Université Mohamed El Bachir El Ibrahimi de Bordj Bou Arréridj

Faculté des Mathématiques et de l'Informatique

Département de Mathématiques

Mémoire

Présenté par

NEFNAF KENZA

Pour l'obtention du diplôme de

Master

Filière : Mathématiques

Spécialité :Analyse Mathématique et applications

Thème

Les méthodes proximales pour l'optimisation non différentiable

---

Soutenu publiquement .Septembre 2020 devant le jury composé de

		Université de Bordj Bou Arréridj	Président
Addoune. Ismaile	*MCA*	Université de Bordj Bou Arréridj	Encadreur
		Université de Bordj Bou Arréridj	Examinateur

Promotion 2019/2020

Remerciement

En premier lieu, je tiens à remercier tout particulièrement, mon directeur de mémoire Mr Addoune.Ismaile, pour m'avoir donné l'opportunité de découvrir le monde passionnant de la recherche.

Je tiens à remercier les membres de jury pour leur participation à l'évaluation de ce travail et qui ont bien voulu ma honneur par leurs présences.

Je voudrais remercier tous les enseignements du Master Mathématique et Applications de l'université BBA qui ont contribué à notre formation.

Enfin, un grand Merci à ma famille en commençant à bien entendu par mes parents pour m'avoir toujours soutenue et conseillée et je remercier mes amis et mes collèges qui on toujours été là pour moi. Leur soutien inconditionnel et leurs encorbellements on été d'une grande aide.

Table des matières

Introduction générale

Les algorithmes d'optimisation sont généralement écrits pour minimiser une fonction. Si l'on désire maximiser une fonction, il suffira de minimiser son opposée. Nous nous intéressons aux méthodes d'optimisation adaptées au cas où la fonction à minimiser est convexe mais non différentiable. De manière général, les méthodes pour l'optimisation non différentiable s'inspirent largement des méthodes différentiables.

Dans ce mémoire, on s'intéresse aux méthodes proximaux pour résoudre des problèmes d'optimisation non différentiable qui s'écrivent sous la forme :

$$\begin{cases} \text{Minimiser } F(x) = f_1(x) + \cdots + f_m(x), \\ x \in \mathbb{R}^n. \end{cases} \tag{1}$$

Où $f_1, f_2, \cdots, f_m$ sont des fonctions convexes ne sont pas toutes différentiable. Il y a une classe spécifique d'algorithmes qui peuvent résoudre le problème d'optimisation (1).

Pour $m = 2$, la fonction F que nous voulons minimiser n'est pas différentiable, car certains algorithmes dits de splitting permettent de minimiser la somme de fonctions en alternant des opérations élémentaires utilisant chacune des fonctions prises séparément.

Ce mémoire est divisé en deux chapitres. Le premier chapitre est consacré au rappel de quelques notions de base sur la convexité et d'autre part on présente les outils spécifiques d'optimisation non lisse comme le sous différentiel et les opérateurs monotones. Nous étudierons ensuite les conditions d'optimalité dans le cadre non lisse. Dans le deuxième chapitre, on s'intéresse à l'opérateur proximal et ses propriétés. Nous étudierons plusieurs algorithmes proximaux et leur cadre d'applications, l'algorithme du point proximal, l'algorithme Forward-Backward et l'algorithme Douglas-Rachford ainsi que l'algorithme des directions alternées. Pour conclure cette partie nous verrons comment un même problème peut être réécrit de plusieurs manières et ainsi être résolu via divers algorithmes. On termine ce mémoire par une conclusion finale.

Chapitre 1

Notions de base

1.1 Éléments d'analyse convexe

1.1.1 Ensemble convexe

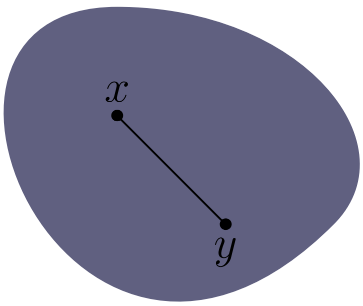
Définition 1.1
Soit A une partie de $\mathbb{R}^n$. On dit que A est convexe s'il contient tout segment dont les extrémités sont dans A, c'est a dire :

$$\forall x, y \in A \text{ et } \lambda \in [0,1], \qquad \lambda x + (1-\lambda)y \in A. \tag{1.1}$$

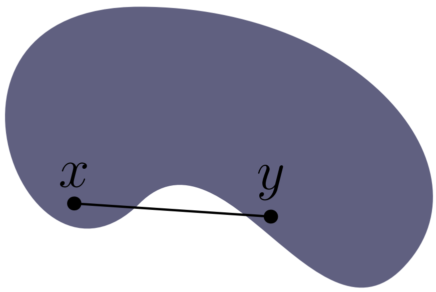
Exemple 1.1

1) Tout ensemble affine pour qui la relation (1.1) est vraie pour tout $\lambda \in \mathbb{R}$, est convexe.

2) Les boules de $\mathbb{R}^n$ sont convexes.



(a) Un ensemble convexe.



(b) Un ensemble non convexe.

