

CORRECTION

ANALYSE CONVEXE & OPTIMISATION

Durée approximative : 2h.*Seuls le polycopié et les notes de cours sont autorisés.**Il est vivement recommandé de lire tout l'énoncé au départ, car la majorité des questions sont dépendantes.**Les questions précédées d'une (*) sont plus difficiles (mais abordables). La question (***) est bien plus ouverte et complexe.***Régularisation de Moreau-Yosida**

La régularisation de Moreau-Yosida est une technique générique pour obtenir des approximations régulières de fonctions non lisses. Le but de ce problème est d'établir quelques propriétés de cette régularisation dans le cas convexe.

Soit

$$\Gamma_0(\mathbb{R}^n) = \{f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}, f \text{ est convexe fermée}\}.$$

Soit $f \in \Gamma_0(\mathbb{R}^n)$. On note pour $\epsilon \geq 0$:

$$f_\epsilon(x) = \min_{u \in \mathbb{R}^n} f(u) + \frac{1}{2\epsilon} \|x - u\|_2^2, \quad (1)$$

où $\|\cdot\|_2^2$ désigne la norme ℓ^2 usuelle sur \mathbb{R}^n . L'application $f \mapsto f_\epsilon$ est appelée régularisation de Moreau-Yosida.

Partie I (Questions de cours ~11pts)

1. Soit $g_\epsilon(u) = f(u) + \frac{\epsilon}{2} \|x - u\|_2^2$. La fonction g_ϵ est-elle différentiable pour tout $f \in \Gamma_0(\mathbb{R}^n)$? Est-elle fortement convexe pour tout $f \in \Gamma_0(\mathbb{R}^n)$?

La fonction g_ϵ n'est pas forcément différentiable. Par exemple pour $f(u) = |u|$, g_ϵ a une non différentiabilité en 0.

La fonction g_ϵ est toujours fortement convexe, car la somme d'une fonction convexe et d'une fonction fortement convexe est fortement convexe.

2. Le problème (??) est-il convexe? Admet-il une solution? Si oui, est-elle unique? Le problème est convexe et il admet une et une seule solution comme g_ϵ est fortement convexe.
3. Calculer la régularisée de Moreau-Yosida de la fonction $h(u) = |u|$, où $u \in \mathbb{R}$.

On a vu en TP et en TD que :

$$\begin{aligned} & \arg \min_{u \in \mathbb{R}^n} |u| + \frac{1}{2\epsilon} \|x - u\|_2^2 \\ &= \arg \min_{u \in \mathbb{R}^n} \epsilon |u| + \frac{1}{2} \|x - u\|_2^2 \\ &= \begin{cases} 0 & \text{si } |x| \leq \epsilon \\ x - \epsilon & \text{si } x > \epsilon \\ x + \epsilon & \text{si } x < -\epsilon. \end{cases} \end{aligned}$$

En remplaçant cette expression dans $g_\epsilon(u)$, on obtient :

$$h_\epsilon(x) = \begin{cases} \frac{1}{2\epsilon} x^2 & \text{si } |x| \leq \epsilon \\ |x| - \frac{\epsilon}{2} & \text{sinon} \end{cases}$$

4. Montrer que pour $h(u) = |u|$, la fonction h_ϵ est C^1 .

Clairement, h_ϵ est C^∞ sur $] -\epsilon, \epsilon[$, sur $] -\infty, -\epsilon[$ et sur $]\epsilon, \infty[$. Il faut donc analyser uniquement la régularité en ϵ et $-\epsilon$. On se contente du cas ϵ , l'autre étant identique.

On a $h_\epsilon(\epsilon^-) = h_\epsilon(\epsilon^+) = \frac{\epsilon}{2}$, la fonction est donc continue en ϵ .

