Optimisation et simulation numérique – Devoir Maison 1

à déposer dans le casier de F. Santambrogio avant le 23 octobre ; résoudre en détail 3 exercices, pas plus.

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 $h : \mathbb{R}^3 \to \mathbb{R}$ définie par
$$h(x, y, z) := \frac{x^2 + y^2 + z^2}{2} + \sqrt{x^2 + y^2 + z^2} - \sin(\sqrt{x^2 + y^2 + z^2})$$
est convexe. Est-elle strictement convexe ?
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 $\tilde{h}(x, y, z) := h(x, y, z) + |x|$.

Solution Si f est convexe on a $f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$ et, en composant avec g qui est convexe, on trouve $(g \circ f)((1-t)x + ty) \leq g((1-t)f(x) + tf(y))$. Mais g est également convexe, donc $g((1-t)f(x) + tf(y)) \leq (1-t)(g \circ f)(x) + t(g \circ f)(y)$, ce qui donne la convexité de $g \circ f$.

Pour prouver la convexité de h, on prend $f(x, y, z) = \sqrt{x^2 + y^2 + z^2}$ (qui est convexe) et $g(t) = t^2/2 + t - \sin t$. En calculant $g'(t) = t + 1 - \cos t \geq t \geq 0$ (sur $t \geq 0$, mais on l'applique seulement à $t = \sqrt{x^2 + y^2 + z^2} \geq 0$) et $g''(t) = 1 + \sin t \geq 0$ on trouve que g est convexe et croissante. Donc, h est convexe. Remarquons que g est strictement croissante ($g' > 0$ sauf en $t = 0$) et strictement convexe (g' est strictement croissante, parce que $g'' > 0$ sauf en $t = \pi/2 + 2k\pi$). Analysons les cas d'égalité dans les inégalités précédentes. Pour avoir

$$(g \circ f)((1-t)X + tX') = g((1-t)f(X) + tf(X')) = (1-t)(g \circ f)(X) + t(g \circ f)(X'),$$

(avex $x \neq y$ et $t \in]0, 1[$ et $X = (x, y, z)$, $X' = (x', y', z')$) il faudrait $f(X) = f(X')$ (dans la dernière inégalité). Mais dans ce cas, X et X' ont le mÎme module, et f est strictement convexe sur le segment $[X, X']$. Donc la toute première inégalité est stricte, et on a (g étant strictement croissante) $(g \circ f)((1-t)X + tX') < g((1-t)f(X) + tf(X'))$.

La fonction h est différentiable partout. En effet, $g \in C^\infty(\mathbb{R}_+)$ et $g(t) = O(t^2)$ en $t = 0$ (parce que $t - \sin t = t - t + O(t^3)$). La fonction f est C^∞ en dehors de l'origine, ce qui donne la différentiabilitÈ sauf en $X = 0$. Mais en ce point, f étant Lipschitz, on a $|g(f(X)) - g(f(0))| \leq C|f(X) - f(0)|^2 \leq C|X|^2$, ce qui montre que $g \circ f$ est différentiable avec gradient nul en $X = 0$. Le sous-différentiel est donc donné par
$$\partial h(X) = \left\{ \nabla h(X) = g'(|X|) \frac{X}{|X|} \right\}.$$

Sachant que le sous-différentiel de $X = (x, y, z) \mapsto |x|$ est donnÈ par $\{(1, 0, 0)\}$ si $x > 0$, par $\{(-1, 0, 0)\}$ si $x < 0$, et par $[-1, 1] \times \{0\} \times \{0\}$ si $x = 0$, on trouve

$$\partial h(x, y, z) = \begin{cases} \left\{ ((|X| + 1 - \cos(|X|)) \frac{x}{|X|} + 1, (|X| + 1 - \cos(|X|)) \frac{y}{|X|}, (|X| + 1 - \cos(|X|)) \frac{z}{|X|}) \right\} & \text{si } x > 0 \\ \left\{ ((|X| + 1 - \cos(|X|)) \frac{x}{|X|} - 1, (|X| + 1 - \cos(|X|)) \frac{y}{|X|}, (|X| + 1 - \cos(|X|)) \frac{z}{|X|}) \right\} & \text{si } x < 0 \\ \left\{ (t, (|X| + 1 - \cos(|X|)) \frac{y}{|X|}, (|X| + 1 - \cos(|X|)) \frac{z}{|X|}) \ , \ t \in [-1, 1] \right\} & \text{si } x = 0 \end{cases}$$

Exercice 2. 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 de sa convergence en s'appuyant sur les résultats vus en cours.

Exercice 3. 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 ?

Solution S'agissant d'une optimisation sous contrainte, un algorithme adapté est, par exemple, celui d'Uzawa, dont la convergence a été prouvée en classe. L'algorithme du gradient projeté par contre est complètement inadapté parce qu'il demanderait à utiliser la projection sur K, ce qui est exactement ce qu'on cherche à calculer !!