Définition 1.2 *(Demi-espaces)*
Soit H un hyperplan affine, soient f une forme linéaire non nulle sur $\mathbb{R}^n$ et a un réel tels que

$$H = \{x \in \mathbb{R}^n | f(x) = a\}.$$

On appelle demi-espaces associés à l'hyperplan H les ensembles

$$D_+ = \{x \in \mathbb{R}^n | f(x) \leq a\} \text{ et } D_- = \{x \in \mathbb{R}^n | f(x) \geq a\}.$$

Les demi-espaces stricts associés à l'hyperplan H sont les ensembles

$$D'_+ = \{x \in \mathbb{R}^n | f(x) < a\} \text{ et } D'_- = \{x \in \mathbb{R}^n | f(x) > a\}.$$

Ces ensembles sont convexes.

Proposition 1.1
On a les propriétés élémentaires suivantes :

1. *Si A et B sont deux convexes de $\mathbb{R}^n$ et λ_1, λ_2 deux réels, alors $\lambda_1 A + \lambda_2 B$ est un convexe de $\mathbb{R}^n$.*
2. *Si $(A_j)_{j\in J}$ est une famille quelconque de convexes de $\mathbb{R}^n$, alors $\cap_{j\in J} A_j$ est un convexe de $\mathbb{R}^n$.*
3. *Si A est un convexe de $\mathbb{R}^n$ et $f : \mathbb{R}^n \to \mathbb{R}^m$ est une application affine, i.e. $f(x) = Cx + b$ pour certains $C \in \mathbb{R}^{m\times n}$ et $b \in \mathbb{R}^m$, alors $f(A)$ est un convexe de $\mathbb{R}^m$.*
4. *On appelle polyèdre convexe de $\mathbb{R}^n$ un ensemble de la forme $\{x \in \mathbb{R}^n : Ax \leq b\}$, où $A : \mathbb{R}^n \to \mathbb{R}^m$ est une application linéaire, $b \in \mathbb{R}^m$ et l'inégalité $Ax \leq b$ se comprend composante par composante : $(Ax)_i \leq b_i$, pour tout $i \in \{1, \cdot, \cdot, \cdot, \cdot, m\}$. C'est donc l'intersection d'un nombre fini de demi-espaces. D'après ce qui précède, c'est un convexe.*

Définition 1.3 (Combinaison convexe)
Soient $x_1, \ldots, x_k$ un nombre fini des points de $\mathbb{R}^n$ et $\lambda_1, \ldots, \lambda_k$ des réels tels que

$$\lambda_i \geq 0 \quad \forall i = 1, \ldots, k \text{ avec } \sum_{i=1}^{k} \lambda_i = 1.$$

On dit que

$$x = \sum_{i=1}^{k} \lambda_i x_i.$$

est une combinaison convexe *des points x_i.*

1.1.2 Enveloppe convexe

Définition 1.4
Soit S un ensemble de $\mathbb{R}^n$. L'enveloppe convexe de S est l'intersection de toutes les parties convexes contenant S, par conséquent c'est le plus petit convexe contenant S, on le note $\text{conv}(S)$.

Proposition 1.2
L'enveloppe convexe de S est l'ensemble de toutes les combinaisons convexes d'éléments de S où

$$\text{conv}(S) = \Big\{ \sum_{i=1}^{k} \lambda_i x_i, \quad \lambda_i \geq 0, \quad x_i \in S, \sum_{i=1}^{k} \lambda_i = 1 \Big\}.$$

Preuve

Comme $\text{conv}(S)$ est convexe, cet ensemble contient toutes les combinaisons convexes de ses éléments, donc en particulier celles des éléments de S.
Inversement, il est facile de voir que l'ensemble des combinaisons convexes des éléments de S est un convexe. Il contient donc $\text{conv}(S)$ qui est le plus petit convexe contenant S. ■

Théorème 1.1 (Carathéodory)
Soit S un ensemble non vide d'un espace vectoriel de dimension finie n, alors tout élément de $\text{conv}(S)$ *s'écrit comme combinaison convexe d'au plus $n+1$ points d'éléments de S.*

1.1.3 Fonctions convexes

Dans cette section, on présente quelques propriétés des fonctions convexes. Les fonctions considérées seront définies sur $\mathbb{R}^n$ à valeurs dans $\overline{\mathbb{R}}$ où $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$.

On définit le domaine de la fonction f de la façon suivante :

Définition 1.5 (Domaine d'une fonction)
Le domaine d'une fonction f, noté $\text{dom}(f)$, *est défini par*

$$\text{dom}(f) = \{x \in \mathbb{R}^n : f(x) < +\infty\}.$$

Exemple 1.2 *Soit $f : \mathbb{R} \to \overline{\mathbb{R}}$*

$$f(x) = \begin{cases} -\infty & si \quad x < 0 \\ x^2 & si \quad 0 \leq x \leq 1 \\ +\infty & si \quad x > 0 \end{cases}$$

alors $\text{dom}(f) =]-\infty, 1]$.

Définition 1.6
Soit $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ une fonction, l'épigraphe de f est l'ensemble de $\mathbb{R}^n \times \mathbb{R}$ défini par :

$$\text{epi}(f) = \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq r\}.$$

Définition 1.7
Soit $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, on dit que f est convexe si et seulement si son épigraphe est un ensemble convexe de l'espace $\mathbb{R}^{n+1}$.

