

Moreau 邻近算子 MODULE A5

Opérateur proximal de MOREAU

Sauf mention contraire, \mathcal{X} (resp. \mathcal{Y}) est un espace de HILBERT, muni du produit scalaire noté $\langle \cdot, \cdot \rangle$ et on note $\|\cdot\|$ la norme qui découle du produit scalaire

Dans ce module, on va introduire la notion de *proximité* et d'*opérateur proximal*. Cet opérateur, proposé par Jean-Jacques MOREAU en 1965, joue un rôle prépondérant en optimisation non différentiable.
不可微 优化问题

1 Opérateur proximal 邻近算子

1.1 Définition et exemples 定义和例子

Définition 1 (Opérateur proximal) 邻近算子

Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction s.c.i. et de domaine non vide. On appelle *opérateur proximal* (ou encore : *opérateur de proximité*) associé à J l'opérateur noté prox_J qui à tout point de $x^0 \in \mathcal{X}$ associe l'ensemble des points

$$\text{prox}_J(x^0) = \underset{x \in \mathcal{X}}{\operatorname{argmin}} \left\{ J(x) + \frac{1}{2} \|x - x^0\|^2 \right\}$$

Si cet ensemble est non vide, alors ses éléments sont appelés *points proximaux* de x^0 relativement à la fonction J .

On remarquera que, par définition d'un minimiseur, on a toujours

$$\forall x^0 \in \mathcal{X}, \quad \text{prox}_J(x^0) \subset \operatorname{dom} J$$

À l'origine, cet opérateur a été défini pour des fonctions convexes; il présente dans ce cas une propriété très intéressante :

Proposition 1 (Cas convexe) 下凸函数情形

On suppose que \mathcal{X} est de dimension finie. Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe, s.c.i. et propre. Soit $x^0 \in \mathcal{X}$. Alors $\text{prox}_J(x^0)$ est non vide et contient un unique élément.

DÉMONSTRATION : Il suffit de noter que la fonction

$$\begin{cases} \mathcal{X} & \rightarrow \mathbb{R} \cup \{+\infty\} \\ x & \mapsto J(x) + \frac{1}{2} \|x - x^0\|^2 \end{cases}$$

est fortement convexe (proposition 13 du module A1 : Éléments de topologie). On en déduit l'existence d'un unique minimiseur (proposition 8 du module B1 : Méthodes d'optimisation du premier ordre). ■

REMARQUE : Dans ce cas, lorsque le **point proximal** est **défini** de manière **unique**, on identifie le **singleton** et le **point**, et on écrira

$$x^+ = \text{prox}_J(x^0)$$

au lieu de

$$\{x^+\} = \text{prox}_J(x^0)$$

Donnons un premier exemple central d'opérateur proximal.

EXEMPLE

Opérateur proximal. Considérons la fonction

$$J : \begin{cases} \mathcal{X} & \rightarrow \mathbb{R} \\ x & \mapsto \frac{1}{2} \|x\|^2 \end{cases}$$

Soit $x^0 \in \mathcal{X}$. Puisque J est une fonction **convexe** et **différentiable** (donc propre), le **point proximal** de x^0 relativement à la fonction J est, par définition, l'**unique minimiseur**, de la fonction

$$\frac{1}{2} \|x\|^2 + \frac{1}{2} \|x - x^0\|^2$$

Celui-ci est caractérisé au premier par la **règle de FERMAT**, qui s'écrit ici

$$2 \text{prox}_J(x^0) - x^0 = 0 \quad \text{soit} \quad \text{prox}_J(x^0) = \frac{x^0}{2}$$

Dans le **cas non convexe** mais **minoré**, on a le résultat suivant :

对于非上凸函数, 但是有下界的情形

Proposition 2

On suppose que \mathcal{X} est de dimension finie. Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction **minorée**, **s.c.i.** et de domaine **non vide**. Soit $x^0 \in \mathcal{X}$. Alors $\text{prox}_J(x^0)$ est **non vide**.

