Université Claude Bernard - Lyon 1
M2 Statistique, Modélisation et Science des données
8 décembre 2025

Optimisation Convexe : Algorithmes et Applications en Apprentissage

Contrôle terminal. Durée : 2h. Tous les documents sont autorisés, mais pas les objets connectés et les moyens de communication. Le barème dépasse largement 20, il est conseillé de ne pas traiter tous les exercices.
Dans tout le sujet ci-dessous, $||x||$ désigne la norme euclidienne : $||x|| := (\sum_i |x_i|^2)^{1/2}$.

Exercice 1 (6 points)**.** Soit $f : \mathbb{R}^3 \to \mathbb{R}$ la fonction définie par

$$f(x_1, x_2, x_3) := \frac{1}{4}x_1^4 + \frac{1}{4}(x_2^2 + x_3^2)^2$$

Calculer f^*. Est-ce que f et/ou f^* sont elliptiques ?

Solution : On a

$$f^*(y_1, y_2, y_3) = \sup_{x_1,x_2,x_3} y_1x_1 - \frac{1}{4}x_1^4 + (y_2x_2 + y_3x_3) - \frac{1}{4}(x_2^2 + x_3^2)^2.$$

On remarque qu'il y a deux parties indépendantes. La première correpond à calculer la transformé de la fonction $x \mapsto \frac{1}{4}x^4$ en dimension 1. Le calcul a déjà été fait en classe : on sait que pour $p > 1$ la transformé de la fonction $x \mapsto \frac{1}{p}|x|^p$ est $y \mapsto \frac{1}{q}|y|^q$ où l'exposant q est caractérisé par $\frac{1}{p} + \frac{1}{q} = 1$. Ici on trouve donc $q = 4/3$ et donc le premier sup donne $\frac{3}{4}|y_1|^{4/3}$.

La deuxième partie peut être exprimée, dans $\mathbb{R}^2$, comme la transformé de $z \mapsto \frac{1}{4}||z||^4$, à calculer en $y' = (y_2, y_3)$. Dans la maximisation de $z \mapsto y' \cdot z - \frac{1}{4}||z||^4$ en dérivant on trouve $y' = ||z||^2 z$ donc y et z sont alignés. Si $y' \neq 0$ en écrivant $z = ty/||y||$ on retrouve la maximisation de $t||y'|| - \frac{1}{4}t^4$, ce qui nous permet d'utiliser le résultat 1D précedent. On trouve donc $\frac{3}{4}||y'||^{4/3}$, e la même formule vaut pour $y' = 0$.

Le résultat est donc $f^*(y) = \frac{3}{4}|y_1|^{4/3} + \frac{3}{4}(y_2^2 + y_3^2)^{2/3}$.

La fonction f n'est pas elliptique parce qu'on a $f(x) = O(||x||^4)$ près de $x = 0$. La fonction f^* n'est pas elliptique parce qu'on a $f(y) \leq C||y||^{4/3} \ll ||y||^2$ pour $||y|| \to \infty$.

Exercice 2 (6 points)**.** Étant donné une matrice $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ définie positive, un point "cible" $(\bar{x}, \bar{y}) \in \mathbb{R}^2$ et un point "départ" $(x_0, y_0) \in \mathbb{R}^2$, ainsi qu'un nombre $\tau > 0$, considérer la suite (x_k, y_k) définie par recurrence comme suit

$$\begin{aligned} x_{k+1} &= \min\{1, \max\{-1, x_k - \tau(a(x_k - \bar{x}) + b(y_k - \bar{y}))\}\}, & (1) \\ y_{k+1} &= (|y_k - \tau(b(x_k - \bar{x}) + c(y_k - \bar{y}))| - \tau)_+ \text{signe}(y_k - \tau(b(x_k - \bar{x}) + c(y_k - \bar{y})). & (2) \end{aligned}$$

1. S'agit-il de la suite obtenue en appliquant un algorithme connu pour résoudre un problème d'optimisation ? lequel ?
2. Prouver que si τ est suffisamment petit alors la suite $(x_k, y_k)_k$ converge vers un point (x_∞, y_∞) et le caractériser comme solution d'un problème d'optimisation.
3. Donner une estimation précise de la vitesse de convergence de la suite $(x_k, y_k)_k$.

Solution :

1. Il s'agit-il de la suite obtenue en appliquant le gradient proximal pour la fonction $F = f + g$ où

$$f(x, y) = \frac{1}{2}a(x - \bar{x})^2 + b(x - \bar{x})(y - \bar{y}) + \frac{1}{2}c(y - \bar{y})^2, \qquad g(x, y) = \begin{cases} |y| & \text{si } |x| \leq 1, \\ +\infty & \text{sinon.} \end{cases}.$$

2. La fonction f étant quadratique, son gradient est Lipschitzien d'une constante L égale à la plus grande valeur propre de la matrice A, donc l'algorithme converge dès que $\tau < \frac{1}{2L}$ vers le point (x_∞, y_∞) qui est l'unique minimseur de F (unique parce que f est strictement convexe).
3. La suite convege expenentiellement : $||(x_k, y_k)-(x_\infty, y_\infty)|| \leq ||(x_0, y_0)-(x_\infty, y_\infty)||\lambda^k$ où $\lambda = \min\{\tau L - 1, 1-\alpha\tau\}$, où $\alpha > 0$ est la plus petite valeur propre de la matrice A.