Exemple 1.3 *Soit $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ une fonction définie par*

$$f(x) = \begin{cases} +\infty & si \quad x < 0 \\ -\sqrt{x} & si \quad x \geq 0. \end{cases}$$

La fonction f est convexe car son épigraphe est un ensemble convexe de $\mathbb{R}^2$ (voir Figure 1.2).

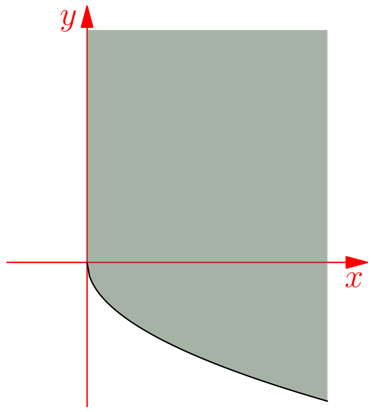


FIGURE 1.2 – L'épigraphe de la fonction f de l'exemple (1.3).

Théorème 1.2
Soit $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ une fonction. f est convexe si est seulement si

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \quad \forall x, y \in \mathrm{dom}(f),\ \forall \lambda \in]0,1[. \tag{1.2}$$

Si l'inégalité (1.2) est stricte si $x \neq y$, on dit que f est strictement convexe.

Preuve
On suppose que f est convexe. Soient $x, y \in \mathrm{dom}(f)$ et $\lambda \in]0,1[$, si $f(x) = -\infty$ ou $f(y) = -\infty$ (on ne peut pas écrire $(x, -\infty) \in \mathrm{epi}(f)$ ou $(y, -\infty) \in \mathrm{epi}(f)$) dans ce cas il faut écrire $(x, r) \in \mathrm{epi}(f) \quad \forall r \in \mathbb{R}$,

$$\lambda(x, r) + (1-\lambda)(y, f(y)) \in \mathrm{epi}(f) \quad \text{pour tout } r$$

Alors,

$$f(\lambda x + (1-\lambda)y) \leq \lambda r + (1-\lambda) f(y) \quad \text{pour tout } r \in \mathbb{R},$$

en faisant, tendre r vers $-\infty$, on obtient $f(\lambda x + (1-\lambda)y) = -\infty$. Par conséquent l'inégalité (1.2) est vérifiée.
les points $(x, f(x))$ et $(y, f(y))$ appartiennent à l'épigraphe de f qui est convexe, donc

$$\lambda(x, f(x)) + (1-\lambda)(y, f(y)) \in \mathrm{epi}(f).$$

Par suite

$$\lambda(x, f(x)) + (1-\lambda)(y, f(y)) = (\lambda x + (1-\lambda)y, \lambda f(x) + (1-\lambda) f(y)) \in \mathrm{epi}(f),$$

on obtient alors

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$

Réciproquement, Soient (x, r_1) et (y, r_2) deux éléments de l'épigraphe de f et $\lambda \in]0,1[$, alors

$$f(x) \leq r_1. \tag{1.3}$$

$$f(y) \leq r_2. \tag{1.4}$$

En multipliant l'équation (1.3) par λ, et l'équation (1.4) par $(1-\lambda)$ et en additionnant le tout on obtient

$$\lambda f(x) + (1-\lambda) f(y) \leq \lambda r_1 + (1-\lambda) r_2.$$

D'après la relation (1.1), on a

$$f(\lambda x + (1-\lambda) y) \leq \lambda r_1 + (1-\lambda) r_2.$$

Donc

$$(\lambda x + (1-\lambda) y, \lambda r_1 + (1-\lambda) r_2) = \lambda(x, r_1) + (1-\lambda)(y, r_2) \in \operatorname{epi}(f).$$

■

Exemple 1.4

1) La fonction constante $f \equiv \infty$ est un cas particulier de fonction convexe, dont l'épigraphe et le domaine sont vides.

2) La fonction $f : x \in \mathbb{R} \mapsto x^2$ est une fonction strictement convexe car géométriquement la corde entre deux points distincts du graphe se trouve strictement en dessous du graphe. Mais $f : x \in \mathbb{R} \mapsto |x|$ n'est pas strictement convexe.

Théorème 1.3 *(Caractérisation différentielle de la convexité)*
Soit $S \subseteq \mathbb{R}^n$ un ensemble convexe ouvert non vide et $f : S \to \mathbb{R}$ une fonction différentiable. Les propriétés suivantes sont équivalentes :

(a) f est convexe.

(b) Les hyperplans tangents sont des minorants

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \quad \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n. \tag{1.5}$$

(c)

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0 \quad \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n. \tag{1.6}$$

La relation (1.6) nous dit que l'opérateur gradient d'une fonction convexe est monotone.

