Optimisation

Master 1 Statistique & Data Science, Ingénierie Mathématique, 2020-2021

Examen du mercredi 5 mai 2021

Nombre de pages du sujet : 3. Durée 2h. **Aucun document n'est autorisé**

Exercice 1. *(sur environ 9 points)*

On considère la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par

$$f(x_1, x_2) = x_1^2 + x_2.$$

1. Etude de f sur $\mathbb{R}^2$: Répondre rigoureusement aux questions suivantes.
   1.a) La fonction f est-elle convexe sur $\mathbb{R}^2$?
   1.b) La fonction f est-elle strictement convexe sur $\mathbb{R}^2$?
   1.c) La fonction f atteint-elle une valeur minimale sur $\mathbb{R}^2$?

On considère maintenant l'ensemble

$$U = \{x = (x_1, x_2) \in \mathbb{R}^2,\ x_1^2 + x_2^2 \leqslant 13 \text{ et } x_1 + x_2 \geqslant -1\}.$$

2. Propriétés de l'ensemble U :
   2.a) Représenter graphiquement U et préciser les coordonnées des points d'intersection qui apparaissent naturellement.
   2.b) Montrer que U est un ensemble convexe et compact.
   2.c) Introduire deux fonctions g_1 et g_2 pour décrire U comme un ensemble de contraintes inégalités et vérifier que tous les points de U vérifient la qualification de contraintes de Kuhn-Tucker (QCKT).

On cherche désormais à résoudre le problème d'optimisation sous contraintes :

$$\text{Trouver } x^\star \text{ tel que } x^\star \in U \text{ et } \quad f(x^\star) = \min_{x \in U} f(x) \qquad (\mathcal{P})$$

3. Justifier que le problème ($\mathcal{P}$) admet au moins une solution.
4. Justifier que $x \in U$ est solution de ($\mathcal{P}$) si et seulement si x vérifie les conditions de Karush-Kuhn-Tucker (KKT) que l'on précisera.
5. Déterminer toutes les solutions de ($\mathcal{P}$) à l'aide des conditions KKT.

Exercice 2. *(sur environ 16 points)*

Dans cet exercice on considère $g : \mathbb{R}^n \to \mathbb{R}$ une fonctionnelle quadratique

$$g(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + c,$$

où $A \in \mathcal{M}_n(\mathbb{R})$ est une matrice carrée symétrique définie positive, $b \in \mathbb{R}^n$ et $c \in \mathbb{R}$.

Partie I : Propriétés générales

On commence par **redémontrer** des résultats de cours.

1. Justifier que g est de classe $\mathcal{C}^\infty$ sur $\mathbb{R}^n$ et prouver que pour tout $x \in \mathbb{R}^n$,

$$\nabla g(x) = Ax - b \quad \text{et} \quad \nabla^2 g(x) = A.$$

2. Redémontrer rigoureusement que g admet un unique minimum global $x^\star = A^{-1}b$.

Partie II : Convergence de l'algorithme de gradient à pas fixe

Dans cette partie de l'exercice on étudie la convergence de l'algorithme de descente de gradient à pas fixe appliqué à la fonctionnelle quadratique g. Pour une fonction quelconque $f : \mathbb{R}^n \to \mathbb{R}$, cet algorithme est décrit par le pseudo-code suivant :

Algorithme 1 : Algorithme de descente de gradient à pas fixe

Données : Un point initial $x^{(0)} \in \mathbb{R}^n$, un seuil de tolérance $\varepsilon > 0$, un pas fixe $t > 0$
Résultat : Un point $x \in \mathbb{R}^n$ proche de $x^\star$
Initialiser x :
  $x \leftarrow x^{(0)}$;
  $k \leftarrow 0$;
tant que $\|\nabla f(x)\| > \varepsilon$ **faire**
  1. Mettre à jour x avec le pas fixe t dans la direction de descente $-\nabla f(x^{(k)})$:
    $x \leftarrow x^{(k+1)} = x^{(k)} - t\nabla f(x^{(k)})$;
    $k \leftarrow k + 1$;
fin

Dans la suite de cette partie $(x^{(k)})_{k\in\mathbb{N}}$ désigne la suite de vecteurs obtenues en appliquant l'Algorithme 1 à la fonction g avec un pas $t > 0$ et en partant de d'un point initial $x^{(0)} \in \mathbb{R}^n$.

