Optimisation Convexe : Algorithmes et Applications en Apprentissage

Exercice 1.

1. Prouver que si $f : \mathbb{R}^n \to \mathbb{R}_+$ est une fonction convexe et $g : \mathbb{R}_+ \to \mathbb{R}$ est une fonction convexe croissante, alors $g \circ f$ est convexe.
2. Prouver que la fonction $h : \mathbb{R}^3 \to \mathbb{R}$ ci-dessous est convexe. Est-elle strictement convexe ?

$$h(x,y,z) := \frac{x^2+y^2+z^2}{2} + \sqrt{x^2+y^2+z^2} - \sin(\sqrt{x^2+y^2+z^2})$$

3. Déterminer le sous-différentiel ∂h en tout point de $\mathbb{R}^3$. En quels points h est-elle différentiable ?
4. Déterminer également le sous-différentiel de la fonction $\tilde{h}$ définie par $\tilde{h}(x,y,z) := h(x,y,z) + |x|$.

Exercice 2. Considérer les ensembles

$$A = \{(x,y) \in \mathbb{R}^2 \,:\, |x+y| \leq 1\} \quad \text{et} \quad B = \overline{B(0,1)} = \{(x,y) \in \mathbb{R}^2 \,:\, x^2+y^2 \leq 1\}.$$

1. Dessiner A, B et $A \cap B$. A et B sont-ils convexes ?
2. Écrire une formule pour la projection sur $A \cap B$, du type $P_{A\cap B}(x,y) = \ldots$, en distinguant éventuellement des cas (il suffit de la justifier avec un dessin).
3. Lesquelles des relations suivantes sont-elles vraies ?

$$P_{A\cap B} = P_A \circ P_B, \quad P_A \circ P_B = P_B \circ P_A, \quad P_{A\cap B} = P_{A\cap B} \circ P_B.$$

Exercice 3. Considérer la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ donnée par

$$f(x,y) := x^2 - 2x + |x-1|^3 + \frac{y^4}{4} + \frac{y^3}{3} + \frac{y^2}{2} + y.$$

1. Prouver que f est une fonction convexe. Écrire sa matrice Hessienne. Dire si f est elliptique.
2. Dire si ∇f est Lipschitzien sur $B_2 = \{(x,y) \in \mathbb{R}^2 \,:\, x^2+y^2 \leq 4\}$ et donner une estimation (même grossière) de sa constante de Lipschitz M.
3. Déterminer $\min\{f(x,y) \,:\, (x,y) \in \mathbb{R}^2\}$.
4. Prouver que $\min\{f(x,y) \,:\, (x,y) \in B_1\}$ existe, où $B_1 = \{(x,y) \in \mathbb{R}^2 \,:\, x^2+y^2 \leq 1\}$.
5. Suggérer un algorithme pour approcher la solution du problème de minimisation de f sur B_1, en donner sa description explicite (des formules explicites pour x_{k+1} et y_{k+1} et, s'il y a des paramètres à choisir, en donner une valeur admissible) et justifier sa convergence.

Exercice 4. Donner une formule pour l'opérateur proximal de la fonction $g : \mathbb{R} \to \mathbb{R}$ définie par $g(x) = \frac{2}{3}|x|^{3/2}$ ainsi que pour la fonction $G : \mathbb{R}^n \to \mathbb{R}$ définie par $G(x) = \frac{2}{3}||x||_{3/2}^{3/2}$.

Exercice 5. Soit $P_{\tau,g}$ l'opérateur proximal associé à une fonction convexe $g : \mathbb{R}^n \to \mathbb{R}$ telle que g et ∇g sont Lipschitziens. Prouver

$$|P_{\tau,g}[y] - (y - \tau\nabla g(y))| \leq C\tau^2.$$

Si $g \in C^2$ avec dérivées secondes Lipschitziennes, prouver

$$|P_{\tau,g}[y] - (y - \tau\nabla g(y) + \tau^2 D^2 g(y)\nabla g(y))| \leq C\tau^3.$$

Exercice 6 (Algorithme du gradient à pas optimal)**.** Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction convexe C^1, coercive, et strictement convexe. On définit une suite $(x_k)_k$ comme suit

$$x_{k+1} = \operatorname{argmin}\{f(x) \,:\, x \in x_k + \mathbb{R}\nabla f(x_k)\}.$$

1. Prouver que le point x_{k+1} est bien défini et il est de la forme $x_k - t\nabla f(x_k)$ pour $t \geq 0$.
2. Prouver que la suite satisfait $\nabla f(x_{k+1}) \cdot \nabla f(x_k) = 0$ pour tout k ainsi que $f(x_{k+1}) \leq f(x_k)$.
3. Prouver $|\nabla f(x_{k+1})| \leq |\nabla f(x_{k+1}) - \nabla f(x_k)|$.