Preuve
On montre, d'abord l'équivalence entre (a) et (b). Supposons que la fonction f soit convexe sur S. Soient x et y deux points de S, on a alors

$$f(\lambda y + (1-\lambda) x) \leq \lambda f(y) + (1-\lambda) f(x), \quad \forall \lambda \in]0, 1[. \tag{1.7}$$

La relation (1.7) est équivalent à la suivante

$$f(x + \lambda(y - x)) - f(x) \leq \lambda(f(y) - f(x)), \quad \forall \lambda \in]0, 1[,$$

donc

$$\frac{f(x + \lambda(y - x)) - f(x)}{\lambda} \leq (f(y) - f(x)).$$

En passant à la limite quand $\lambda \to 0^+$, on obtient le résultat annoncée.

Réciproquement, supposons la relation (1.5) vérifiée et montrons la convexité de la fonction f sur S. Soient x et y deux éléments de S, $\lambda \in]0,1[$ et $z = \lambda x + (1-\lambda)y \in S$. En utilisant l'inégalité (1.5) on obtient

$$\langle \nabla f(z), x - z\rangle + f(z) \leq f(x).$$

$$\langle \nabla f(z), y - z\rangle + f(z) \leq f(y).$$

Soit encore :

$$\langle \nabla f(z), (1-\lambda)(x-y)\rangle + f(\lambda x + (1-\lambda)y) \leq f(x).$$

$$\langle \nabla f(z), \lambda(y-x)\rangle + f(\lambda x + (1-\lambda)y) \leq f(y).$$

En multipliant les deux inégalités par λ et $(1-\lambda)$ respectivement puis en additionnant, on trouve l'inégalité de convexité,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y), \text{ pour tout } \lambda \in]0,1[.$$

Ensuite, montrons $(b) \Leftrightarrow (c)$
Supposons que f est convexe sur S et soient x et y deux points de S. On obtient d'après (1.5) que

$$f(x) \geq f(y) + \langle \nabla f(y), x - y\rangle. \tag{1.8}$$

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle. \tag{1.9}$$

Le résultat s'obtient en sommant les inégalités (1.8) et (1.9).
Réciproquement, soient x et y deux points de S. Nous allons montrer que l'inégalité (1.6) a lieu par conséquent f sera convexe d'après l'inégalité (1.5). D'après le théorème de la moyenne, il existe $z = \lambda x + (1-\lambda)y \in]x, y[$ tel que

$$\langle \nabla f(z), y - x\rangle = f(y) - f(x). \tag{1.10}$$

En appliquant la relation (1.8) aux points x et z on obtient

$$\langle \nabla f(z) - \nabla f(x), z - x\rangle \geq 0. \tag{1.11}$$

Or $z - x = (1-\lambda)(y-x)$, on en déduit donc que l'inégalité (1.11) est équivalente à

$$\langle \nabla f(z) - \nabla f(x), y - x\rangle \geq 0. \tag{1.12}$$

Par conséquent, on obtient

$$\langle \nabla f(z), y - x\rangle \geq \langle \nabla f(x), y - x\rangle. \tag{1.13}$$

Les relations (1.10) et (1.13) donnent

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle.$$

Ce qui prouve la convexité de f sur S. ∎

Proposition 1.3 *(Caractérisation de second ordre de la convexité)*
Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction deux fois différentiable, elle est convexe si et seulement si $\nabla^2 f(x) \succeq 0$ pour tout $x \in \mathbb{R}^n$ (i.e. la hessienne de f est semi-définie positive en tout point $x \in \mathbb{R}^n$.)

Définition 1.8 *Une fonction $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ est dite fortement convexe s'il existe $\mu > 0$ tel que*

$$\forall (x,y) \in \mathbb{R}^n \times \mathbb{R}^n, f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) - \frac{\mu}{2}\lambda(1-\lambda) \parallel x - y \parallel^2 .$$

Remarque 1.1 *Toute fonction fortement convexe est strictement convexe.*

Proposition 1.4
Soit $S \subseteq \mathbb{R}^n$ un ensemble convexe ouvert non vide et $f : S \to \mathbb{R}$ une fonction.

1) Si la fonction f est différentiable, et f est μ-fortement convexe sur S si et seulement si

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y - x\|^2 \quad \text{pour tout } x, y \in S.$$

2) Si f est deux fois différentiable et f est μ-fortement convexe sur S si et seulement si pour tout $x \in S$ on a

$$\langle h, \nabla^2 f(x) h \rangle \geq \mu \|h\|^2 \quad \text{pour tout } h \in \mathbb{R}^n. \tag{1.14}$$

Exemple 1.5 *On donne quelques exemples de fonction fortement convexe*

a) La fonction $f(x) = \frac{1}{2}\|x\|^2$ est fortement convexe.

b) La fonction $f(x) = -log(x)$ n'est pas fortement convexe sur $\mathbb{R}_+^*$. Par contre, elle l'est sur tout intervalles $[a,b], 0 < a < b$. Sa constante de forte convexité est alors égale à 1.

