TP2 : Méthodes de gradient

1 Lignes de niveau

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction. Soit $t \in \mathbb{R}$. On appelle *lignes de niveau inférieur* t (ou *de sous-niveau*) l'ensemble des points suivant :

$$\mathrm{niv}_{\leqslant t} f = \left\{ x \in \mathcal{X} \mid f(x) \leqslant t \right\} = f^{-1}(]-\infty, t[)$$

et *ligne de niveau* t l'ensemble des points suivant :

$$\mathrm{niv}_t f = \left\{ x \in \mathcal{X} \mid f(x) = t \right\} = f^{-1}(t).$$

Exercice 1. Tracé des lignes de niveaux, des gradients

Soit $(a, b) \in \mathbb{R}^2$. On considère la fonction quadratique

$$J : \left\{ \begin{array}{ccc} \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ (x, y) & \longmapsto & \frac{1}{2} a\, x^2 + \frac{1}{2} b\, y^2 \end{array} \right.$$

1. Pour différentes valeurs de a et b, tracer des lignes de niveau de J sur le carré $[-5, 5] \times [-5, 5]$.
   Outils recommandés : `np.meshgrid, plt.contour`
2. Pour $(x, y) \in \mathbb{R}^2$, calculer $\nabla J(x, y)$.
3. Sur la même figure, tracer des lignes de niveau de J, puis le vecteur $\nabla J(x_0, y_0)$ au point $(x_0, y_0) \in \mathbb{R}^2$.
   Outils recommandés : `plt.quiver` pour tracer des vecteurs.
4. Reprenez la question précédente en traçant le champs de vecteurs $\nabla J(x, y)$ aux points $(x, y) \in [-5, 5]^2$.

2 Méthode de gradient à pas fixe

Le principe de la méthode de gradient à pas fixe est de faire, à partir d'un point initial $v_0 \in \mathbb{R}^n$, les itérations

$$v_{k+1} = v_k - \rho \nabla F(v_k)$$

où ρ est le pas, qui est constant.

Exercice 2. Gradient à pas fixe. On s'intéresse au cas "d'école" suivant

$$F(x, y) = x^2 + y^2.$$

1. Tracer quelques lignes de niveaux.
2. Montrer que la fonction est α-convexe, et donner α.
3. Montrer que la fonction est à gradient Lipschitz, on donnera la constante de Lipschitz.
4. Ecrire une fonction `gf(GradF, v0, rho, tol, NitMax)` qui implémente la méthode de gradient à pas fixe où
   `GradF` est une fonction qui calcule le gradient de F au point $v = (x, y)$
   `v0` le point initial
   `rho` le pas fixe choisi
   `tol` la tolérance demandée qui servira pour le critère d'arrêt
   `NitMax` le nombre maximal d'itérations autorisé

Cette fonction retourne
le minimum trouvé
la suite des itérés $v_k = (x_k, y_k)_k$
le nombre d'itérations pour atteindre la tolérance demandée
une variable booléenne qui prend la valeur `True` si la méthode a convergé et `False` sinon.

5. D'après votre cours, à quelle condition sur `rho` la méthode converge-t-elle ? Quelle est la valeur optimale de `rho` ?
6. Tester votre fonction pour 3 valeurs différentes de `rho`, que constatez vous ? Dans chaque cas, comparer le nombre d'itérations pour la même tolérance.

Exercice 3. Gradient à pas fixe - critère d'arrêt. On s'intéresse maintenant au cas suivant

$$F(x, y) = x^2 + 100\, y^2.$$

1. Appliquer votre fonction `gf` avec `rho = 1.9999/1000` et $v_0 = (1, 1)$.
2. Afficher sur un graphique la suite de points $(x_k, y_k)_k$.
3. Afficher sur un deuxième graphique la fonction $k \longmapsto (F(x_k, y_k))_k$.
4. Afficher sur un troisième graphique la fonction $k \longmapsto (\|(x_k, y_k)\|)_k$.
5. Ecrire une fonction `gf2(GradF, v0, rho, tol, NitMax)` comme dans l'exercice précédent, mais en choisissant un autre critère d'arrêt (lorsque le gradient de F est suffisamment petit, ou bien lorsque les itérés ne progressent plus suffisamment, ou bien lorsque les valeurs de F ne changent plus suffisamment)
6. Appliquer votre fonction `gf2` pour minimiser la fonction F. Observez vous une différence dans le résultat en fonction du critère d'arrêt ? Comparez la valeur de la fonction aux points solutions donnés par `gf` et `gf2`.

3 Méthode de gradient à pas optimal

Dans la méthode de gradient à pas optimal, le pas est variable, à chaque itération, on choisit le pas qui permet de faire décroitre la fonction F le plus possible :

$$v_{k+1} = v_k - \rho_k \nabla F(v_k) \quad \text{où } \rho_k = \arg\min_{\rho > 0} F(v_k - \rho \nabla F(v_k))$$

Exercice 4. Gradient à pas optimal.

On s'intéresse au cas $F(x, y) = ax^2 + by^2$, avec $a, b \in \mathbb{R}$.