On a :

$$h'_\epsilon(x) = \begin{cases} \frac{x}{\epsilon} & \text{si } |x| \leq \epsilon \\ 1 & \text{si } x > \epsilon \\ -1 & \text{si } x < -\epsilon. \end{cases}$$

Donc, $h'_\epsilon(\epsilon^-) = h'_\epsilon(\epsilon^+) = 1$, la fonction a donc bien une dérivée continue : elle est C^1 .

5. (*) Calculer la constante de Lipschitz de la dérivée de h_ϵ .

La constante de Lipschitz de h'_ϵ est donnée par :

$$L_\epsilon = \sup_{x_1, x_2 \in \mathbb{R}} \frac{|h'_\epsilon(x_1) - h'_\epsilon(x_2)|}{|x_1 - x_2|}.$$

Si $|x_1| \leq \epsilon$ et $|x_2| \leq \epsilon$, alors $\frac{|h'_\epsilon(x_1) - h'_\epsilon(x_2)|}{|x_1 - x_2|} = \frac{1}{\epsilon}$. Dans tous les autres cas, $\frac{|h'_\epsilon(x_1) - h'_\epsilon(x_2)|}{|x_1 - x_2|} \leq \frac{1}{\epsilon}$. On en déduit que $L_\epsilon = \frac{1}{\epsilon}$.

Note : une autre façon de calculer la constante de Lipschitz pour une fonction C^2 est d'évaluer $\max_{t \in \mathbb{R}} h''_\epsilon(t)$. Malheureusement, h_ϵ n'est pas C^2 ici aux points $\pm\epsilon$. Il faudrait donc être un peu prudent. Les gens ayant eu adopté cette approche auront tout de même les points de la question.

6. Calculer la régularisée de Moreau-Yosida de la fonction $f(u) = \|u\|_1$, où $u \in \mathbb{R}^n$ (vous pourrez utiliser les résultats sur h_ϵ directement).

On a

$$\begin{aligned} f_\epsilon(x) &= \min_{u \in \mathbb{R}^n} \sum_{i=1}^n |u_i| + \frac{\epsilon}{2} (u_i - x_i)^2 \\ &= \sum_{i=1}^n \min_{u_i \in \mathbb{R}} |u_i| + \frac{\epsilon}{2} (u_i - x_i)^2 \\ &= \sum_{i=1}^n h_\epsilon(x_i). \end{aligned}$$

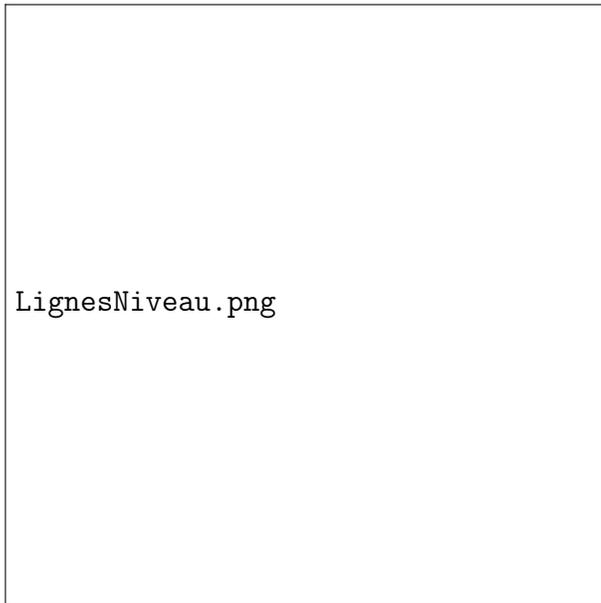


FIGURE 1 – Lignes de niveau 1,2,3 de f et de f_ϵ .

7. Dessiner les lignes de niveau 1 et 2 de la fonction $f(u) = \|u\|_1$ dans \mathbb{R}^2 . Sur le même graphe, dessiner les lignes de niveau de sa régularisée f_ϵ pour $\epsilon = 1$.

Voir Figure ??.

8. Déterminer le minimiseur (l'argmin) du problème (??) pour la fonction $f(u) = \frac{1}{2}\|Au - b\|_2^2$, où $A \in \mathbb{R}^{m \times n}$ et $b \in \mathbb{R}^m$.

