Optimisation avancée
2018-2019. Exercices d'entraînement (Chapître 4: Méthodes géométriques)

Exercice 1

Montrer que pour tout $\alpha, L > 0$, il n'existe pas de fonction à la fois α-fortement convexe et L-Lipschitzienne sur $\mathbb{R}^d$.

Exercice 2 Résolution exacte de systèmes linéaires: Méthode du gradient conjugué

Soit $A \in \mathbb{R}^{d\times d}$ une matrice symétrique définie positive et $b \in \mathbb{R}^d$. On cherche à résoudre le systéme linéaire $Ax = b$.

1. Pour $x, y \in \mathbb{R}^d$, on note $\langle x, y\rangle_A = x^\top Ay$ et $\|x\|_A = (x^\top Ax)^{1/2}$. Montrer que $\langle\cdot,\cdot\rangle_A$ définit un produit scalaire. Dans la suite, on considère une base $(p_0, \ldots, p_{d-1})$ orthogonale pour le produit scalaire $\langle\cdot,\cdot\rangle_A$. Soit $x_0 \in \mathbb{R}^d$ arbitraire et, pour $t = 1, \ldots, d$, soit $x_t \in \underset{x\in\{x_{t-1}+\lambda p_{t-1}:\lambda\in\mathbb{R}\}}{\operatorname{argmin}} f(x)$. Le but du problème est de montrer que $\hat{x} := x_d$ est égal à la solution x^* du système $Ax = b$.
2. Montrer que pour tout $t = 1, \ldots, d$, $x_t = x_{t-1} - \nabla f(x_{t-1})^\top \dfrac{p_{t-1}}{\|p_{t-1}\|_A^2}$.
3. Montrer que pour tout $t = 1, \ldots, d$ et tout $i = 0, \ldots, t-1$, $\nabla f(x_t)^\top p_i = 0$.
4. En utilisant la relation $Ax^* = b$, en déduire que $\langle\hat{x}, p_i\rangle_A = \langle x^*, p_i\rangle_A$, pour tout $i = 0, \ldots, d-1$, puis que $\hat{x} = x^*$.
5. En pratique, on n'a pas nécessairement accès à une famille $(p_0, \ldots, p_{d-1})$ orthogonale pour $\langle\cdot,\cdot,\rangle_A$. Montrer qu'on en obtient bien une en définissant $p_0 = \nabla f(x_0)$ et, pour $t = 1, \ldots, d-1$, $p_t = \begin{cases}\nabla f(x_t) - \langle\nabla f(x_t), p_{t-1}\rangle_A \frac{p_{t-1}}{\|p_{t-1}\|_A^2} \text{ si } p_{t-1} \neq 0\\ 0 \text{ sinon.}\end{cases}$
6. On veut conclure qu'en définissant, par récurrence: $x_0 \in \mathbb{R}^d, p_0 = \nabla f(x_0)$ et, pour $t = 1, \ldots, d-1$:

$$x_t = \begin{cases} x_{t-1} - \nabla f(x_{t-1})^\top \frac{p_{t-1}}{\|p_{t-1}\|_A^2} \text{ si } p_t \neq 0\\ x_{t-1} \text{ sinon}\end{cases}$$

et

$$p_t = \begin{cases}\nabla f(x_t) - \langle\nabla f(x_t), p_{t-1}\rangle_A \frac{p_{t-1}}{\|p_{t-1}\|_A^2} \text{ si } p_{t-1} \neq 0\\ \nabla f(x_t) \text{ sinon}\end{cases}$$

on obtient nécessairement $x_d = x^*$.

a) Montrer que si $p_t \neq 0$ pour tout $t \leq d-1$, on peut conclure directement en utilisant les questions précédentes.

b) Sinon, soit $t = \min\{s \geq 0 : p_s = 0\}$. Montrer que $\nabla f(x_t)$ est nécessairement une combinaison linéaire de $\nabla f(x_0), \ldots, \nabla f(x_{t-1})$.

c) Montrer que les familles $(p_0, \ldots, p_{t-1})$ et $(\nabla f(x_0), \ldots, \nabla f(x_{t-1}))$ engendrent le même sous-espace vectoriel E_t de $\mathbb{R}^d$.

d) Montrer que x_t minimise f sur E_t et en déduire que $\nabla f(x_t)^\top z = 0$ pour tout $z \in E_t$.

e) Conclure que $x_t = x^*$ et que pour tout $s \geq t, x_s = x_t = x^*$, et donc, $\hat{x} = x^*$.

7. Sans donner de justification, comment généraliseriez-vous l'algorithme précédent au cas où f n'est pas nécessairement une fonction quadratique ?