Méthodes numériques : optimisation.
Partiel du 8 mars 2017

Recommandations

Le partiel est assez 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 Jupyter 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 Gradient à pas fixe pour une fonction quadratique

On considère la fonction f de $\mathbb{R}^n$ dans $\mathbb{R}$ donnée par $f(x) = \frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle$, où A est une matrice symétrique, pas forcément définie positive et où b est un vecteur de $\mathbb{R}^n$. On note $(e_1, \ldots, e_n)$ une base orthonormale de vecteurs propres de la matrice A associée aux valeurs propres $(\lambda_1, \ldots, \lambda_n)$.

1. En décomposant $b = \sum_{i=1}^n b_i e_i$, montrer qu'une condition nécessaire et suffisante sur A et b pour que la fonction f admette un minimum sur $\mathbb{R}^n$ est que pour tout $1 \leqslant i \leqslant n$, on ait $\lambda_i > 0$ ou $\lambda_i = b_i = 0$.(on pourra calculer $f(x)$ pour $x = \sum_{i=1}^n t_i e_i$). Montrer que cette condition est équivalente à A symétrique positive et $b \in \mathrm{Im}(A)$.
2. Calculer le gradient de f. Donner une condition nécessaire et suffisante sur A pour que la fonction f admette un unique minimum sur $\mathbb{R}^n$.

On effectue la méthode de descente de gradient à pas fixe α à f, à partir d'un point $x_0 \in \mathbb{R}^n$.

3. Écrire la relation de récurrence entre les itérées x_k de cette méthode. Montrer, en notant $r_k = \nabla f(x_k)$, qu'on a $r_{k+1} = r_k - \alpha A r_k$.
4. On décompose r_k dans la base $(e_1, \ldots, e_n)$ en écrivant $r_k = \sum_{i=1}^n r_{k,i} e_i$. Pour $i \in \{1, \ldots, n\}$ fixé, donner la relation de récurrence entre les réels $r_{k,i}$.

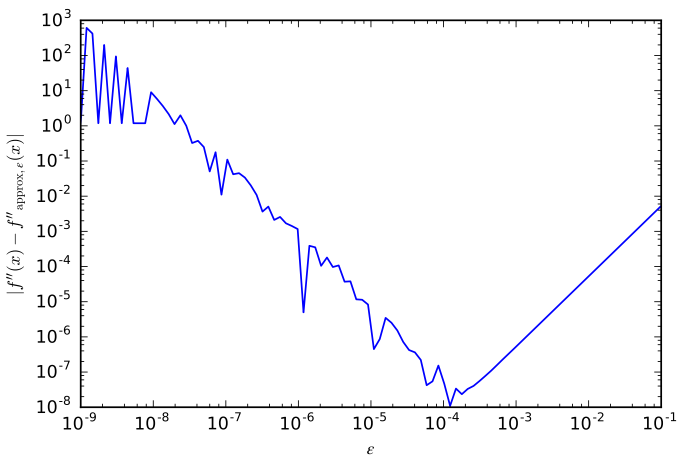
On se place pour les questions suivantes dans le cas où A est symétrique positive (mais pas forcément définie) et où $b \in \mathrm{Im}(A)$, et on note ℓ la plus petite valeur propre non-nulle de A, L sa plus grande valeur propre, et $\rho(\alpha) = \max\{|1 - \alpha\ell|, |1 - \alpha L|\}$. On suppose de plus que A est non-nulle, de sorte que ℓ est bien définie, et que $0 < \ell \leqslant L$.

5. Montrer que $r_{0,i} = 0$ dès que $\lambda_i = 0$. En déduire que l'on a $\|r_{k+1}\| \leqslant \rho(\alpha)\|r_k\|$ pour tout $k \in \mathbb{N}$. Montrer qu'il existe des conditions initiales x_0 telles que $r_0 \neq 0$ et où l'inégalité précédente est une égalité pour tout $k \in \mathbb{N}$.
6. En déduire que si $\alpha \in]0, \frac{2}{L}[$, la suite $(x_k)_{k\in\mathbb{N}}$ converge linéairement vers un point de minimum x_* de f, à un taux inférieur ou égal à $\rho(\alpha)$. Quel est le pas α qui minimise $\rho(\alpha)$?
7. * Montrer (toujours pour $\alpha \in]0, \frac{2}{L}[$) que pour presque tout x_0 (au sens de la mesure de Lebesgue sur $\mathbb{R}^n$), le taux de convergence de x_k vers x_* est exactement $\rho(\alpha)$.