Preuve

a) Pour montrer que la fonction $x \mapsto \frac{1}{2}\|x\|^2$ est fortement convexe, il suffit d'utiliser(1.14) de la proposition précédente. En effet, $\nabla^2 f(x) = I$ et $\langle h, Ih \rangle = \|h\|^2 \quad \forall h$. Donc $x \mapsto \frac{1}{2}\|x\|^2$ est fortement convexe de rapport 1.

b) On utilise la même proposition pour montrer que la fonction $f(x) = -log(x)$ n'est pas fortement convexe sur $]0, +\infty[$. En effet, $f'(x) = -\frac{1}{x}$ et $f''(x) = \frac{1}{x^2}$.
Soit μ tel que

$$\begin{aligned} \langle h, f''(x) h \rangle \geq \mu h^2 \quad &\Rightarrow \quad h^2 f''(x) \geq \mu h^2 \qquad \forall x > 0, \forall h \in \mathbb{R} \\ &\Rightarrow \quad \frac{1}{x^2} \geq \mu \quad \forall x > 0 \\ &\Rightarrow \quad 0 \geq \mu.(\text{ en faisant tendre x vers } +\infty) \end{aligned}$$

Donc, il n'existe pas de $\mu > 0$ tel que (1.14) est vraie.
Si $S = [a,b]$ avec $a > 0, \quad 1/x^2 \geq \mu \quad \forall x \in [a,b]$ alors, $\mu \leq \inf_{a \leq x \leq b} 1/x^2 = 1/b^2$.

■

Définition 1.9 (Fonction propre)
Une fonction convexe $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *est dite propre si elle n'est pas identiquement égale à* $+\infty$ $(f \not\equiv +\infty)$ *et* $f(x) > -\infty$ *pour tout* $x \in \mathbb{R}^n$*. Autrement dit, si* $epi(f) \neq \emptyset$ *et ne contient pas de droite verticale .*

Définition 1.10
Soit $f : S \to \mathbb{R}$ *une fonction. On dit que* f *est semi-continue inférieurement (s.c.i), en un point* $x_0 \in \operatorname{dom}(f)$ *si*

$$\liminf_{x \to x_0} f(x) \geq f(x_0),$$

ou, en utilisant le langage des suites : pour toute suite (x_n) *convergeant vers* x_0 *on a*

$$\liminf_{n \to +\infty} f(x_n) \geq f(x_0).$$

Remarque 1.2

⋆ *Toute fonction continue sur* $\mathbb{R}^n$ *est semi-continue inférieurement.*
⋆ *L'ensemble des fonctions convexes, s.c.i et propres est noté* $\Gamma_0(\mathbb{R}^n)$*.*

Théorème 1.4 *Soit* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$*. Il y a équivalence entre*
(a) f *est s.c.i.*
(b) $\operatorname{epi}(f)$ *est un ensemble fermé de* $\mathbb{R}^{n+1}$*.*
(c) *pour tout* $\alpha \in \mathbb{R}$*, les ensembles de niveau* $lev_\alpha(f) = \{x \in \mathbb{R}^n, \quad f(x) \leq \alpha\}$ *sont fermés.*

Preuve
On montre que (a) implique (b). On considère une suite $(x_k, \alpha_k) \subset \operatorname{epi}(f)$ qui converge vers un point(x, α) alors,

$$\alpha = \lim_{k \to +\infty} \alpha_k = \liminf_{k \to +\infty} \alpha_k \geq \liminf_{k \to +\infty} f(x_k) \geq f(x).$$

donc, $(x, \alpha) \in \operatorname{epi}(f)$. Ensuite, on montre que (b) implique (c). Soit $\alpha \in \mathbb{R}$ tel que $lev_\alpha(f)$ soit non vide et soit $(x_k) \subset lev_\alpha(f)$ une suite qui converge vers un point x. Alors,

$$\forall k \in \mathbb{N}^*, \quad (x_k, \alpha) \in \operatorname{epi}(f).$$

la condition (b) implique que $(x, \alpha) \in \operatorname{epi}(f)$, c'est à dire que $x \in lev_\alpha(f)$. Enfin, on montre que (c) implique (a). Supposons donc qu'il existe un point x tel que f ne soit pas s.c.i en x. Alors, il existe une suite (x_k) telle que

$$x_k \to x \quad \text{lorsque} \quad k \to +\infty \quad \text{et} \quad f(x) > \liminf_{k \to +\infty} f(x_k)$$

donc, il existe une sous-suite (x_{k_l}) et un réel β tels que

$$x_{k_l} \to x \quad \text{lorsque } l \to +\infty \text{ et } f(x_{k_l}) \to \beta < f(x).$$

Soit $\alpha \in]\beta, f(x)[$. pour l suffisamment grand, posons $l \geq L, f(x_{k_l}) \leq \alpha < f(x)$, donc $lev_\alpha(f)$ contient la suite convergente (x_{k_l}) et non sa limite. ■

Propriétés 1.1

Soient f_1 et f_2 deux fonctions convexes s.c.i et $\alpha > 0$. Les fonctions suivantes sont convexes et semi-continue inférieurement :

1) $f(x) = \alpha f_1(x)$.

2) $f(x) = (f_1 + f_2)(x)$ *et* $\mathrm{dom}(f) = \mathrm{dom}(f_1) \cap \mathrm{dom}(f_2)$.

3) $f(x) = \max\{f_1(x), f_2(x)\}$ *et* $\mathrm{dom}(f) = \mathrm{dom}(f_1) \cap \mathrm{dom}(f_2)$.

