Méthodes numériques : optimisation.
L3 2016–2017 — 2e semestre.
Feuille de TD n° 2 : Méthodes de descente de gradient.

1 Retour en dimension un.

(a) Montrer que les méthodes de Newton et de la sécante pour résoudre l'équation $f'(x) = 0$ peuvent être vues dans certains cas comme des méthodes de descente de gradient. Quel est alors la valeur du pas α_k, et quelles sont les conditions à satisfaire pour rentrer dans le cadre des méthodes de descente ?

(b) On considère la fonction $f : x \mapsto x^2 - \frac{1}{4}x^4$. Écrire la relation de récurrence vérifiée par les points x_k correspondant à une descente de gradient à pas fixe α pour la fonction f.

On suppose que aucun des x_k n'est nul (sinon la suite est stationnaire). Pour quelles valeurs de α est-on sûr que la suite des $|x_k|$ est décroissante dès que x_0 est suffisamment proche de 0 ?

Montrer que dans tous ces cas, la suite des x_k converge vers 0. Dans quels cas la convergence est-elle linéaire ou superlinéaire (donner alors le taux de convergence linéaire, ou l'ordre de convergence s'il existe) ?

(c) * On conjecturera en TP un équivalent de $|x_n|$ lorsque $n \to \infty$, dans le cas particulier où la suite est convergente mais ne converge pas linéairement. Démontrer cette conjecture.

Indication : calculer la limite lorsque $n \to \infty$ de $\frac{1}{x_{n+1}^2} - \frac{1}{x_n^2}$.

(d) On considère la fonction $f : x \mapsto \frac{x^2}{1+x^2}$. Écrire la relation de récurrence vérifiée par les points x_k pour la méthode de descente de gradient à pas fixe α. On suppose encore qu'aucun des x_k n'est nul.

Pour quelles valeurs de α la suite converge-t-elle lorsque x_0 est suffisamment proche de 0 ?

Que va-t-il se passer si l'on prend x_0 trop grand ?

En séance de TP, on représentera graphiquement ces suites pour visualiser leur vitesse de convergence.

2 Méthode de gradient à pas fixe, étude de cas en dimension 2.

On va tester, en séance de TP, la méthode de descente de gradient sur la fonction $f_a : (x_0, x_1) \mapsto 1 - \frac{1}{1+ax_0^2+x_1^2}$, où $a > 0$ est un paramètre qu'on pourra changer pour voir comment se comporte la méthode en fonction de a.

Attention au changement de notation : ici x_i est la $i$ème coordonnée d'un vecteur $X \in \mathbb{R}^n$. On commence par $i = 0$ pour être cohérent avec la numérotation en python. On notera les itérées X_k. On utilisera deux indices si on veut préciser les coordonnées, par exemple en dimension 2 on écrit $X_k = (x_{0,k}, x_{1,k})$.

(a) Calculer le gradient de f_a en tout point, et rappeler l'expression des itérées de la méthode de descente de gradient à pas fixe.

(b) Écrire un code en python comprenant les parties suivantes :
- Définition de la fonction ∇f_a, avec inclusion d'un compteur pour pouvoir compter le nombre d'appels.
- Définition d'une fonction qui, pour un pas fixe α, un point initial X_0, et une tolérance ε donnée, calcule les points X_k suivant la méthode de descente de gradient à pas fixe, en s'arrêtant dès que $\|\nabla f_a\| < \varepsilon$, et renvoie le dernier point calculé.
- Test de cette fonction pour différentes valeurs de ε, et affichage d'un graphique pour illustrer le taux de convergence effectif de la méthode.

On s'attachera à essayer de minimiser le nombre d'appels à la fonction ∇f_a.

(c) Calculer la Hessienne de f_a en 0. Quel pas prendre pour appliquer la méthode de descente de gradient à pas fixe ?

Quel est le taux de convergence auquel on s'attend ?

(d) On veut approximer le gradient par différence finies. On peut choisir une des approximations suivantes (e_i étant le vecteur numéro i de la base), que l'on doit calculer pour $0 \leqslant i < n$.
- Différences finies à droite : $\partial_i f(X) \approx \dfrac{f(X + \varepsilon e_i) - f(X)}{\varepsilon}$.
- Différences finies centrées : $\partial_i f(X) \approx \dfrac{f(X + \varepsilon e_i) - f(X - \varepsilon e_i)}{2\varepsilon}$.