DÉMONSTRATION : Il suffit de noter que la fonction

$$\begin{cases} \mathcal{X} & \rightarrow \mathbb{R} \cup \{+\infty\} \\ x & \mapsto J(x) + \frac{1}{2} \|x - x^0\|^2 \end{cases}$$

est **minorée** par la fonction

$$\begin{cases} \mathcal{X} & \rightarrow \mathbb{R} \cup \{+\infty\} \\ x & \mapsto \inf J + \frac{1}{2} \|x - x^0\|^2 \end{cases}$$

qui est **coercive**. Ainsi, la **première fonction** est également **coercive**, et **admet** donc un **minimiseur** d'après la proposition 6 du module **B1 : Méthodes d'optimisation du premier ordre**. ■

EXERCICE

Translation. Soit $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction **s.c.i.** Soit $z \in \mathcal{X}$. Considérons la fonction

$$J : \begin{cases} \mathcal{X} & \rightarrow \mathbb{R} \cup \{+\infty\} \\ x & \mapsto f(x - z) \end{cases}$$

Montrer que $\forall x^0 \in \mathcal{X}, \quad \text{prox}_J(x^0) = z + \text{prox}_f(x^0 - z)$

EXERCICE

Changement d'échelle. Soit $f: \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction s.c.i. et $\alpha \neq 0$. Considérons la fonction

$$J: \begin{cases} \mathcal{X} & \rightarrow \mathbb{R} \cup \{+\infty\} \\ x & \mapsto f(\alpha x) \end{cases}$$

Montrer que $\forall x^0 \in \mathcal{X}, \quad \text{prox}_J(x^0) = \frac{1}{\alpha} \text{prox}_{\alpha^2 f}(\alpha x^0)$

Le résultat suivant permet d'exprimer l'opérateur proximal d'une fonction séparable à l'aide des opérateurs proximaux des termes qui la composent :

Proposition 3 邻近算子 可以表达为 可分函数

Soit $n \in \mathbb{N}$ et $(\mathcal{X}_i)_{1 \leq i \leq n}$ des espaces de HILBERT. Pour tout $i \in \llbracket 1; n \rrbracket$, on considère $f_i: \mathcal{X}_i \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction s.c.i. Définissons la fonction séparable

$$J: \begin{cases} \mathcal{X}_1 \times \cdots \times \mathcal{X}_n & \rightarrow \mathbb{R} \cup \{+\infty\} \\ x = (x_{(i)})_{1 \leq i \leq n} & \mapsto \sum_{i=1}^n f_i(x_{(i)}) \end{cases}$$

Soit $x^0 \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$. Alors

$$\text{prox}_J(x^0) = \text{prox}_{f_1}(x_{(1)}^0) \times \cdots \times \text{prox}_{f_n}(x_{(n)}^0)$$

REMARQUE : Notons que cette formule n'est pas valable dans le cas général de la somme de fonctions quelconques.

DÉMONSTRATION : Il suffit de vérifier que, par définition des points proximaux, on a pour tout $x^+ = (x_{(i)}^+)_{1 \leq i \leq n}$

$$x^+ \in \text{prox}_J(x^0) \iff \forall i \in \llbracket 1; n \rrbracket, \quad x_{(i)}^+ \in \text{prox}_{f_i}(x_{(i)}^0)$$

On obtient alors le résultat désiré par définition du produit cartésien. ■

Avant de nous intéresser aux propriétés de l'opérateur proximal, soulignons qu'il peut être interprété comme une généralisation de la projection orthogonale.

Proposition 4 邻近算子 可以为 投影算子的形式

Soit $\mathcal{C} \subset \mathcal{X}$ un ensemble convexe fermé non vide. Alors

$$\text{prox}_{\chi_{\mathcal{C}}} = \text{proj}_{\mathcal{C}}$$

DÉMONSTRATION : On a en effet vu dans la module A1 : Éléments de topologie que la fonction indicatrice $\chi_{\mathcal{C}}$ est bien convexe, propre et s.c.i., de sorte que l'opérateur proximal de $\chi_{\mathcal{C}}$ est bien défini. Remarquons alors que

$$\text{dom } \chi_{\mathcal{C}} = \mathcal{C}$$

La définition de l'opérateur proximal pour $\chi_{\mathcal{C}}$ s'écrit donc

$$\text{prox}_{\chi_C}(x^0) = \underset{x \in \text{dom } \chi_C}{\text{argmin}} \left\{ \chi_C(x) + \frac{1}{2} \|x - x^0\|^2 \right\} = \underset{x \in C}{\text{argmin}} \left\{ \frac{1}{2} \|x - x^0\|^2 \right\}$$

ce qui achève la preuve. ■

Cette interprétation fournit une preuve alternative de l'unicité de la projection sur un ensemble convexe fermé non vide, en appliquant la proposition 1 qui assure que le point proximal est défini de manière unique, c'est-à-dire, pour le cas de l'indicatrice d'un ensemble convexe fermé non vide, que le projeté orthogonal est défini de manière unique.