Pour minimiser $\frac{1}{2}|x-p|^2 : x \in K = \{x : g(x) \leq 1\}$, l'algorithme d'Uzawa utilise la dualité, et cherche à résoudre

$$\max_{\lambda \geq 0} \min_x \frac{1}{2}|x-p|^2 + \lambda(g(x)-1).$$

Si on appelle $F(\lambda) = \min_x \frac{1}{2}|x-p|^2 + \lambda(g(x)-1)$, l'algorithme d'Uzawa est un algorithme de gradient projeté pour la fonction concave F (qu'on cherche à maximiser), soumise à la contrainte $\lambda \geq 0$ (qui est simple à gérer). On sait que $F'(\lambda) = g(x_\lambda)-1$ où x_λ est le x optimal correspondant à λ, celui qui minimise $\frac{1}{2}|x-p|^2+\lambda(g(x)-1)$. Ce x_λ peut être trouvé en imposant $x-p+\lambda\nabla g(x) = 0$. La suite des λ produite par l'algorithme doit satisfaire $\lambda_{k+1} = (\lambda_k + \rho(g(x_{\lambda_k})-1))_+$, où $\rho > 0$ est le pas de l'algorithme de gradient projeté. Ce ρ doit être suffisamment petit, et éventuellement variable.

Explicitons le calcul de x_λ : pour chaque composante i, on doit imposer $x_i - p_i + \lambda\partial_{x_i}g(x) = 0$. Comme $\partial_{x_i}g(x) = \pm 3x_i^2$ (le signe étant celui de x_i, il est nécessaire que x_i ait le même signe de p_i. Donc pour $p_i > 0$ on cherche $x_i > 0$ tel que $x_i - p_i + \lambda x_i^2 = 0$ et pour $p_i < 0$ on cherche $x_i < 0$ tel que $x_i - p_i - \lambda x_i^2 = 0$, ce qui donne

$$x_i = \frac{-1+\sqrt{1+12\lambda p_i}}{6\lambda} \text{ si } p_i \geq 0, \quad x_i = \frac{1-\sqrt{1-12\lambda p_i}}{6\lambda} \text{ si } p_i < 0.$$

L'algorithme itératif est donc obtenu en partant d'un couple $(x^{(0)}, \lambda_0)$ quelconque et prenant ensuite

$$x^{(k+1)} := \text{ donné par la formule précédente avec } \lambda = \lambda_k, \quad \lambda_{k+1} = \Big(\lambda_k + \rho(g(x^{(k+1)})-1)\Big)_+.$$

La suite $x^{(k)}$ si obtenue converge alors vers la projection de p sur K.

Exercice 4. Trouver la formulation duale de ce problème d'optimisation linéaire, et démontrer que tant le problème primal que le problème dual admettent une solution : étant donné $f \in \mathbb{R}^{N+1}$

$$\max\left\{\sum_{i=0}^N f_i x_i : x \in \mathbb{R}^{N+1}, |x_i| \leq 1 \text{ pour tout } i = 0,\dots,N, |x_i - x_{i-1}| \leq 1 \text{ pour tout } i = 1,\dots,N\right\}.$$

Exercice 5. Trouver un exemple de fonction convexe $f : \mathbb{R} \to \mathbb{R}$ bornée inférieurement telle que l'algorithme du gradient à pas constant donne une estimation exactement d'ordre $1/k$, i.e. $f(x_k) - \inf f = O(1/k)$, quelque soit le pas choisi. Le minimum est-il atteint ?

Solution On peut prendre $f(x) = e^x$. L'algorithme du gradient nous donne $x_{k+1} = x_k - \tau e^{x_k}$. Soit $y_k = e^{x_k}$. On veut montrer $y_k = O(1/k)$. Il est clair qu'on a $x_k \to -\infty$ (x_k est décroissante, et si on avait $x_k \to \ell$ on devrait avoir $\ell = \ell - \tau e^\ell$, d'où $\ell = -\infty$). Donc $y_k \to 0$ En utilisant

$$y_{k+1} = e^{x_k - \tau e^{x_k}} = \frac{y_k}{e^{\tau y_k}}$$

on peut obtenir

$$\frac{y_k}{1+2\tau y_k} \leq y_{k+1} \leq \frac{y_k}{1+\tau y_k}$$

(on a utilisé $y_k \geq 0$, $e^y \geq 1+y$, qui est toujours vraie, et $e^y \leq 1+2y$, qui est vraie pour y petit et positif). Ceci nous donne (en passant aux réciproques)

$$\frac{1}{y_k} + 2\tau \geq \frac{1}{y_k} \geq \frac{1}{y_k} + \tau,$$

ce qui montre que $1/y_k$ croit à chaque étape de quelque chose de l'ordre de τ, et que donc $1/y_k = O(k)$.

Exercice 6. Donner un algorithme de type ISTA pour résoudre

$$\min \left\{ ||x - x_0||_2^2 + \lambda ||x||_p^p \ : \ x \in \mathbb{R}^n \right\}$$

en décrivant en détail les étapes ($x_0 \in \mathbb{R}^n$, $\lambda > 0$ et $p \geq 1$ sont fixés). Pour quelles valeurs de p vaut-il mieux utiliser un tel algorithme au lieu d'un algorithme de gradient à pas fixe ? discuter aussi les liens entre ce problème et celui de l'exercice 3.