1.1.4 Sous-différentiel de fonctions convexes

Le sous-différentiel joue un rôle fondamental dans l'optimisation non différentiable. On s'intéresse dans ce qui suit à l'étude des propriétés de cette application ainsi que la relation qui l'a lie à la transformée de Fenchel.

Définition 1.11 (Sous-gradient et sous-différentiel)

Soit $f : \mathbb{R}^n \to]-\infty, +\infty]$ une fonction convexe et propre, un vecteur $\eta \in \mathbb{R}^n$ est appelé sous gradient de f au point $x_0 \in \mathrm{dom}(f)$ si

$$f(x) \geq f(x_0) + \langle \eta, x - x_0 \rangle \quad \text{pour tout } x \in \mathbb{R}^n. \tag{1.15}$$

L'ensemble de tous les sous gradients de f en x_0 est appelé sous différentiel de f en x_0. Il est noté par $\partial f(x_0)$.

Interprétation géométrique : *La fonction affine définie par $\phi(x) = f(x_0) + \langle \eta, x - x_0 \rangle$ est majorée par f. Le sous-différentiel $\partial f(x_0)$ est formé par toutes les directions des hyperplans d'appui qui passent par le point $(x_0, f(x_0))$ et restent sous le graphe de la fonction f.*

Définition 1.12 *On dit que f est sous-différentiable en x_0 si $\partial f(x_0) \neq \varnothing$.*

Exemple 1.6

- ⋆ *Toute fonction convexe différentiable est sous-différentiable et on a $\partial f(x_0) = \{\nabla f(x)\}$. Si $f = (\frac{1}{2})\|\cdot\|^2$. Alors $\partial f(x) = Id(x) = \{x\}$.*
- ⋆ *Soit $A \subset \mathbb{R}^n$ un ensemble convexe et fermé, on définit la fonction indicatrice de A comme*

$$\chi_A(x) = \begin{cases} 0 & \text{si} \quad x \in A \\ +\infty & \text{sinon.} \end{cases}$$

Le sous-différentiel de χ_A au point x est le cone normal de A en x : $\partial \chi_A(x) = N_A(x)$ où $N_A(x_0) = \{\eta \in \mathbb{R}^n, \langle \eta, x - x_0 \rangle \leq 0 \quad \forall x \in A\}$.

- ⋆ *Soit $f : x \in \mathbb{R} \mapsto |x|$, on calcule les sous-gradients de f en tout point $x \in \mathbb{R}$ (voir Figure 1.3)*

$$\partial f(x) = \begin{cases} \{-1\} & \text{si} \quad x < 0 \\ \{+1\} & \text{si} \quad x > 0 \\ [-1, +1] & \text{si} \quad x = 0. \end{cases}$$

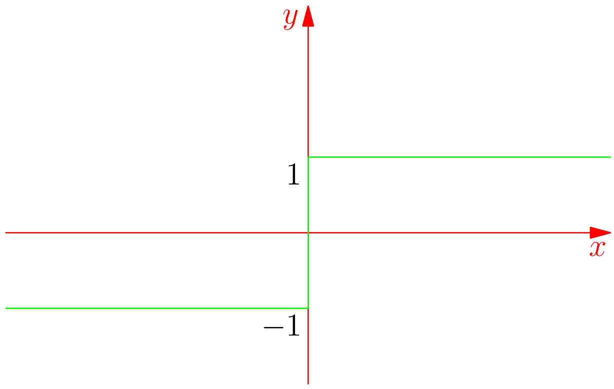


FIGURE 1.3 – Sous différentiel de la fonction $x \mapsto |x|$.

Proposition 1.5
Le sous-différentiel $\partial f(x_0)$ *est un ensemble convexe et fermé.*

Preuve
Soient g_1 et g_2 deux éléments de $\partial f(x_0)$. On a $\forall y \in \mathbb{R}^n$:

$$f(y) \geq f(x_0) + \langle g_1, y - x_0 \rangle$$

$$f(y) \geq f(x_0) + \langle g_2, y - x_0 \rangle.$$

Soit $\alpha \in [0,1]$. En multipliant la première inégalité par α et la deuxième par $(1-\alpha)$ et en sommant, on obtient :

$$f(y) \geq f(x_0) + \langle \alpha g_1 + (1-\alpha) g_2, y - x_0 \rangle.$$

Alors, $\alpha g_1 + (1-\alpha) g_2 \in \partial f(x_0)$. Reste à montrer que $\partial f(x_0)$ est fermé. Soit $(\xi_k)_{k \in \mathbb{N}^*} \subset \partial f(x_0)$ une suite qui converge vers un point ξ. Pour tout $y \in \mathbb{R}^n$ et tout $k \in \mathbb{N}^*$,

$$f(y) \geq f(x) + \langle \xi_k, y - x_0 \rangle.$$

puisque $\langle ., y - x_0 \rangle$ est continue, on peut passer à la limite dans l'inégalité ci-dessus, alors $\xi \in \partial f(x_0)$. ■

Définition 1.13 (Dérivée directionnelle d'une fonction convexe)
Soit $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *une fonction. Soit* $x_0 \in \mathrm{dom}(f)$ *et* d *un vecteur non nul de* $\mathbb{R}^n$*. La dérivée directionnelle de* f *en* x_0 *dans la direction* d*, qu'on note* $f'(x_0, d)$*, est donnée par la limite suivante*

$$f'(x_0, d) = \lim_{t \to 0^+} \frac{f(x_0 + td) - f(x_0)}{t}.$$

Proposition 1.6
Soit $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *une fonction convexe. Si* $x_0 \in \mathbf{int}(\mathrm{dom}(f))$ *alors* $f'(x_0, d)$ *existe pour tout* $d \in \mathbb{R}^n$.