下凸函数情形, 邻近点的特征

1.2 Caractérisation du point proximal dans le cas convexe

Lorsque la fonction J est convexe, le point proximal peut être caractérisé de manière intéressante à l'aide du sous-différentiel de J :

Proposition 5 (Caractérisation du point proximal dans le cas convexe)

Soit $J: \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe, s.c.i. et propre. Pour tout $x^0 \in \mathcal{X}$, on a l'équivalence entre les deux énoncés suivants :

- (i) $x^+ = \text{prox}_J(x^0)$
- (ii) $x^0 - x^+ \in \partial J(x^+)$

DÉMONSTRATION : Soit $x^0 \in \mathcal{X}$. Par définition du point proximal, on a

$$x^+ = \underset{x \in \mathcal{X}}{\text{argmin}} \left\{ J(x) + \frac{1}{2} \|x - x^0\|^2 \right\}$$

Or, on a fait remarquer que la fonction objectif de ce problème de minimisation est (fortement) convexe ; en appliquant la règle de FERMAT dans le cas convexe, on a la caractérisation suivante de x^+ :

$$0 \in \partial \left(x \mapsto J(x) + \frac{1}{2} \|x - x^0\|^2 \right) (x^+) = \partial J(x^+) + (x^+ - x^0)$$

ce qui achève la preuve. ■

绝对值函数的邻近算子

EXEMPLE

Opérateur proximal de la valeur absolue. Considérons la fonction

$$J: \begin{cases} \mathbb{R} & \rightarrow \mathbb{R} \\ x & \mapsto |x| \end{cases}$$

Soit $x^0 \in \mathbb{R}$. la fonction J étant convexe, le point proximal $x^+ = \text{prox}_J(x^0)$ est caractérisé par

$$x^0 - x^+ \in \partial J(x^+) = \begin{cases} \left\{ \frac{x^+}{|x^+|} \right\} & \text{si } x^+ \neq 0 \\ [-1; 1] & \text{si } x^+ = 0 \end{cases}$$

On en déduit que

$$\text{prox}_J(x^0) = \begin{cases} 0 & \text{si } x^0 \in [-1; 1] \\ \frac{x^0}{|x^0|} (|x^0| - 1) & \text{si } x^0 \notin [-1; 1] \end{cases}$$

Cette fonction est connue sous le nom de *seuillage doux* (soft thresholding).

软阈值函数

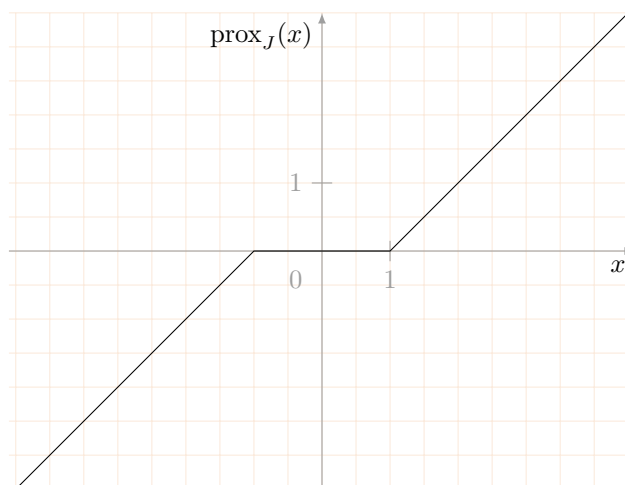


FIGURE 1 – Opérateur proximal de la valeur absolue.