2 Méthode de Newton approchée par différence finies

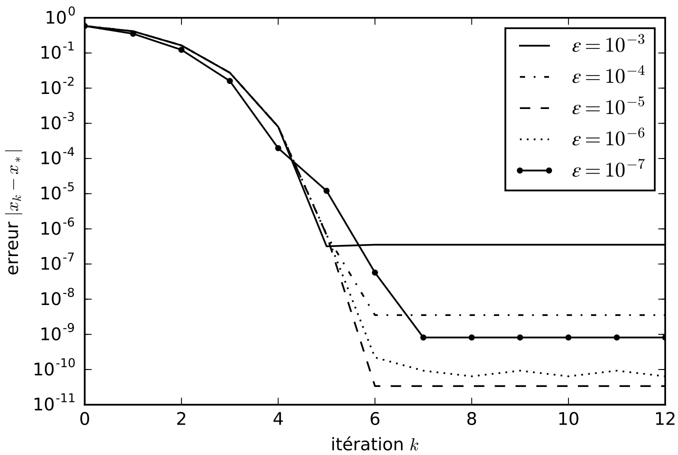
On cherche à remplacer les calculs de dérivées et dérivées secondes dans la méthode de Newton par des approximations par différences finies. On veut tester cette méthode sur la fonction $f : x \mapsto x + \frac{2}{x}$.

1. Étudier les variations de f sur $]0, +\infty[$, et en déduire l'unique point de minimum x_*. Calculer la dérivée seconde de f.
2. On approxime $f''(x)$ par la formule $\frac{f(x+\varepsilon)+f(x-\varepsilon)-2f(x)}{\varepsilon^2}$. Montrer que cette formule correspond à approximer $f''(x)$ par différence finie centrée (de paramètre $\frac{\varepsilon}{2}$), après avoir approximé $f'(x+\frac{\varepsilon}{2})$ et $f'(x-\frac{\varepsilon}{2})$ également par différences finies centrées (de paramètre $\frac{\varepsilon}{2}$).
3. Coder en python les fonctions f et f'', puis écrire un code correspondant à évaluer, pour une liste de ε allant entre 10^{-9} et 10^{-1}, l'erreur faite en approximant numériquement $f''(x)$ par cette formule, et afficher (dans des échelles appropriées) ces erreurs en fonction de ε (on se fixera un x au préalable). Voici ce que l'on obtient pour $x = \frac{3}{2}$; expliquer brièvement ce que l'on observe et quelle en est la raison (l'explication théorique précise est l'objet de la question suivante).



4. On se donne une fonction $f \in C^4([a,b])$ telle que $|f^{(4)}(x)| \leqslant C_4$ pour tout $x \in [a,b]$, où C_4 est une constante de l'ordre de grandeur de 1. Montrer que si $[x-\varepsilon, x+\varepsilon] \subset [a,b]$, alors
$$\left|\frac{f(x+\varepsilon)+f(x-\varepsilon)-2f(x)}{\varepsilon^2} - f''(x)\right| \leqslant \frac{C_4}{12}\varepsilon^2.$$
On suppose que la machine ne calcule pas exactement f, mais une approximation (notée $\hat{f}$) à une erreur numérique η près (avec $\eta \ll 1$) : $|\hat{f}(x) - f(x)| \leqslant \eta$ pour $x \in [a,b]$. Montrer que l'erreur totale d'approximation est majorée par l'estimation suivante :
$$\left|\frac{\hat{f}(x+\varepsilon)+\hat{f}(x-\varepsilon)-2\hat{f}(x)}{\varepsilon^2} - f''(x)\right| \leqslant 4\frac{\eta}{\varepsilon^2} + \frac{C_4}{12}\varepsilon^2.$$
En déduire l'ordre de grandeur de ε à choisir pour minimiser cette estimation d'erreur totale. Interpréter alors plus précisément le graphique de la question précédente.

5. Rappeler la formule de récurrence liant les itérées de la méthode de Newton pour la minimisation d'une fonction d'une variable. On veut effectuer une méthode approchée par différence finie, expliquer en quoi la formule $x_{k+1} = x_k - \frac{\varepsilon[f(x_k+\varepsilon)-f(x_k-\varepsilon)]}{2[f(x_k+\varepsilon)+f(x_k-\varepsilon)-2f(x_k)]}$ correspond à cela.
6. Coder en python la variante de la méthode de Newton donnée par la formule précédente, en gardant en mémoire la liste des itérées, pour un nombre fixe N d'itérations (on se fixera un x_0 et un ε au préalable). Afficher dans une échelle appropriée la suite des erreurs $|x_k - x_*|$ en fonction de k. On sera attentif à effectuer, pour chaque itération, le moins d'appels possibles à la fonction f. Voici ce que l'on obtient, pour plusieurs valeurs de ε (ici on a pris $x_0 = 2$ et $N = 12$).



