Université Claude Bernard - Lyon 1
M2 Statistique, Modélisation et Science des données
5 décembre 2022

Optimisation Convexe : Algorithmes et Applications en Apprentissage

Contrôle terminal. Durée : 2h. Tous les documents sont autorisés.
Dans tout le sujet ci-dessous, $||x||_p$ désigne la norme ℓ_p : $||x||_p := (\sum_i |x_i|^p)^{1/p}$ si $p < \infty$, $||x||_\infty := \max_i |x_i|$.

Exercice 1 (6 points)**.** Soit $L : \mathbb{R}^n \to \mathbb{R}$ la fonction définie par $L(x) := \sqrt{1 + ||x||_2^4}$. Prouver que L est convexe, que ses dérivées secondes sont bornées, mais qu'elle n'est pas elliptique. Calculer en particulier la Hessienne $D^2L(0)$. Quelle est la vitesse de convergence de l'algorithme de gradient à pas fixe appliqué à L ?

Exercice 2 (6 points)**.** Considérer la fonction $f : \mathbb{R}^{d+1} \to \mathbb{R} \cup \{+\infty\}$ donnée par

$$f(x,t) := \begin{cases} \frac{||x||_2^2}{2t} & \text{si } t > 0; \\ 0 & \text{si } t = 0, x = 0; \\ +\infty & \text{si } t = 0, x \neq 0; \\ +\infty & \text{si } t < 0; \end{cases}$$

où les points de $\mathbb{R}^{d+1}$ sont écrits sous la forme (x,t), avec $x \in \mathbb{R}^d$ et $t \in \mathbb{R}$.

Prouver que f est convexe et calculer son sous-différentiel en tout point.

Exercice 3 (6 points)**.** Soit $h : \mathbb{R}^n \to \mathbb{R}$ la fonction définie par $h(x) := \sqrt{1 + ||x||_2^2}$. Étant donnée une matrice $A \in \mathbb{R}^{n\times n}$ symétrique et définie positive, un vecteur $c = (c_1, \dots, c_n) \in \mathbb{R}^n$ et un nombre $\tau > 0$, considérer la suite $(x[k])_k$ de points de $\mathbb{R}^n$ définie comme suit : $x[0] = 0$ et, pour tout $k \geq 0$ et tout index $i = 1, \dots, n$, on définit

$$x[k+1]_i := \min\left\{ i, x[k]_i - \tau\left((Ax[k])_i + c_i + \frac{x[k]_i}{h(x[k])} \right) \right\}.$$

Dans l'expression ci-dessus les indices i représentent les composantes de chaque vecteur (et donc $(Ax)_i := \sum_j A^{ij} x_j$.). Prouver que si τ est suffisamment petit ($\tau \leq \tau_0$ où τ_0 dépend de A seulement) cette suite converge (s'agiti-il d'une suite obtenue par un algorithme d'optimisation ? lequel ?), donner une estimation de τ_0 en termes des valeurs propres de A, et estimer également la vitesse de convergence. Si on appelle $\bar{x}$ la limite de cettte suite, prouver que l'on a

$$\frac{1}{2}(A\bar{x}) \cdot \bar{x} + h(\bar{x}) \leq 1 + c \cdot \bar{x}.$$

Exercice 4 (12 points)**.** Étant donnés une matrice $A \in \mathbb{R}^{m\times n}$, un vecteur $c \in \mathbb{R}^n$, un vecteur $b \in \mathbb{R}^m$ et un nombre $\delta \geq 0$, considérer le problème de minimisation

$$\inf\left\{ \frac{1}{2}||Ax - b||_2^2 + c \cdot x + \delta||x||_1 \ : \ x \in \mathbb{R}^n \right\}.$$

1. Si $\delta = 0$, prouver que la valeur de cette borne inférieure est finie si et seulement si c est orthogonal au noyau de A, c'est-à-dire s'il existe $y \in \mathbb{R}^m$ tel que $c = A^t y$. Prouver dans ce cas que le minimum est alors atteint. Dans quel cas le point de minimum est-il unique ?
2. Si $\delta > 0$, prouver que la valeur de cette borne inférieure est finie si c s'écrit sous la forme $c = A^t y + \tilde{c}$, pour $y \in \mathbb{R}^m$ et $\tilde{c} \in \mathbb{R}^n$ satisfaisant $||\tilde{c}||_\infty \leq \delta$. Prouver également que le minimum est atteint si en plus l'inégalité $||\tilde{c}||_\infty < \delta$ est stricte.
3. Calculer les transformées de Legendre-Fenchel des fonctions $g : \mathbb{R}^m \to \mathbb{R}$ et $h : \mathbb{R}^n \to \mathbb{R}$ données par $g(z) := \frac{1}{2}||z - b||^2$ et $h(x) := \delta||x||_1 + c \cdot x$, et écrire le problème dual du problème de minimisation considéré.
4. Suggérer un algorithme pour résoudre ce problème, en décrivant explicitement ses différentes étapes.