EXERCICE

Opérateur proximal de la **norme ℓ_1** . Considérons la fonction

$$J : \begin{cases} \mathbb{R}^n & \rightarrow \mathbb{R} \\ x = \{x_{(i)}\}_{1 \leq i \leq n} & \mapsto \|x\|_1 = \sum_{i=1}^n |x_{(i)}| \end{cases}$$

Calculer prox_J .

La **proposition 5** permet de donner une **nouvelle écriture** à l'opérateur proximal dans le **cas convexe**, car, puisque

$$\begin{aligned} x^+ = \text{prox}_J(x^0) & \iff x^0 - x^+ \in \partial J(x^+) \\ & \iff x^0 \in x^+ + \partial J(x^+) = (\text{Id} + \partial J)(x^+) \end{aligned}$$

on peut écrire le **point proximal** comme une **image réciproque** (en confondant à nouveau le singleton et le point) :

$$x^+ = \text{prox}_J(x^0) \iff x^+ = (\text{Id} + \partial J)^{-1}(x^0)$$

在非下凸函数情形, Fermat 准则 只能提供 优化的必要条件

Dans le cas **non convexe**, la **règle de FERMAT** ne donne qu'une **condition nécessaire** d'optimalité; aussi, on a uniquement l'implication suivante :

Proposition 6

Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction **s.c.i.** Soit $x^0 \in \mathcal{X}$. On suppose qu'il existe $x^+ \in \text{prox}_J(x^0)$. Alors on a $x^0 - x^+ \in \partial J(x^+)$. Autrement dit,

$$x^+ \in \text{prox}_J(x^0) \implies x^0 - x^+ \in \partial J(x^+)$$

ce qui ne fournit pas une caractérisation des points proximaux, attendu que **certain points vérifiant** l'inclusion de **droite** peuvent ne pas être des **points proximaux**.

DÉMONSTRATION : Laissé en exercice.

1.3 Propriétés de l'opérateur proximal 邻近算子的性质

Commençons par établir quelques propriétés intéressantes de l'opérateur proximal dans le cas convexe.

Proposition 7 (Ferme non-expansivité) 非扩张闭包

Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe, s.c.i. et propre. Alors

$$\forall (x, x') \in \mathcal{X}^2, \quad \langle \text{prox}_J(x) - \text{prox}_J(x'), x - x' \rangle \geq \|\text{prox}_J(x) - \text{prox}_J(x')\|^2$$

On dit que prox_J est un opérateur fermement non-expansif.

非扩张闭算子

DÉMONSTRATION : Soit $(x, x') \in \mathcal{X}^2$. La caractérisation du point proximal (proposition 5) permet d'écrire

$$x - \text{prox}_J(x) \in \partial J(\text{prox}_J(x)) \quad \text{et} \quad x' - \text{prox}_J(x') \in \partial J(\text{prox}_J(x'))$$

En utilisant la monotonie de ∂J (qui découle de la convexité de J), on obtient que

$$\langle x - \text{prox}_J(x) - x' + \text{prox}_J(x'), \text{prox}_J(x) - \text{prox}_J(x') \rangle \geq 0$$

En développant le produit scalaire et en réarrangeant les termes obtenus, on retrouve le résultat annoncé. ■

Cette proposition permet également de démontrer que prox_J est lipschitzien :

Corollaire 1

Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe, s.c.i. et propre. Alors prox_J est 1-lipschitzien.

On dit également que prox_J est un opérateur non-expansif.

DÉMONSTRATION : Il suffit d'appliquer l'inégalité de CAUCHY-SCHWARZ à la définition de forte non-expansivité de l'opérateur proximal. ■

Le résultat suivant est très utilisé dans les preuves de convergence d'algorithmes d'optimisation impliquant l'opérateur proximal d'une fonction convexe :

Proposition 8 (Inégalité proximale) 邻近不等式 (convexe)

Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe, propre et s.c.i. et soit $x^0 \in \mathcal{X}$. Soit $\tau > 0$. On pose

$$x^+ = \text{prox}_{\tau J}(x^0)$$

$$\text{Alors} \quad \forall x \in \mathcal{X}, \quad J(x) \geq J(x^+) - \frac{1}{\tau} \langle x - x^+, x^+ - x^0 \rangle$$