On cherche

$$\arg \min_{u \in \mathbb{R}^n} \frac{1}{2}\|Au - b\|_2^2 + \frac{1}{2\epsilon}\|x - u\|_2^2.$$

Les conditions d'optimalité sont :

$$A^T(Au - b) + (u - x)/\epsilon = 0,$$

soit encore

$$u = (A^T A + I/\epsilon)^{-1}(A^T b + x/\epsilon).$$

A noter que la matrice $A^T A + I/\epsilon$ est bien inversible, puisque c'est la somme d'une matrice symétrique semi-définie positive ($A^T A$) et de l'identité (i.e. une matrice symétrique définie positive). Par exemple, en diagonalisant $A^T A$, on obtient une matrice orthogonale $U \in \mathbb{R}^{n \times n}$ et une matrice diagonale à coefficients diagonaux positifs ou nuls $\Sigma \in \mathbb{R}^{n \times n}$ telle que $A^T A = U\Sigma U^T$. Ainsi $A^T A + I/\epsilon = U(\Sigma + I/\epsilon)U^T$, dont les valeurs propres sont toutes supérieures ou égales à $1/\epsilon$.

Partie II (Problème ~8pts + X) Pour analyser la fonction f_ϵ , on va se servir de relations de dualité. On rappelle que la transformée de Fenchel de f est notée f^* et que pour $f \in \Gamma_0(\mathbb{R}^n)$, elle satisfait pour tout $u \in \mathbb{R}^n$:

$$f(u) = \sup_{y \in \mathbb{R}^n} \langle u, y \rangle - f^*(y).$$

On peut donc écrire que :

$$\begin{aligned} f_\epsilon(x) &= \min_{u \in \mathbb{R}^n} \sup_{y \in \mathbb{R}^n} \langle u, y \rangle - f^*(y) + \frac{1}{2\epsilon} \|x - u\|_2^2 \\ &= \sup_{y \in \mathbb{R}^n} \min_{u \in \mathbb{R}^n} \langle u, y \rangle - f^*(y) + \frac{1}{2\epsilon} \|x - u\|_2^2. \end{aligned}$$

La possibilité d'intervertir le min et le sup est un résultat de dualité de Fenchel-Rockefeller (non vue en cours). Ce résultat fondamental est équivalent à la dualité Lagrangienne vue dans la première partie du cours.

1. Montrer que

$$f_\epsilon(x) = \sup_{y \in \mathbb{R}^n} \langle x, y \rangle - f^*(y) - \frac{\epsilon}{2} \|y\|_2^2. \quad (2)$$

Le minimiseur en u de $\langle u, y \rangle - f^*(y) + \frac{1}{2\epsilon} \|x - u\|_2^2$ satisfait $y + (x - u)/\epsilon = 0$, soit encore $u = x - \epsilon y$. En utilisant cette expression, on obtient le résultat annoncé.

2. Montrer que f_ϵ est une fonction convexe.

Soient x_1 et x_2 deux éléments de \mathbb{R}^n . On a :

$$\begin{aligned} f\left(\frac{x_1 + x_2}{2}\right) &= \sup_{y \in \mathbb{R}^n} \langle (x_1 + x_2)/2, y \rangle - f^*(y) - \frac{\epsilon}{2} \|y\|_2^2 \\ &= \sup_{y \in \mathbb{R}^n} \frac{1}{2} \left(\langle x_1, y \rangle - f^*(y) - \frac{\epsilon}{2} \|y\|_2^2 \right) + \frac{1}{2} \left(\langle x_2, y \rangle - f^*(y) - \frac{\epsilon}{2} \|y\|_2^2 \right) \\ &\leq \sup_{y \in \mathbb{R}^n} \frac{1}{2} \left(\langle x_1, y \rangle - f^*(y) - \frac{\epsilon}{2} \|y\|_2^2 \right) + \sup_{y \in \mathbb{R}^n} \frac{1}{2} \left(\langle x_2, y \rangle - f^*(y) - \frac{\epsilon}{2} \|y\|_2^2 \right) \\ &= \frac{1}{2} (f(x_1) + f(x_2)). \end{aligned}$$