Exemple 1.7 $f(x) = \begin{cases} +\infty & si \quad x < 0 \\ -\sqrt{x} & si \quad x \geq 0. \end{cases}$ *où* $f : \mathbb{R} \to \mathbb{R}$ *est une fonction convexe tel que*

$\mathrm{dom}(f) = [0, +\infty[$ *et* $\mathbf{int}(\mathrm{dom}(f)) =]0, +\infty[$.
On prend $x_0 = 0, d = 1$ *donc*

$$\lim_{t\to 0^+} \frac{f(x_0+td)-f(x_0)}{t} = \lim_{t\to 0^+} \frac{f(t)-f(0)}{t} = \lim_{t\to 0^+} \frac{-\sqrt{t}}{t} = -\infty.$$

Proposition 1.7
Soit $S \subset \mathbb{R}^n$ *un ensemble convexe d'intérieur non vide et* $f : S \to \mathbb{R}$ *une fonction convexe, alors pour tout* $x_0 \in \mathbf{int}(\mathrm{dom}(f))$ *on a*

$$f'(x_0, d) = \max\{\langle \eta, d\rangle \mid \eta \in \partial f(x_0)\}, \quad \forall d \in \mathbb{R}^n.$$

Proposition 1.8
Soit $f : \mathbb{R}^n \to]-\infty, +\infty]$ *une fonction convexe et propre, soit* $x \in \mathrm{dom}(f)$ *alors*

i) $\mathrm{dom}(\partial f) \subset \mathrm{dom}(f)$.
ii) on suppose que $x \in \mathrm{dom}(\partial f)$, *alors* f *est semi continue inférieurement.*

Preuve

i) Puisque f est propre, $f(x) = +\infty \Rightarrow \partial f(x) = \varnothing$.
ii) On prend $u \in \partial f(x)$, donc $f(y) \geq f(x) + \langle u, y-x\rangle$ pour tout $y \in \mathbb{R}^n$. D'où

$$\liminf_{y\to x} f(y) \geq \liminf_{y\to x} f(x) + \langle u, y-x\rangle = f(x).$$

■

Remarque 1.3
Le sous-différentiel peut ne pas exister sur le bord du domaine car si $x_0 \in \mathbf{int}(\mathrm{dom}(f))$ *Alors* $\partial f(x_0) \neq \emptyset$, *par exemple la fonction* $f(x) = -\sqrt{x}$ *sur* $\mathbb{R}^+$ *est convexe et fermée, mais le sous-différentiel n'existe pas en* 0 *car si*

$$u \in \partial f(0) \quad f(x) \geq f(0) + \langle u, x-0\rangle \quad \forall x.$$

Donc, $f(x) \geq ux \quad x \in \mathbb{R} \Rightarrow -\sqrt{x} \geq ux \quad \forall x > 0 \Rightarrow -\frac{1}{\sqrt{x}} \geq u$. *Alors, si* $x \to 0^+, u \leq -\infty$ *impossible.*

Théorème 1.5
Soit $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *une fonction convexe et s.c.i. Le point* x_0 *est un minimum de* f *sur* $\mathbb{R}^n$ *du problème* $f(x_0) = \min_{x\in\mathbb{R}^n} f(x)$ *si et seulement si* $0 \in \partial f(x_0)$.

Preuve
Si $0 \in \partial f(x_0)$, alors $f(x) \geq f(x_0) + \langle 0, x-x_0\rangle \geq f(x_0)$ pour tout $x \in \mathrm{dom}(f)$ donc x_0 est un minimum de la fonction f. Réciproquement, si $f(x) \geq f(x_0)$ et d'après l'inégalité (1.5) on obtient $0 \in \partial f(x_0)$. ■

1.1.5 Transformée de Fenchel ou fonctions conjugués

La fonction conjuguée, aussi appelée transformée de Fenchel ou transformée de Legendre-Fenchel est utilisée par exemple pour calculer le sous-différentiel d'une fonction convexe. Elle a été introduite par Mandelbrojt en 1939 puis précisée par Fenchel en 1949, cette transformée généralise la transformation de Legendre(1787).