DÉMONSTRATION : D'après la caractérisation du point proximal (proposition 5),

$$x^+ \in x^0 - \tau \partial J(x^+) \quad \text{soit} \quad \frac{1}{\tau} (x^0 - x^+) \in \partial J(x^+)$$

de sorte que, par définition des sous-gradients, on a

$$\forall x \in \mathcal{X}, \quad J(x) \geq J(x^+) + \frac{1}{\tau} \langle x^0 - x^+, x - x^+ \rangle$$

Un réarrangement des termes donne alors le résultat désiré. ■

Enfin, signalons le résultat suivant, valable dès que l'opérateur proximal est bien défini (la convexité n'est pas requise) :

Proposition 9 邻近不等式 (一般)

Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction s.c.i. Soit $x^0 \in \mathcal{X}$ tel que $\text{prox}_J(x^0)$ soit non vide. On pose

$$x^+ \in \text{prox}_J(x^0)$$

Alors
$$J(x^+) \leq J(x^0) + \frac{1}{2} \|x^+ - x^0\|^2 \leq J(x^0)$$

DÉMONSTRATION : Par définition d'un minimiseur global, on a pour tout $x \in \text{dom } J$

$$J(x^+) + \frac{1}{2} \|x^+ - x^0\|^2 \leq J(x) + \frac{1}{2} \|x - x^0\|^2$$

En particulier, en prenant $x = x^0$, on a

$$J(x^+) + \frac{1}{2} \|x^+ - x^0\|^2 \leq J(x^0) + \frac{1}{2} \|x^0 - x^0\|^2 = J(x^0)$$

On en déduit alors le résultat souhaité. ■

1.4 Points fixes de l'opérateur proximal 邻近算子的固定点

Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe, s.c.i. et propre. On s'intéresse dans cette section aux points fixes de prox_J , c'est-à-dire aux points $\bar{x} \in \mathcal{X}$ tels que

$$\bar{x} = \text{prox}_J(\bar{x})$$

(On rappelle que, dans le cas convexe, prox_J est une application définie sur \mathcal{X} et à valeurs dans \mathcal{X} .) On peut alors démontrer le résultat suivant :

Proposition 10 邻近算子的固定点 等价于 J 的最小值点

Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe, s.c.i. et propre. Soit $\bar{x} \in \mathcal{X}$. Alors on a équivalence entre les deux énoncés suivants :

- (i) \bar{x} est un point fixe de prox_J ;
固定点
- (ii) \bar{x} est un minimiseur de J .
最小值点

DÉMONSTRATION : Il s'agit d'une conséquence directe de la proposition 5 en notant que les points critiques d'une fonction convexe sont ses minimiseurs. ■

Dans le cas ^{非凸情形} non convexe, on a un résultat plus faible, puisque la règle de FERMAT ne caractérise plus les minimiseurs et que l'opérateur proximal n'est pas nécessairement bien défini ni de manière unique :

Proposition 11 — 非凸情形, 邻近算子的固定点 是 J 的驻点

Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction s.c.i. Soit $\bar{x} \in \mathcal{X}$ tel que $\bar{x} \in \text{prox}_J(\bar{x})$. Alors \bar{x} est un point critique ^{驻点} de J .

DÉMONSTRATION : Il s'agit cette fois-ci d'une conséquence immédiate de la proposition 6. ■

Cette propriété, essentielle, est au cœur de nombreux algorithmes d'optimisation non lisse. C'est le cas en particulier de l'algorithme du point proximal (module B3 : Algorithme du point proximal), qui cherche un point critique de J en appliquant les itérations du point fixe à l'opérateur proximal.

2 Enveloppes de MOREAU ^{Moreau 包络}

On va à présent s'intéresser de manière plus spécifique à l'opérateur proximal dans le cas d'une fonction convexe. ^{邻近算子}

2.1 Définition

Définition 2 (Enveloppe de MOREAU) — Moreau 包络

Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe, s.c.i. et propre. Soit $\tau > 0$. On appelle ^{Moreau 包络} *enveloppe de MOREAU* d'indice τ de J la fonction notée ${}^\tau J$ définie par

$$\forall x^0 \in \mathcal{X}, \quad {}^\tau J(x^0) = \min_{x \in \mathcal{X}} \left\{ J(x) + \frac{1}{2\tau} \|x - x^0\|^2 \right\}$$

Lorsque J est convexe, s.c.i. et propre, la fonction ${}^\tau J$ est bien définie d'après la proposition 1 puisque le minimiseur est atteint quel que soit $x^0 \in \mathcal{X}$.