3. On suppose que $\text{dom}(f^*)$ est borné, c'est-à-dire qu'il existe $R \geq 0$ tel que $\forall x \in \mathbb{R}^n, \|x\|_2 > R, f^*(x) = +\infty$. Montrer l'encadrement suivant :

$$f(x) - \frac{\epsilon}{2} R^2 \leq f_\epsilon(x) \leq f(x).$$

Ce résultat établit que la suite $(f_\epsilon)_{\epsilon \in \mathbb{R}_+}$ converge uniformément vers f lorsque $\epsilon \rightarrow 0$.

$$\begin{aligned} f_\epsilon(x) &= \sup_{y \in \mathbb{R}^n} \langle x, y \rangle - f^*(y) - \frac{\epsilon}{2} \|y\|_2^2 \\ &\leq \sup_{y \in \mathbb{R}^n} \langle x, y \rangle - f^*(y) \\ &= f(x). \end{aligned}$$

De plus, pour tout $y \in \text{dom}(f^*)$, on a $\|y\|_2 \leq R$ et donc :

$$\langle x, y \rangle - f^*(y) - \frac{\epsilon}{2} \|y\|_2^2 \geq \langle x, y \rangle - f^*(y) - \frac{\epsilon}{2} R^2.$$

Ainsi

$$\begin{aligned} f_\epsilon(x) &= \sup_{y \in \mathbb{R}^n} \langle x, y \rangle - f^*(y) - \frac{\epsilon}{2} \|y\|_2^2 \\ &\geq \sup_{y \in \mathbb{R}^n} \langle x, y \rangle - f^*(y) - \frac{\epsilon}{2} R^2 \\ &\geq f(x) - \frac{\epsilon}{2} R^2. \end{aligned}$$

4. Soit $y(x) \in \mathbb{R}^n$ le point qui réalise le suprémum dans (??). Justifier que ce point est unique.

Le point est unique car la fonction $y \mapsto \langle x, y \rangle - f^*(y) - \frac{\epsilon}{2} \|y\|_2^2$ est fortement concave.

5. (*) Soit $x_0 \in \mathbb{R}^n$. On rappelle qu'un élément $\eta \in \partial f_\epsilon(x_0)$ satisfait pour tout $x \in \mathbb{R}^n$:

$$f_\epsilon(x) \geq f_\epsilon(x_0) + \langle \eta, x - x_0 \rangle.$$

Montrer que $y(x_0) \in \partial f_\epsilon(x_0)$.

On a pour tout $x \in \mathbb{R}^n$ et tout $y \in \mathbb{R}^n$:

$$\begin{aligned} f_\epsilon(x) &\geq \langle y, x \rangle - f^*(y) - \frac{\epsilon}{2} \|y\|_2^2 \\ &= \langle y, x_0 \rangle + \langle y, x - x_0 \rangle - f^*(y) - \frac{\epsilon}{2} \|y\|_2^2. \end{aligned}$$

Comme cette égalité est valide pour tout y , elle vaut en particulier pour $y = y(x_0)$, d'où :

$$\begin{aligned} f_\epsilon(x) &\geq \langle y(x_0), x_0 \rangle + \langle y(x_0), x - x_0 \rangle - f^*(y(x_0)) - \frac{\epsilon}{2} \|y(x_0)\|_2^2 \\ &= f_\epsilon(x_0) + \langle x - x_0, y(x_0) \rangle. \end{aligned}$$

6. Le but de cette question est de montrer que f_ϵ est différentiable et de déterminer son gradient. On donne pour ce faire le théorème de Danskin (1967).