3. Soit $k \in \mathbb{N}$. Montrer que

$$x^{(k+1)} - x^\star = (I_n - tA)(x^{(k)} - x^\star).$$

En déduire une expression de $x^{(k)} - x^\star$ en fonction de A, t, et $x^{(0)} - x^\star$.

4. Soit $0 < \lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$ les valeurs propres de A et $u_1, u_2, \ldots, u_n$ des vecteurs propres associés formant une base orthonormée de $\mathbb{R}^n$. En introduisant la décomposition du vecteur $x^{(0)} - x^\star$ dans la base $(u_1, u_2, \ldots, u_n)$, montrer que la suite $(x^{(k)})_{k\in\mathbb{N}}$ converge vers $x^\star$ quelque soit la valeur de $x^{(0)}$ si et seulement si

$$\forall i \in \{1, \ldots, n\}, \quad |1 - t\lambda_i| < 1.$$

5. Montrer que
$$\forall i \in \{1,\dots,n\}, \quad |1-t\lambda_i| \leqslant \max(|1-t\lambda_1|, |1-t\lambda_n|).$$
6. En déduire l'intervalle des valeurs de t pour lesquelles l'Algorithme 1 converge et montrer que
$$\|x^{(k)}-x^\star\| \leqslant \max(|1-t\lambda_1|, |1-t\lambda_n|)^k \|x^{(0)}-x^\star\|.$$
7. Tracer l'allure des graphiques des fonctions $t \mapsto |1-t\lambda_1|$ et $t \mapsto |1-t\lambda_n|$. Pour quelle valeur de t le membre de droite de l'inégalité précédente tend-il le plus vite vers 0 ?

Partie III : Algorithme de gradient à pas fixe avec moment

On considère à nouveau une fonction $f : \mathbb{R}^n \to \mathbb{R}$ générique. L'algorithme de gradient à pas fixe $t > 0$ avec moment $\beta > 0$ et point initial $x^{(0)} \in \mathbb{R}^n$ consiste à construire la séquence de vecteurs $(x^{(k)})_{k\in\mathbb{N}}$ telle que
$$\begin{aligned} x^{(1)} &= x^{(0)} - t\nabla f(x^{(0)}), \\ \forall k \geqslant 1, \quad x^{(k+1)} &= x^{(k)} - t\nabla f(x^{(k)}) + \beta(x^{(k)} - x^{(k-1)}). \end{aligned}$$
En pratique l'algorithme est arrêté avec le même critère d'arrêt que l'Algorithme 1.

8. ***(sur 4 points)*** On suppose que les fonctions

   `function fx = f(x)` et `function gfx = gradf(x)`

   sont définies dans scilab. Ecrire une fonction scilab

   ```
   function [Xk,FXk] = algo_moment(x0, eps, t, bet, f, gradf)
   ```

   qui prend en entrées un point initial $x^{(0)} \in \mathbb{R}^n$, un seuil de tolérance $\varepsilon > 0$, et deux paramètres $t > 0$ et $\beta > 0$, les fonctions pour évaluer f et ∇f. La fonction renvoie la matrice `Xk` (de taille n par nombre d'itérations) contenant la liste des vecteurs $x^{(k)}$ donnée par l'algorithme de gradient à pas fixe $t > 0$ avec moment $\beta > 0$ ainsi que la liste des valeurs `Fxk` de la fonctionnelle f aux points $x^{(k)}$. On veillera à ajouter une condition pour limiter le nombre d'itérations à 10 000.

On considère de nouveau la fonction quadratique $g(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + c$. Soit $(x^{(k)})$ la suite de vecteurs obtenue par l'algorithme de gradient à pas fixe $t > 0$ avec moment $\beta > 0$ appliqué à g. On introduit la suite de vecteurs
$$y^{(k)} = \begin{pmatrix} x^{(k+1)} \\ x^{(k)} \end{pmatrix} \in \mathbb{R}^{2n}$$
et on pose également $y^\star = \begin{pmatrix} x^\star \\ x^\star \end{pmatrix} \in \mathbb{R}^{2n}$.

9. Montrer que la suite $y^{(k)} - y^\star$ vérifie une équation de récurrence de la forme
$$y^{(k+1)} - y^\star = T(y^{(k)} - y^\star)$$
(bien préciser l'expression de T).
10. Que faut-il faire pour établir la convergence de l'algorithme de gradient à pas fixe $t > 0$ avec moment $\beta > 0$ appliqué à g ? Peut-on faire le même raisonnement qu'à la question 4 ?