Les enveloppes de MOREAU appartiennent à la classe des fonctions définies par *inf-convolution*. L'inf-convolution (ou encore *convolution infimale*) de deux fonctions ^{下卷积} f et g , notée $f \square g$, est définie comme suit :

$$\forall x^0 \in \mathcal{X}, \quad f \square g(x^0) = \inf_{x \in \mathcal{X}} \left\{ f(x) + g(x^0 - x) \right\}$$

Ainsi, l'enveloppe de MOREAU ${}^\tau J$ d'indice $\tau > 0$ de J est l'inf-convolution de J et de la fonction $\|\cdot\|^2/(2\tau)$:

$${}^\tau J = J \square \left(\frac{1}{2\tau} \|\cdot\|^2 \right)$$

L'inf-convolution possède des propriétés intéressantes, parmi lesquelles la préservation de la convexité et de la différentiabilité. C'est pourquoi elle apparaît fréquemment dans la littérature de l'optimisation convexe.

到非空凸集的距离

EXEMPLE

Distance à un ensemble convexe non vide. Soit $\mathcal{C} \subset \mathcal{X}$ un ensemble convexe fermé non vide. Alors, pour tout $\tau > 0$, l'enveloppe de MOREAU d'indice τ de $\chi_{\mathcal{C}}$ définit la distance euclidienne au carré au convexe \mathcal{C}

$$\forall x^0 \in \mathcal{X}, \quad \text{dist}(x^0, \mathcal{C}) = \min_{x \in \mathcal{C}} \|x - x^0\| = \sqrt{2\tau \chi_{\mathcal{C}}(x^0)}$$

2.2 Régularisation de MOREAU–YOSIDA Moreau-Yosida 正则化

Établissons quelques propriétés intéressantes à propos de l'enveloppe de MOREAU. Commençons par remarquer que

$$\forall x^0 \in \mathcal{X}, \quad {}^\tau J(x^0) = \frac{1}{\tau} \min_{x \in \mathcal{X}} \left\{ \tau J(x) + \frac{1}{2} \|x - x^0\|^2 \right\}$$

Proposition 12 Moreau 包络的梯度

Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe, s.c.i. et propre. Soit $\tau > 0$. Alors ${}^\tau J$ est une fonction différentiable. Son gradient est donné par

$$\forall x \in \mathcal{X}, \quad \nabla({}^\tau J)(x) = \frac{1}{\tau} (x - \text{prox}_{\tau J}(x))$$

Il s'agit d'un cas particulier du *théorème de l'enveloppe*, qui permet de mesurer l'influence d'un paramètre sur le minimum d'une fonction dépendant de ce paramètre. Dans le cas de l'enveloppe de MOREAU, le paramètre en question est le point x^0 .

DÉMONSTRATION : On admet pour l'instant ce résultat, qui sera plus facile à démontrer une fois qu'on aura abordé les notions de dualité de FENCHEL dans le module **A8 : Dualité de FENCHEL et conjuguée convexe**.

Proposition 13 Moreau 包络函数的正则性

Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe, s.c.i. et propre. Soit $\tau > 0$. Alors ${}^\tau J$ est $(1/\tau)$ -régulière.

DÉMONSTRATION : Ce résultat est également provisoirement admis et sera démontré dans le module **A8 : Dualité de FENCHEL et conjuguée convexe**.

Un corollaire presque immédiat de la preuve du lemme 7 est la convexité des enveloppes de MOREAU :

Corollaire 2 Moreau 包络函数的凸性

Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe, s.c.i. et propre. Soit $\tau > 0$. Alors ${}^\tau J$ est une fonction convexe.