Exercice 3 (6 points)**.** Étant donnés une matrice $A \in \mathbb{R}^{m\times n}$ et un vecteur $y \in \mathbb{R}^m$ on considère l'ensemble $K = \{x \in \mathbb{R}^n \,:\, x \geq 0, Ax = b\}$, où l'inégalité $x \geq 0$ est à prendre composant par composante. On suppose par la suite que K est non-vide. Étant donnéun point $y \in \mathbb{R}^n$ on considère

$$\min\{||x-y|| \,:\, x \in K\}.$$

1. Prouver que ce problème admet une unique solution.
2. Décrire comment approcher la solution du problème par l'algorihme d'Uzawa en donnant les étapes de manière explicite.

Solution :

1. L'ensemble K est onvexe et fermé et il s'agit de trouver la projection de y sur K. Cette projection existe et est unique.
2. Au lieu de minimiser $||x-y||$ on minimise $\frac{1}{2}||x-y||^2$, ce qui est équivalent. On considère

$$\min_{x\geq 0} \frac{1}{2}||x-y||^2 + \sup_{z\in\mathbb{R}^m} z\cdot(Ax-b)$$

et on définit $G(z) = \inf_{x\geq 0} 12||x-y||^2 + z\cdot(Ax-b)$. On a $\nabla G(z) = Ax(z) - b$ où $x(z)$ est la solution de ce problème d'optimisation définissant $G(z)$. Ce problème peut se réécrire comme $\min_{x\geq 0} \frac{1}{2}||x-(y-A^t z)||^2 + C$ où la constant C dépend de z, b et A mais pas de x. Le point $x(z)$ est donc la projection de $y - A^t z$ sur l'ortant positif, donc $x(z) = (y - A^t z)_+$ où la partie positive est à prendre composante par composante. L'algorihme d'Uzawa correspond donc dans ce cas à

$$x_k = (y - A^t z_k)_+, \quad z_{k+1} = z_k + \tau(Ax_k - b).$$

Exercice 4 (9 points)**.** 1. Prouver que pour tout $p > 1$ la fonction $h_p : \mathbb{R}_+ \to \mathbb{R}$ donnée par $h_p(s) = (1+s^p)^{1/p}$ est convexe et Lipschitzienne de constante 1.

2. On définit ensuite la fonction $f : \mathbb{R}^n \to \mathbb{R}$ par

$$f(x_1, \dots, x_n) = \sum_{i=1}^n h_{p_i}\left(\sqrt{x_i^2 + x_{i+1}^2}\right),$$

où on pose $x_{n+1} = x_1$ et les exposants p_i sont des exposants strictement plus grands que 1 et fixés. Prouver que l'on a $f(x) \geq \sum_{i=1}^n |x_i|$.

3. Étant donné un vecteur $v \in \mathbb{R}^n$ avec des composantes $(v_i)_i$ telles que $|v_i| \leq 1$ pour tout i, on considère le problème

$$(P) \qquad \min\{f(x) + v\cdot x \,:\, x \in \mathbb{R}^n\}.$$

Prouver que ce problème admet une unique solution.

4. Décrire comment approcher la solution du problème par l'algorithme de gradient stochastique et quel type de convergence on obtient.
5. Peut-on améliorer la condition $f(x) \geq \sum_{i=1}^n |x_i|$ en obtenant en fait $f(x) \geq \sqrt{2}\sum_{i=1}^n |x_i|$? Que se passe -t-il dans le problème (P) si les composantes v_i de v satisfont $|v_i| \leq \sqrt{2}$?

Solution :