Théorème 1. Soit Y un sous-ensemble compact de \mathbb{R}^m . Soit $\phi : \mathbb{R}^n \times Y \rightarrow \mathbb{R}$ une fonction convexe en la variable x . Soit $g(x) = \sup_{y \in Y} \phi(x, y)$. On suppose que le suprémum est atteint en un seul point $y(x)$.

Alors g est convexe et différentiable. De plus $\nabla g(x) = \frac{\partial \phi(x, y(x))}{\partial x}$.

En utilisant ce résultat, montrer que f_ϵ est différentiable et déterminer son gradient.

Dans la question II.5, on a déjà vu que $y(x_0) \in \partial f(x_0)$. Si f_ϵ est différentiable, son gradient est donc nécessairement égal à $y(x_0)$.

Pour utiliser le théorème de Danskin, il suffit de poser

$$\phi(x, y) = \langle x, y \rangle - f^*(y) - \frac{\epsilon}{2} \|y\|_2^2,$$

et $Y = \text{dom}(f^*)$ qui est borné par hypothèse. Ainsi, $f_\epsilon(x) = \sup_{y \in Y} \phi(x, y)$. Le théorème de Danskin nous apprend que f_ϵ est différentiable et que son gradient est déterminé par $\nabla f_\epsilon(x) = \partial_x \phi(x, y(x)) = y(x)$.

7. (***) Hors-barème) Montrer que ∇f_ϵ satisfait :

$$\|\nabla f_\epsilon(x_1) - \nabla f_\epsilon(x_2)\|_2 \leq \frac{1}{\epsilon} \|x_1 - x_2\|_2.$$

Indication : on peut utiliser les conditions d'optimalité $x_1 - \partial f^*(y(x_1)) - \epsilon y(x_1) \ni 0$. On a $\nabla f_\epsilon(x_i) = y(x_i)$. Les conditions d'optimalité sur $y(x_i)$ donnent :

$$\begin{aligned} x_1 - \partial f^*(y(x_1)) - \epsilon y(x_1) &\ni 0 \\ x_2 - \partial f^*(y(x_2)) - \epsilon y(x_2) &\ni 0 \end{aligned}$$

Soit $\eta_i = x_i - \epsilon y(x_i)$. D'après les inclusions ci-dessus, on a $\eta_i \in \partial f^*(y(x_i))$, de plus :

$$\begin{aligned} \langle x_1 - \eta_1 - \epsilon y(x_1), y(x_2) - y(x_1) \rangle &= 0 \\ \langle x_2 - \eta_2 - \epsilon y(x_2), y(x_2) - y(x_1) \rangle &= 0. \end{aligned}$$

En soustrayant les deux égalités, on obtient :

$$\langle y(x_2) - y(x_1), x_1 - x_2 - \eta_1 + \eta_2 - \epsilon y(x_1) + \epsilon y(x_2) \rangle = 0,$$

soit encore :

$$\epsilon \|y(x_2) - y(x_1)\|_2^2 = \langle y(x_2) - y(x_1), x_1 - x_2 \rangle + \langle y(x_2) - y(x_1), \eta_2 - \eta_1 \rangle.$$

Comme la fonction f^* est convexe, l'opérateur ∂f^* est monotone, c'est-à-dire que pour tout y_1, y_2 , et tout $\nu_1 \in \partial f^*(y_1)$ et $\nu_2 \in \partial f^*(y_2)$, il satisfait :

$$\langle \nu_2 - \nu_1, y_2 - y_1 \rangle \geq 0.$$

On en déduit que $\langle y(x_2) - y(x_1), \eta_2 - \eta_1 \rangle \geq 0$. Ainsi :

$$\epsilon \|y(x_2) - y(x_1)\|_2^2 \leq \langle y(x_2) - y(x_1), x_1 - x_2 \rangle.$$