DÉMONSTRATION : En effet, la **monotonie** du **gradient** de ${}^\tau J$ est une **conséquence** de la **ferme non-expansivité** de l'opérateur proximal (proposition [7](#)) :

$$\begin{aligned}\langle \nabla({}^\tau J)(x) - \nabla({}^\tau J)(x'), x - x' \rangle &= \left\langle \frac{1}{\tau}(x - \text{prox}_{\tau J}(x)) - \frac{1}{\tau}(x' - \text{prox}_{\tau J}(x')), x - x' \right\rangle \\ &= \frac{1}{\tau} \langle x - \text{prox}_{\tau J}(x) - x' + \text{prox}_{\tau J}(x'), x - x' \rangle \geq 0\end{aligned}$$

qui prouve que ${}^\tau J$ est convexe. ■

Proposition 14 — J et ${}^\tau J$ ont les mêmes minimiseurs et le même minimum

Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction **convexe**, **s.c.i.** et **propre**. Soit $\tau > 0$. Alors J et ${}^\tau J$ ont **mêmes minimiseurs** et **même minimum**.

DÉMONSTRATION : On démontre séparément les deux propriétés.

- J et ${}^\tau J$ ont **mêmes minimiseurs**. Il suffit de **combiner** les propositions [10](#) et [12](#), en remarquant que, si $\tau > 0$, alors J et ${}^\tau J$ ont **mêmes minimiseurs**.
- J et ${}^\tau J$ ont **même minimum**. On commence par remarquer que, puisque $\|x - x^0\|^2 \geq 0$, on a pour tout $x^0 \in \mathcal{X}$

$${}^\tau J(x^0) = \min_{x \in \mathcal{X}} J(x) + \frac{1}{2\tau} \|x - x^0\|^2 \geq \inf_{x \in \mathcal{X}} J(x)$$

En passant à la **borne inférieure** sur x^0 , il s'ensuit que

$$\inf_{x^0 \in \mathcal{X}} {}^\tau J(x^0) \geq \inf_{x \in \mathcal{X}} J(x)$$

Par ailleurs, par **optimalité**, on a

$${}^\tau J(x^0) = \min_{x \in \mathcal{X}} J(x) + \frac{1}{2\tau} \|x - x^0\|^2 \leq J(x^0) + \frac{1}{2\tau} \|x^0 - x^0\|^2 = J(x^0)$$

En passant à nouveau à la **borne inférieure** sur x^0 , on obtient

$$\inf_{x^0 \in \mathcal{X}} {}^\tau J(x^0) \leq \inf_{x \in \mathcal{X}} J(x)$$

La **conclusion** suit **immédiatement** si J atteint son minimum. ■

Cette dernière proposition justifie que l'on **cherche** parfois à **remplacer** le problème de la **minimisation** J par celui de la **minimisation** d'une **enveloppe** de MOREAU ${}^\tau J$.

L'enveloppe de MOREAU est également connue sous le nom de *régularisation de MOREAU-YOSIDA* de la fonction J . Il s'agit en effet d'une famille de fonctions qui présentent des caractéristiques intéressantes, à savoir une **forte régularité** (chaque enveloppe est différentiable à gradient lipschitzien) et une **convexité**. Par ailleurs, il est aisé d'observer que, lorsque l'indice τ tend vers $+\infty$, l'enveloppe de MOREAU associée "**converge**" (dans un sens qu'il faudrait préciser) vers la **fonction** J . De plus, on verra plus bas que, par ailleurs, J et ses **enveloppes** de MOREAU ont **mêmes minimiseurs**.

On termine ce module avec un exemple d'enveloppes de MOREAU.

绝对值函数的 Moreau - Yosida 正则化

EXEMPLE

Régularisation de MOREAU-YOSIDA de la valeur absolue. Considérons la fonction

$$J : \begin{cases} \mathcal{X} & \rightarrow \mathbb{R} \\ x & \mapsto |x| \end{cases}$$

Cette fonction est **convexe**, **continue** et **non différentiable en 0**. Intéressons-nous à sa régularisation de MOREAU-YOSIDA. On cherche donc à **calculer le minimiseur de la fonction**

$$f_\tau : x \mapsto |x| + \frac{1}{2\tau} (x - x^0)^2$$

pour $\tau > 0$. En reprenant les **calculs** de l'exemple de l'opérateur proximal de J , on vérifie que l'**unique point x^*** vérifiant $0 \in \partial f_\tau(x^*)$ est

$$x^* = \begin{cases} x^0 - \tau & \text{si } x^0 > \tau \\ x^0 + \tau & \text{si } x^0 < -\tau \\ 0 & \text{si } x^0 \in [-\tau; \tau] \end{cases}$$