1. On calcule la dérivée de h_p : on a $h_p'(s) = (1+s^p)^{\frac{1}{p}-1}s^{p-1}$, donc $|h_p'(s)| \leq (s^p)^{\frac{1}{p}-1}s^{p-1} = 1$ donc h_p est Lipschitzienne de constante 1. Quant à sa convexité, on calcule h_p'' et on obtient,
$$h_p''(s) = (p-1)(1+s^p)^{\frac{1}{p}-1}s^{p-2} + (\frac{1}{p}-1)(1+s^p)^{\frac{1}{p}-2}ps^{p-1}s^{p-1}$$
$$= (p-1)(1+s^p)^{\frac{1}{p}-2}s^{p-2}(1+s^p-s^p) = (p-1)(1+s^p)^{\frac{1}{p}-2}s^{p-2} \geq 0,$$
donc h_p est convexe. On observe aussi que $h_p''(s)$ ne s'annule qu'en $s=0$, donc h est strictement convexe.
2. On observe que l'on a $h_p(s) \geq s$ et $\sqrt{x_i^2+x_{i+1}^2} \geq |x_i|$, donc on a
$$f(x_1,\dots,x_n) = \sum_{i=1}^n h_{p_i}\left(\sqrt{x_i^2+x_{i+1}^2}\right) \geq \sum_{i=1}^n |x_i|.$$
3. Supposons $\max_i |v_i| = c < 1$. Alors on a $f(x)+v\cdot x \geq (1-c)\sum_{i=1}^n |x_i|$, ce qui montre que la fonction à minimiser est coercive. Elle est aussi continue (parce que somme de fonctions Lipschitziennes) et admet donc un minimiseur. Ce minimiseur est unique à cause de la stricte convexité des h_p. En effet, les fonctions h_p sont strictement convexes et strictement croissantes sur $\mathbb{R}_+$, donc $x \mapsto h_p(||x||)$ est aussi convexe sur $\mathbb{R}^2$ (par composition avec la norme, qui est convexe). Cette fonction est aussi strictement convexe, parce que le seul cas d'égalité dans l'inégalité de convexité sur un segment $[a,b]$ demanderai en même temps l'égalité de la norme $||a|| = |lb||$ (parce que h_p est strictement convexe) et $a//b$ (parce que sinon on aurait une inégalité stricte dans l'inégalité de convexité de la norme), donc $a=b$. On déduit que deux éventuels minimiseurs auraient toutes les composantes qui coïncident. Reste à voir le cas $c=1$, qui est en fait inclus dans la dernière question.
4. On commence par $X_0 = 0$. Étant donnée une variable aléatoire X_k on tire au hasard, de manière indépendante de tous les tirages précédents, un indice $i \in \{1,\dots,n\}$. On update ensuite $X_{k+1} = X_k - \tau_k \nabla f_i(X_k)$ où f_i est la fonction $x \mapsto h_{p_i}(\sqrt{x_i^2+x_{i+1}^2}) + v_i x_i$, donc
$$X_{k+1,j} = \begin{cases} X_{k,j} & \text{si } j \neq i, i+1, \\ X_{k,j} - \tau_k(v_i + \frac{n_i^{p_i-2}}{(1+n_i^{p_i})^{1-1/p_i}}X_{k,j}) & \text{si } j = i \\ X_{k,j} - \tau_k\frac{n_i^{p_i-2}}{(1+n_i^{p_i})^{1-1/p_i}}X_{k,j} & \text{si} j = i+1 \end{cases} \quad \text{où } n_i = \sqrt{x_i^2+x_{i+1}^2},$$
pour une suite $\tau_k \to 0$ fixée, notamment $\tau_k = k^{-1/2}$. La suite $\tilde{X}_k$ définie par $\gamma_k := \sum_{j=0}^k \tau_j$ et $\tilde{X}_k := \frac{\sum_{j=0}^k \tau_j X_j}{\gamma_k}$ converge presque sûrement vers la seule solution du problème et satisfait
$$\mathbb{E}[f(\tilde{X}_k) - \min f] \leq \frac{C+\log k}{\sqrt{k}}.$$
5. Pour améliorer la condition $f(x) \geq \sum_{i=1}^n |x_i|$ en obtenant en fait $f(x) \geq \sqrt{2}\sum_{i=1}^n |x_i|$ iol faut remplacer l'inégalité $\sqrt{x_i^2+x_{i+1}^2} \geq |x_i|$ par $\sqrt{x_i^2+x_{i+1}^2} \geq \frac{1}{\sqrt{2}}(|x_i|+|x_{i+1}|$. Cela donne donc
$$f(x_1,\dots,x_n) = \geq \frac{1}{\sqrt{2}}\sum_{i=1}^n (|x_i|+|x_{i+1}| = \sqrt{2}\sum_{i=1}^n |x_i|.$$
On obtient donc la coercivité pour tout v tel que $|v_i| \leq c < \sqrt{2}$. Le cas $c=\sqrt{2}$ ne peut pas être traité de la même manière et en effet le minimum n'existe pas forcement, notamment si on prend $v = -\sqrt{2}e$ où $e = (1,1,\dots,1)$. Dans ce cas on a quand même $f(x)+v\cdot x \geq 0$ pour tout x mais on peut prendre $x = te$ et on a $\lim_{t\to\infty} f(te) + v\cdot(te) = 0$. Pour voir cela, il suffit de considérer un seul terme de la somme, donc $h_{p_i}(t\sqrt{2}) - t\sqrt{2}$ et, en posant $s = t\sqrt{2}$ on observe que l'on a
$$\lim_{s\to\infty}(1+s^p)^{1/p} - s = \lim_{s\to\infty} s\left[(s^{-p}+1)^{1/p}-1\right] = \lim_{s\to\infty} sO(s^{-p}) = 0.$$
On en déduit donc que l'infimum est 0 mais n'est pas atteint.