les instructions suivantes en python :
   — Définir une fonction `fcompteur` qui prend en argument un réel x et qui renvoie $\sqrt{1+x^2}$ tout en incrémentant une variable globale jouant le rôle de compteur.
   — Utiliser cette fonction pour effectuer la méthode de la section dorée sur cette fonction entre -1 et 2, pour différentes tolérances données, et utiliser la variable de compteur pour obtenir le nombre de fois où la fonction `fcompteur` aura été utilisée dans chaque cas.
   — Visualiser la convergence linéaire effective vers 0 du résultat de l'algorithme en fonction du nombre d'appels n de la fonction f. Afficher également sur ce même graphique les tolérances utilisées ainsi que les points (n, α^n), où $\alpha = \frac{\sqrt{5}-1}{2}$.

   Voilà ce que l'on devrait obtenir :



4. Expliquer l'allure du graphique, et en particulier expliquer pourquoi les erreurs se stabilisent autour de 10^{-8}, même si les tolérances demandées sont bien plus petites.
5. * Que se passerait-il si on voulait essayer de minimiser $x \mapsto x^4 + 1$ sur $[-1, 2]$ par la même méthode ? Autour de quelle valeur les erreurs se stabiliseraient ?

2 Séparation des variables pour la descente de gradient à pas fixe

1. On suppose que l'on a une fonction $f : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ qui s'écrit

$$f(x, y) = g(x) + h(y),$$

avec $g : \mathbb{R}^n \to \mathbb{R}$ et $h : \mathbb{R}^p \to \mathbb{R}$ de classe C^1, bornées inférieurement. On se fixe (x_0, y_0) dans $\mathbb{R}^n \times \mathbb{R}^p$ et on pose $S_0^g = \{x \in \mathbb{R}^n, g(x) \leqslant g(x_0)\}$ et $S_0^h = \{y \in \mathbb{R}^p, h(y) \leqslant h(y_0)\}$. On suppose de plus que ∇g est L_g-Lipschitz sur S_0^g et que ∇h est L_h-Lipschitz sur S_0^h. Montrer que pour un L bien choisi (dépendant de L_g et L_h) et pour un pas $\alpha < \frac{2}{L}$, l'algorithme de descente de gradient à pas fixe α appliqué à la fonction f et partant du point (x_0, y_0) fournit bien une suite de points $(x_k, y_k) \in \mathbb{R}^n \times \mathbb{R}^p$ qui vérifient :
— Pour tout $k \in \mathbb{N}$, $(x_k, y_k) \in S_0^g \times S_0^h$.
— La suite $(f(x_k, y_k))_{k\in\mathbb{N}}$ est décroissante et converge vers une limite finie.
— La suite $(\nabla f(x_k, y_k))_{k\in\mathbb{N}}$ converge vers 0 dans $\mathbb{R}^n \times \mathbb{R}^p$.

2. On suppose que $g(x) = \frac{1}{2}\langle x, A_g x\rangle_n + \langle b_g, x\rangle_n$ et $h(y) = \frac{1}{2}\langle y, A_h y\rangle_p + \langle b_h, y\rangle_p$, où $\langle \cdot, \cdot\rangle_n$ et $\langle \cdot, \cdot\rangle_p$ sont les produits scalaires dans $\mathbb{R}^n$ et $\mathbb{R}^p$, où A_g et A_h sont des matrices symétriques de taille n et p, et où $b_g \in \mathbb{R}^n$ et $b_h \in \mathbb{R}^p$. On note ℓ_g et ℓ_h (resp. L_g et L_h) les plus petites (resp. les plus grandes) valeurs propres de A_g et de A_h.
   (a) À quelle condition sur ℓ_g et ℓ_h les fonctions g et h admettent-elles un unique minimum ?
   (b) On suppose que A_g et A_h sont symétriques définies positives. Pour quelles valeurs du pas α la descente de gradient à pas fixe pour f (définie comme à la question précédente par $f(x, y) = g(x) + h(y)$) est-elle convergente quel que soit $(x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^p$? Que peut-on dire sur la vitesse de convergence ? Donner la valeur du pas qui minimise le taux de convergence linéaire obtenu.