On en **déduit que**

$$\tau J(x^0) = f_\tau(x^*) = \begin{cases} x^0 - \frac{\tau}{2} & \text{si } x^0 > \tau \\ -x^0 - \frac{\tau}{2} & \text{si } x^0 < -\tau \\ \frac{(x^0)^2}{2\tau} & \text{si } x^0 \in [-\tau; \tau] \end{cases}$$

On reconnaît (à une constante multiplicative près) une **fonction de HUBER** (module **A4 : Fonctions régulières**). On pourra vérifier que cette fonction est bien **convexe** et **dérivable**, de **dérivée $(1/\tau)$ -lipschitzien**. Cette famille de fonctions est illustrée par la figure [2](#)

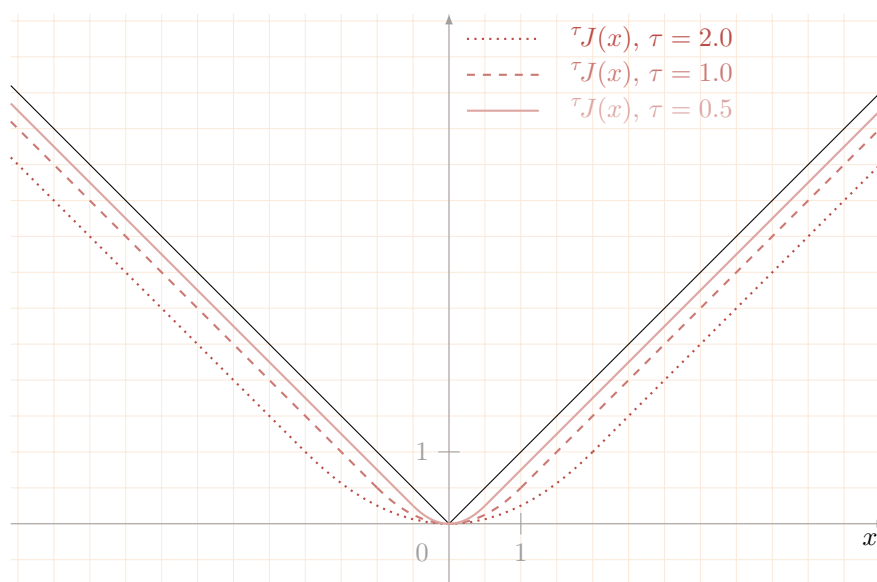


FIGURE 2 – Exemples de régularisation de MOREAU de la valeur absolue (la courbe représentative de cette dernière apparaît en noir).

Pour aller plus loin

L'opérateur proximal dans les algorithmes d'optimisation non lisse.

L'opérateur proximal apparaît dans de nombreux algorithmes d'optimisation du premier ordre, tels que la méthode du point proximal (module **B3 : Algorithme du point proximal**), l'éclatement explicite-implicite, ou encore les méthodes basées sur l'éclatement de DYSKTRA (module **B4 : Éclatement primal d'opérateurs**). Comme on le verra dans les modules dédiés, l'opérateur proximal présente des propriétés très intéressantes qui justifient son utilisation, notamment en optimisation convexe non lisse, bien que des extensions au cas non convexe puissent être également envisagées.

Évaluation numérique de l'opérateur proximal. Comme on l'a vu dans ce module, cependant, le calcul du point proximal peut être fastidieux (il suffit de songer aux cas particuliers des projections orthogonales). L'éclatement de DYSKTRA propose une solution pour remédier à ce problème dans certains cas (module **B4 : Éclatement primal d'opérateurs**).

La régularisation de MOREAU–YOSIDA. Les enveloppes de MOREAU jouent un rôle essentiel dans l'interprétation de concepts connus tels que le *lagrangien augmenté* (module **A6 : Dualité min-max**) ou encore pour établir un lien naturel entre l'algorithme du point proximal et la méthode du gradient explicite (module **B2 : Méthodes du gradient explicite**).