4. Si f est elliptique ($D^2f \geq \alpha I$) prouver que l'on a $f(x_k) \geq f(x_{k+1}) + \frac{\alpha}{2}|x_{k+1} - x_k|^2$. En déduire $\lim_k |x_{k+1} - x_k| = 0$.
5. Si ∇f est $L-$Lipschitzien prouver $|\nabla f(x_{k+1})| \leq L|x_{k+1} - x_k|$.
6. Conclure que, si f est elliptique et ∇f est $L-$Lipschitzien, la suite converge vers l'unique point de minimum de f. Cela peut-il s'étendre au cas f elliptique et C^1 (sans supposer $\nabla f \in \mathrm{Lip}$) ?

Exercice 7. Soit $f : \mathbb{R} \to \mathbb{R}$ la fonction définie par $f(x) = e^x$. Trouver f^*.

Exercice 8. Étant donné une fonction $f : \mathbb{R}^n \to \mathbb{R}$ et un vecteur $v \in \mathbb{R}^n$ soit $g : \mathbb{R}^n \to \mathbb{R}$ définie par $g(x) = f(x) - v \cdot x$. Calculer g^* en termes de f^* et v.

Exercice 9. Considérer le problème d'optimisation suivant : étant donnés N, M deux entiers naturels, des valeurs $c_{ij} \in \mathbb{R}$ ainsi que des nombres $\mu_i > 0$ et $\nu_j > 0$ pour $i = 1, \dots, N$ et $j = 1, \dots, M$, trouver la matrice $\gamma = (\gamma_{ij})_{ij}$ avec cordonnées non-négatives, satisfaisant $\sum_j \gamma_{ij} = \mu_i$ et $\sum_i \gamma_{ij} = \nu_j$ pour tout i, j et minimisant $\sum_{ij} \gamma_{ij} c_{ij}$. Écrire son problème dual.

Exercice 10. Trouver le problème dual du problème de minisation suivant

$$\min \left\{ \frac{1}{p} ||x||^p - c \cdot x \ : \ x \in \mathbb{R}^n, \, a \cdot x = t \right\},$$

où $p > 1$ est un exposant donné, a et c sont deux vecteurs fixés de $\mathbb{R}^n$, t est un réel donné, et $||x||$ représente la norme éuclidienne du vecteur x.

Exercice 11. Considérons l'ensemble $K = \{(x_1, \dots, x_n) \in \mathbb{R}^n \ : \ |x_1|^3 + \dots + |x_n|^3 \leq 1\}$ et $p = (p_1, \dots, p_n) \notin K$ un point à l'extérieur de K. Décrire de manière détaillée et explicite au moins une méthode numérique pour calculer la projection de p sur K et justifier sa convergence.

Cette projection pourrait se configurer comme un problème d'optimisation d'une fonction convexe (la distance à p, ou le carré de cette distance) sous contrainte convexe (K étant convexe). Pourquoi ne serait-il pas raisonnable *du tout* de considérer un algorithme de gradient projeté pour répondre à la question précédente ?

Exercice 12. Choisir un algorithme d'optimisation adapté pour chacun des cas suivants et discuter sa vitesse de convergence

1. $\min\{\sqrt{1 + ||x - a||_2^4} + ||x||_1, \ x \in \mathbb{R}^n\}$
2. $\min\{\sqrt{1 + ||x - a||_2^2} + ||x - b||_2^2 + ||x||_1, \ x \in \mathbb{R}^n\}$
3. $\min\{\sqrt{1 + ||x - x_0||_2^2} + ||x||_1, \ x \in \mathbb{R}^n, |x_i| \leq i \text{ pour } i = 1, \dots, n\}$
4. $\min\{\sqrt{1 + e^{\sum_i x_i}} + x \cdot a, \ x \in \mathbb{R}^n, ||x||_1 \leq 1\}$

Exercice 13. Étant donnée une fonction $f : \mathbb{R}^n \to \mathbb{R}$ $\alpha-$elliptique avec $\alpha = 1$ et telle que ∇f est Lipschitzien, considérer un algorithme de gradient stochastique

$$x_{k+1} = x_k - \eta_k v_k$$

pour une suite de variables aléatoires v_k telles que $\mathbb{E}[v_k | x_k] = \nabla f(x_k)$ et $\mathbb{E}[|v_k|^2] \leq M$. En choisissant $\eta_k = \frac{1}{k+1}$ et en appelant $\bar{x}$ le point de minimum de f, prouver que l'on a

$$\mathbb{E}[|x_k - \bar{x}|^2] \leq C \frac{\log(k+1)}{k+1} \to 0.$$

Exercice 14. Définissions une fonction $f : \mathbb{R}^n \to \mathbb{R}$ comme suit : on prend n dés à six faces et on appelle $d_i \in \{1, \dots, 6\}$ le résultat de chaque dé, et pour chaque $x = (x_1, \dots, x_n)$ on définit $f(x)$ comme la valeur attendue de $\Pi_{i=1}^n d_i^{x_i}$. Décrire en pratique comment mettre en oeuvre un algorithme de gradient stochastique pour minimiser f ou $f + g$, où g est une fonction de régularisation/pénalisation portant sur le vecteur x (par exemple $g(x) = ||x||_2^2$).