Pour conclure, il suffit d'utiliser l'inégalité de Cauchy-Schwartz :

$$\epsilon \|y(x_2) - y(x_1)\|_2^2 \leq \|y(x_2) - y(x_1)\|_2 \|x_1 - x_2\|_2,$$

ce qui donne

$$\|y(x_2) - y(x_1)\|_2 \leq \frac{1}{\epsilon} \|x_1 - x_2\|_2.$$

Partie III (Application algorithmique ~7pts) Dans cette partie, on s'intéresse au problème :

$$\min_{x \in \mathbb{R}^n} f(x)$$

où f est une fonction convexe fermée non-différentiable qui admet au moins un minimum x^* . On suppose que $\text{dom}(f^*)$ est borné de rayon $R > 0$. Pour résoudre ce problème, on se propose de le régulariser en résolvant à place :

$$\min_{x \in \mathbb{R}^n} f_\epsilon(x), \tag{3}$$

et d'utiliser une descente de gradient accélérée sur (??). On note $(x_k)_{k \in \mathbb{N}}$ la suite générée par cette descente.

1. Rappeler le taux de convergence de la suite $(x_k)_{k \in \mathbb{N}}$.

Soit x_ϵ^* un minimiseur de (??). La fonction f_ϵ est différentiable et a un gradient $\frac{1}{\epsilon}$ -Lipschitz. Les théorèmes vus en cours indiquent donc que :

$$f_\epsilon(x_k) - f_\epsilon(x_\epsilon^*) \leq \frac{\|x_0 - x_\epsilon^*\|_2^2}{\epsilon k^2}.$$

2. (*) On note $(x_k)_{k \in \mathbb{N}}$ une suite générée par une descente de gradient accélérée. En utilisant le résultat de la question II.3, montrer que :

$$f(x_k) - f(x^*) \leq \frac{\|x_0 - x_\epsilon^*\|_2^2}{\epsilon k^2} + \frac{\epsilon R^2}{2}.$$

En utilisant la question II.3, l'inégalité précédente et le fait que x_ϵ^* soit le minimiseur de f_ϵ , on a :

$$\begin{aligned} f(x_k) &\leq f_\epsilon(x_k) + \frac{\epsilon R^2}{2} \\ &\leq f_\epsilon(x_\epsilon^*) + \frac{\|x_0 - x_\epsilon^*\|_2^2}{\epsilon k^2} + \frac{\epsilon R^2}{2} \\ &\leq f_\epsilon(x^*) + \frac{\|x_0 - x_\epsilon^*\|_2^2}{\epsilon k^2} + \frac{\epsilon R^2}{2} \\ &\leq f(x^*) + \frac{\|x_0 - x_\epsilon^*\|_2^2}{\epsilon k^2} + \frac{\epsilon R^2}{2}. \end{aligned}$$

3. Soit $c > 0$. On souhaite obtenir un point x_k satisfaisant :

$$f(x_k) - f(x^*) \leq c,$$

avec l'algorithme décrit ci-dessus. Comment choisir ϵ et le nombre d'itérations de façon optimale (i.e. quel est le nombre d'itération minimum nécessaire pour obtenir cette précision) ?

On souhaite maintenant trouver k minimum et ϵ tels que $\frac{\|x_0 - x_\epsilon^*\|_2^2}{\epsilon k^2} + \frac{\epsilon R^2}{2} \leq c$. Le nombre d'itérations doit donc satisfaire :

$$k \geq \left\lceil \frac{\|x_0 - x_\epsilon^*\|_2}{\sqrt{\epsilon(c - \frac{\epsilon R^2}{2})}} \right\rceil.$$

Pour minimiser le nombre d'itérations, il faut maximiser $\epsilon \left(c - \frac{\epsilon R^2}{2} \right)$. Il est facile de voir que le maximum est atteint en $\epsilon = c/R^2$. Ce choix mène à un nombre d'itérations égal à

$$k = \left\lceil \frac{\sqrt{2}R\|x_0 - x_\epsilon^*\|_2}{c} \right\rceil.$$

Il faut donc $O(1/c)$ itérations pour obtenir une précision c . Ceci est équivalent à dire que l'algorithme converge en $O(1/k)$.