Lorsqu'on exécute ce code en prenant $\varepsilon = 10^{-8}$, on peut obtenir pour certains x_0 le message d'erreur suivant : `ZeroDivisionError: float division by zero`. Expliquer pourquoi.
7. * Expliquer rapidement pourquoi l'erreur d'approximation de $f'(x_k)$ par la formule des différences finies centrées $\frac{\hat{f}(x_k+\varepsilon)-\hat{f}(x_k-\varepsilon)}{2\varepsilon}$ est de l'ordre de grandeur de $\frac{\eta}{\varepsilon} + \varepsilon^2$. Expliquer pourquoi, en utilisant la question 3, l'ordre de grandeur du ε pour laquelle la formule de récurrence $x_{k+1} = x_k - \frac{\varepsilon[\hat{f}(x_k+\varepsilon)-\hat{f}(x_k-\varepsilon)]}{2[\hat{f}(x_k+\varepsilon)+\hat{f}(x_k-\varepsilon)-2\hat{f}(x)]}$ donne un résultat le plus proche possible de la méthode de Newton, est $\varepsilon \approx \eta^{\frac{1}{4}}$. Pourquoi la meilleure précision finale d'approximation de x_* sur les exemples du graphique ci-dessus est elle atteinte pour ε entre 10^{-6} et 10^{-5}, et non pas pour ε de l'ordre de 10^{-4}, comme on pourrait alors s'y attendre (et au vu du graphique de la question 2) ?

3 Méthode de descente de gradient à pas fixe pour une fonction convexe

On se donne une fonction $f \in C^2(\mathbb{R}^n, \mathbb{R})$ telle que $0 \leqslant H_f(x) \leqslant L\,\mathrm{I}_n$ pour tout $x \in \mathbb{R}^n$ (on rappelle que H_f est la hessienne de f, et que l'hypothèse signifie que $0 \leqslant \langle h, H_f(x)h\rangle \leqslant L\|h\|^2$ pour tout $h \in \mathbb{R}^n$). On suppose de plus qu'il existe x_* tel que $\nabla f(x_*) = 0$.

On effectue une descente de gradient à pas fixe α avec $0 < \alpha < \frac{2}{L}$, à partir de $x_0 \in \mathbb{R}^n$. Le but de l'exercice est de montrer la convergence des x_k, et une estimation de vitesse de convergence de $f(x_k)$ vers le minimum de f.

1. Montrer que pour tout x de $\mathbb{R}^n$ et tout h de $\mathbb{R}^n$, on a
$$0 \leqslant f(x+h) - f(x) - \langle \nabla f(x), h\rangle \leqslant \frac{L}{2}\|h\|^2.$$
En déduire que x_* est un point de minimum global de f. Est-il unique ?
2. Montrer que $f(x_{k+1}) - f(x_k) \leqslant -\alpha(1 - \frac{L}{2}\alpha)\|\nabla f(x_k)\|^2$.
3. Montrer que $x_{k+1} - x_* = \int_0^1 [\mathrm{I}_n - \alpha H_f(x_* + t(x_k - x_*))](x_k - x_*)\mathrm{d}t$. En déduire que la suite $(\|x_k - x_*\|)_{k\in\mathbb{N}}$ est décroissante. On pourra utiliser sans démonstration le fait que $\| \int_0^1 u(t)\mathrm{d}t\| \leqslant \int_0^1 \|u(t)\|\mathrm{d}t$ pour toute fonction $u \in C^0(\mathbb{R}, \mathbb{R}^n)$.
4. En déduire qu'il existe une valeur d'adhérence $\bar{x}$ de la suite $(x_k)_{k\in\mathbb{N}}$, puis que $\nabla f(\bar{x}) = 0$. En conclure que $x_k \to \bar{x}$ lorsque $k \to \infty$.

On s'intéresse maintenant à une estimation de la convergence de $f(x_k)$ vers le minimum de f. On note $\varepsilon_k = f(x_k) - f(x_*)$.

5. Montrer que $\varepsilon_k \leqslant \|x_0 - x_*\|\|\nabla f(x_k)\|$.
6. En déduire que $\varepsilon_{k+1} \leqslant \varepsilon_k(1 - \frac{\alpha(1-\frac{L}{2}\alpha)}{\|x_0-x_*\|^2}\varepsilon_k)$. En déduire que $\frac{1}{\varepsilon_{k+1}} \geqslant \frac{1}{\varepsilon_k} + \frac{\alpha(1-\frac{L}{2}\alpha)}{\|x_0-x_*\|^2}$.
7. En conclure que $\frac{1}{\varepsilon_k} \geqslant \frac{1}{\varepsilon_0} + k\frac{\alpha(1-\frac{L}{2}\alpha)}{\|x_0-x_*\|^2}$, et que l'on a donc une estimation de la forme
$$f(x_k) - f(x_*) \leqslant \frac{f(x_0) - f(x_*)}{1 + C\,k},$$
où C est une constante dépendant des données du problème (fonction f, pas α) et du point initial x_0.
8. * Montrer qu'on ne peut pas espérer faire bien mieux asymptotiquement, en étudiant la fonction $f : x \in \mathbb{R} \mapsto x^p$ avec p un (grand) entier pair.
Indication : montrer que $\frac{1}{(x_{k+1})^{p-2}} - \frac{1}{(x_k)^{p-2}}$ converge, et que $x_k \sim Ck^{\frac{-1}{p-2}}$.