Discuter de l'intérêt de l'une ou l'autre des méthodes. Quel serait le bon choix d'ordre de grandeur de ε dans chaque cas ?

(e) Quel est le taux de convergence effectif auquel on s'attend lorsque l'on utilise chacune des approximations précédentes, lorsque l'on s'intéresse au nombre d'évaluations de la fonction f_a cette fois-ci ?

Jusqu'à quelle tolérance peut-on aller pour que la méthode donne un résultat ayant du sens ?

3 Problème d'application : recherche de trajectoires fermées sur un billard.

On se donne un convexe de $\mathbb{R}^2$, dont le bord est noté Γ. On cherche à placer n points $M_0, \ldots, M_{n-1}$ sur Γ qui correspondent à une trajectoire de billard parfaite : l'angle entre la tangente au bord et la trajectoire avant rebond doit être le même que celui entre la tangente au bord et la trajectoire après rebond.

Si on se donne M_0 et un angle (donc deux paramètres), alors les points $M_1, \ldots M_{n-1}$ sont uniquement déterminés. On aimerait que la trajectoire après rebond en M_{n-1} passe par M_0 avec le même angle. On a donc deux conditions à satisfaire, avec deux paramètres, donc on espère qu'il puisse y avoir une solution !

(a) On modélise d'abord notre problème. On se donne $t \mapsto \gamma(t) \in \mathbb{R}^2$ une paramétrisation 2π-périodique du bord Γ. Par exemple si le convexe est une ellipse, on prendrait $\gamma(t) = (a \cos t, b \sin t)$. On suppose que γ est de classe C^1 et que $\gamma'(t) \neq 0$ pour tout $t \in \mathbb{R}$.

On pose $M_i = \gamma(t_i)$ pour $0 \leqslant i < n$, et on pose $T_i = \frac{\gamma'(t_i)}{\|\gamma'(t_i)\|}$ le vecteur unitaire tangent à la courbe en M_i. Pour des raisons de notation, on pose $M_n = M_0$ et $T_n = T_0$. Lorsque les points consécutifs sont distincts, on pose $U_i = \frac{M_{i+1} - M_i}{\|M_{i+1} - M_i\|}$ le vecteur unitaire dirigé de M_i à M_{i+1} (et de même on pose $U_n = U_0$).

Faire un dessin.

(b) Montrer que la condition d'être une trajectoire de billard parfaite équivaut à

$$\forall i \in [\![1, n]\!], T_i \cdot (U_i - U_{i-1}) = 0,$$

où on a posé $U_n = U_0$ pour simplifier les notations.

(c) On veut transformer le problème en un problème d'optimisation. On va montrer que les solutions d'un problème d'optimisation bien choisi conduisent à des solutions du problème original.

On pose L la fonction de $\mathbb{R}^n$ dans $\mathbb{R}$ qui donne la longueur totale de la trajectoire passant successivement pas les points (on ne se préoccupe pas de savoir si la trajectoire est une trajectoire de billard).

$$L(t_0, \ldots, t_{n-1}) = \|\gamma(t_0) - \gamma(t_{n-1})\| + \sum_{i=1}^{n-1} \|\gamma(t_i) - \gamma(t_{i-1})\|.$$

Montrer que la fonction L admet un maximum global sur $\mathbb{R}^n$ (on pourra utiliser la périodicité). Que se passe-t-il si n est pair ?

(d) On considère maintenant $(t_0, \ldots, t_{n-1})$ un point de maximum local de la fonction L. Montrer qu'alors on a pour tout $i \in [\![1, n]\!], M_i \neq M_{i-1}$ (on pourra supposer que le billard est strictement convexe pour simplifier la démonstration, mais en fait cela fonctionne tout le temps). En déduire que la fonction L est différentiable en $(t_0, \ldots, t_{n-1})$ et qu'alors la trajectoire est une trajectoire de billard parfaite.

On a donc obtenu que tout point de maximum local (et il en existe) correspond à une trajectoire de billard parfaite. Il existe donc au moins une trajectoire parfaite, et on va essayer d'en approximer numériquement.

4 Séparation des variables pour la descente de gradient à pas fixe (partiel de mars 2016)

(a) 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 $\mathbb{R}^n \times \mathbb{R}^p$.

(b) 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 précédemment 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 ? Quel est le pas qui minimise le taux de convergence linéaire obtenu ?

(c) 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 $k \to \infty$.