1. Montrer que si le pas optimal ρ_k existe alors il vérifie :
$$\langle \nabla F(v_k), \nabla F\big(v_k - \rho_k \nabla F(v_k)\big)\rangle = 0$$
2. En déduire que pour tout $k \in \mathbb{N}$:
$$\rho_k = \frac{a^2 x_k^2 + b^2 y_k^2}{2(a^3 x_k^2 + b^3 y_k^2)}$$
3. Ecrire une fonction qui implémente la méthode de gradient à pas optimal dans le cas considéré ici (utiliser la question précédente) : `gopt(GradF, v0, tol, NitMax)` où,
   `GradF` est une fonction qui calcule le gradient de F au point $v = (x, y)$.
   `v0` le point initial
   `tol` la tolérance demandée qui servira pour le critère d'arrêt
   `NitMax` le nombre maximal d'itérations autorisé
   Cette fonction retourne
   le minimum trouvé
   la suite des itérés $v_k = (x_k, y_k)_k$
   le nombre d'itération pour atteindre la tolérance demandée
   une variable booléenne qui prend la valeur `True` si la méthode a convergé et `False` sinon.
4. Tester votre code avec la fonction $F(x, y) = x^2 + 100\, y^2$, avec pour initialisation $v_0 = (1, 1)$ puis $v_0 = (0, \sqrt{2})$. Que constatez vous ?

5. Comparer les résultats avec ceux obtenus pour le gradient à pas fixe.

Exercice 5. Comparaison des méthodes.
On s'intéresse au cas $G(x,y) = x^2 + 2\,y^2$.

1. Avec l'exercice précédent, montrer par récurrence que le pas optimal est constant ici et vaut $\frac{1}{3}$, avec $v_0 \;=\; (20,10)$
2. Appliquer la méthode du gradient à pas fixe pour $\tau \in \{\frac{1}{3} - 0.01, \frac{1}{3}, \frac{1}{3} + 0.01, 0.9999\}$ avec comme initialisation $v_0 = (20,10)$
3. Appliquer la méthode du gradient à pas optimal ($\tau = \frac{1}{3}$) avec comme initialisation $v_0 \;=\; (20,10)$
4. Comparer les 2 méthodes : nombre d'itérations pour la même tolérance, afficher sur le même graphique les suites v_k. Qu'en pensez vous ?

4 Pour aller plus loin

Dans le cas général, il est difficile de calculer le pas optimal explicitement. Lorsque la fonction objectif J est deux fois différentiable, la formule de TAYLOR permet d'écrire au voisinage de v_k

$$J(v) \approx \underbrace{J(v_k) + \langle \nabla J(v_k), v - v_k\rangle + \frac{1}{2}\langle \text{Hess}\, J(v_k)\,(v - v_k), v - v_k\rangle}_{\tilde{J}_{v_k}(v)}$$

On notera que la fonction $\tilde{J}_{v_k}$ est une fonction quadratique. Dans ce cas, au lieu de calculer le pas optimal pour la fonction J, on se contente de calculer le pas optimal pour la fonction $\tilde{J}_{v_k}$, de sorte que la méthode du gradient à pas optimal devient

$$v_0 \in \mathcal{X} \qquad \text{et} \qquad \forall\, k \in \mathbb{N}, \qquad \begin{cases} \tilde{\rho}_k & \in \arg\min\limits_{\rho \in \mathbb{R}} \tilde{J}_{v_k}\big(v_k - \rho\, \nabla J(v_k)\big) \\ v_{k+1} = v_k - \tilde{\rho}_k\, \nabla J(v_k) \end{cases}$$

Exercice 6. Cas non quadratique.
On s'intéresse au cas $J(x,y) = (x + 2\,y)^4$.

1. Tracer quelques lignes de niveau de la fonction J.
2. Montrer que pour tout $(x,y) \in \mathbb{R}^2$,
$$\nabla J(x,y) = \begin{pmatrix} 4\,(x+2\,y)^3 \\ 8\,(x+2\,y)^3 \end{pmatrix} = 4\,(x+2\,y)^3 \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$
et
$$\text{Hess}\, J(x,y) = \begin{pmatrix} 12\,(x+2\,y)^2 & 24\,(x+2\,y)^2 \\ 24\,(x+2\,y)^2 & 48\,(x+2\,y)^2 \end{pmatrix} = 12\,(x+2\,y)^2 \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$$
3. Montrer que le pas optimal vaut, pour tout $k \in \mathbb{N}$ tel que $x_k + 2\,y_k \neq 0$,
$$\rho_k = \frac{1}{20\,(x_k + 2\,y_k)^2}.$$
4. Montrer que le pas optimal approché vaut, pour tout $k \in \mathbb{N}$ tel que $x_k + 2\,y_k \neq 0$,
$$\tilde{\rho}_k = \frac{1}{60\,(x_k + 2\,y_k)^2}.$$
5. Implémenter la méthode du gradient à pas optimal avec les pas exact ρ_k et approché $\tilde{\rho}_k$. On choisira comme initialisation $(x_0, y_0) = (1,1)$. Comparer les résultats et le nombre d'itérations.
6. Pour cette fonction, comment fonctionne la méthode de gradient à pas fixe (prendre `rho = 0.01` et `NitMax=6000`) ? Comparer le nombre d'itérations de chaque méthode.