3. On suppose que $f : \mathbb{R}^2 \to \mathbb{R}$ est donnée par $f(x, y) = \sqrt{1 + x^2} + \sqrt{1 + y^4}$.
   On note $(x_k, y_k) \in \mathbb{R}^2$ l'itérée de la méthode de gradient à pas fixe α appliquée à f en partant de $(x_0, y_0) \in \mathbb{R}^2$.
   (a) Que faut-il prendre comme pas α pour être assuré de la convergence de la suite des x_k quelle que soit la valeur initiale de x_0 ? Que peut-on alors dire de la vitesse de convergence de $(x_k)_{k\in\mathbb{N}}$?
   (b) Montrer que quel que soit α, si on prend y_0 suffisamment petit, alors la suite des y_k converge vers 0. Que peut-on alors dire de la vitesse de convergence de $(y_k)_{k\in\mathbb{N}}$?
   (c) Montrer que quel que soit (x_0, y_0), on peut trouver un pas α tel que la suite des (x_k, y_k) converge vers le minimum de f. Que peut-on dire de la vitesse de convergence de la suite des (x_k, y_k) ?
   (d) * On suppose que la suite des y_k converge vers 0 avec $y_k \neq 0$ pour tout k. Montrer que $\frac{1}{y_{k+1}^2} - \frac{1}{y_k^2}$ converge vers 2α lorsque $k \to \infty$. En déduire que $y_k \sim \frac{1}{\sqrt{2\alpha k}}$ lorsque $k \to \infty$.

3 Méthodes de Newton et de la sécante en dimension 1

1. On considère la fonction $f : x \mapsto \sqrt{1+x^2}$ de $\mathbb{R}$ dans $\mathbb{R}$. On cherche à minimiser f en approchant un zéro de f' par la méthode de la Newton. On se fixe un réel $x_0 \in \mathbb{R}$.
   (a) Calculer la dérivée et la dérivée seconde de f.
   (b) Montrer que les itérées de la méthode de Newton sont bien définies pour tout $k \geqslant 0$, et sont données par la relation de récurrence $x_{k+1} = -x_k^3$.
   (c) En déduire l'ensemble des x_0 pour lesquels la suite converge, et donner l'ordre de convergence de la suite $(x_k)_{k\in\mathbb{N}}$ dans ces cas-là. Est-ce en contradiction avec les résultats donnés dans le cours ? Pourquoi dit-on que la méthode de Newton n'est pas robuste ?
2. On s'intéresse aux fonctions unimodales sur $\mathbb{R}$, de classe C^2 et dont la dérivée est strictement convexe. On se donne une telle fonction, que l'on note g, et on note x_* son unique minimum sur $\mathbb{R}$.
   (a) La fonction f de la première question satisfait-elle ces hypothèses ? Donner un exemple d'une telle fonction (un bon dessin vaut mieux qu'un long discours).
   (b) Démontrer que pour tout $x \in \mathbb{R}$, on a $g''(x) > 0$.
   (c) En utilisant l'interprétation graphique de la méthode de Newton par rapport au graphe de g', mettre en évidence (à l'aide d'un dessin propre et d'arguments convaincants) le fait que si $x_0 > x_*$, alors les itérées de la méthode pour la fonction g en partant de x_0 sont bien définies et satisfont $x_* < x_{k+1} < x_k$ pour tout $k \geqslant 0$.
   (d) En déduire que la méthode de Newton appliquée à la fonction g définit bien une suite d'itérées qui converge vers x_* quel que soit le point de départ $x_0 \in \mathbb{R}$.
   (e) En utilisant l'interprétation graphique de la méthode de la sécante par rapport au graphe de g', mettre en évidence (à l'aide d'un dessin propre et d'arguments convaincants) le fait que si $x_0 > x_1 > x_*$, alors les itérées de la méthode pour la fonction g en partant de x_0 et x_1 sont bien définies et satisfont $x_* < x_{k+2} < x_{k+1} < x_k$ pour tout $k \geqslant 0$, puis que la suite x_k converge vers x_*.
   (f) * Montrer que dans tous les cas où x_0, x_1 et x_* sont deux à deux distincts, la méthode de la sécante définit bien une suite d'itérées qui converge vers x_*.
3. * On considère la fonction f de la question 1. Montrer que si la suite des itérées de la méthode de la sécante est bien définie jusqu'au rang $k \geqslant 1$ et que $x_k \neq x_{k-1}$, alors elle est bien définie au rang $k+1$ et que l'on a alors

$$x_{k+1} = -x_k x_{k-1} \frac{x_k\sqrt{1+x_{k-1}^2} + x_{k-1}\sqrt{1+x_k^2}}{\sqrt{1+x_{k-1}^2} + \sqrt{1+x_k^2}},$$

   puis en déduire que $|x_{k+1}| \leqslant |x_k||x_{k-1}| \max(|x_k|, |x_{k-1}|)$.
   En déduire que si $0 < |x_1| < |x_0| < 1$ alors la suite des $|x_k|$ est bien définie, strictement décroissante et il existe $\alpha \in]0,1[$ tel que pour tout $k \geqslant 0$, on ait $|x_k| \leqslant \alpha^{2^k}$ (on pourra trouver un α qui convient pour $k=0$ et $k=1$ et montrer le cas général par récurrence, on pourra également admettre — ou démontrer — que $|x_k|$ n'est jamais nul). Que peut-on dire sur l'ordre de convergence de la suite ? Est-ce en contradiction avec les résultats donnés dans le cours ?