Définition 1.14 (Conjuguée Convexe)
Soit $f : \mathbb{R}^n \to [-\infty, +\infty]$ *une fonction. Sa conjuguée* $f^* : \mathbb{R}^n \to [-\infty, +\infty]$ *est définie par :*

$$f^*(u) = \sup_{x \in \mathbb{R}^n} \{\langle u, x\rangle - f(x)\}. \tag{1.16}$$

C'est l'enveloppe supérieure de la famille des fonctions affines

$$l_x(u) = \langle u, x\rangle - f(x).$$

La bi-conjuguée de f *est la conjuguée convexe de* f^**, définie par :*

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

Exemple 1.8

1) Soit $f : x \mapsto \begin{cases} \frac{1}{x} & si \quad x > 0 \\ +\infty & sinon \end{cases} \Rightarrow f^* : u \mapsto \begin{cases} -2\sqrt{-u} & si \quad u \leq 0 \\ +\infty & si \quad u > 0. \end{cases}$

Puisque $f^*(u) = \sup_{x\in\mathbb{R}}\{xu - f(x)\}$*. Donc, on calcule*

$$(xu - f(x)) = \begin{cases} xu - \frac{1}{x} & si \quad x > 0 \\ -\infty & si \quad x \leq 0. \end{cases}$$

$\triangleright$ *Si* $u = 0, \sup_{x>0}(xu - \frac{1}{x}) = \sup_{x>0} -\frac{1}{x} = 0.$

$\triangleright$ *Si* $u < 0, f(x) = ux - \frac{1}{x}$ *alors* $f'(x) = u + \frac{1}{x^2} = 0 \Leftrightarrow \frac{1}{x^2} = -u \Leftrightarrow \frac{1}{x} = \sqrt{-u}$ *alors* $x = -\frac{\sqrt{-u}}{u}$.

Maintenant, on remplace x *dans* $(xu - \frac{1}{x})$*, alors on obtient*

$$f^*(u) = \begin{cases} -2\sqrt{-u} & si \quad u \leq 0 \\ +\infty & si \quad u > 0. \end{cases}$$

2) Si $f : x \mapsto \sqrt{1 + x^2}$ *alors* $f^* : u \mapsto \begin{cases} -\sqrt{1 - u^2} & si \quad |\, u \,| \leqslant 1 \\ +\infty & si \quad |\, u \,| > 1 \end{cases}$

3) Soit $p \in]1, +\infty[$ *et* q *le conjuquée de* p *(i.e* $\frac{1}{p} + \frac{1}{q} = 1$*), Alors* $(\frac{1}{p} \mid \cdot \mid^p)^* = \frac{1}{q} \mid \cdot \mid^q$ *.*

Proposition 1.9

1. *L'épigraphe de f^* est l'intersection des épigraphes des fonctions l_x, l'ensemble epi(f^*) est donc convexe fermée.*

2. *Si $f \in \Gamma_0(\mathbb{R}^n)$ alors $f^* \in \Gamma_0(\mathbb{R}^n)$.*

3. *Si f est paire alors f^* est paire.*

Théorème 1.6 (Théorème de Moreau-Fenchel)
Soit $f : \mathbb{R}^n \to]-\infty, +\infty]$ une fonction propre. f est convexe et s.c.i si et seulement si $f^{**} = f$.

Preuve
Voir [1] ■

Proposition 1.10 *Soient $f, g : \mathbb{R}^n \to [-\infty, +\infty]$ deux fonctions, alors on a*

- ▷ $f^{**} \leq f$.
- ▷ $f \leq g \Rightarrow f^* \geq g^*$ *et* $f^{**} \leq g^{**}$.
- ▷ $f^{***} = f^*$.

Preuve
Voir [1] ■

Théorème 1.7 *Soit $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe et propre. Il y a équivalence entre*

(a) $u \in \partial f(x)$.

(b) $f(x) + f^*(u) = \langle u, x \rangle$.
Si de plus $f^{**} = f$, les conditions (a) et (b) sont équivalentes à

(c) $x \in \partial f^*(u)$.

Preuve
On montre d'abord l'équivalence entre (a) et (b), on utilise la définition de sous gradient.

$$\begin{aligned}
u \in \partial f(x) \quad &\Leftrightarrow \quad \forall z \in \mathbb{R}^n, \quad f(z) \geq f(x) + \langle u, z - x \rangle \\
&\Leftrightarrow \quad \forall z \in \mathbb{R}^n, \quad \langle u, x \rangle - f(x) \geq \langle u, z \rangle - f(z) \\
&\Leftrightarrow \quad \langle u, x \rangle - f(x) \geq \sup_{z \in \mathbb{R}^n} \{\langle u, z \rangle - f(z)\} = f^*(u) \quad (\text{ on a toujours } \langle u, x \rangle \leq f(x) + f^*(u)). \\
&\Leftrightarrow \quad f(x) + f^*(u) = \langle u, x \rangle.
\end{aligned}$$

De plus, on montre l'équivalence entre (a) et (c).

$$\begin{aligned}
u \in \partial f(x) \quad &\Leftrightarrow \quad f^*(u) + f(x) = \langle u, x \rangle. \\
&\Leftrightarrow \quad f^*(u) + f^{**}(x) = \langle u, x \rangle. \quad (\text{car } f^{**} = f) \\
&\Leftrightarrow \quad x \in \partial f^*(u).
\end{aligned}$$

■

Proposition 1.11 *Soit $f : \mathbb{R}^n \to [-\infty, +\infty]$, on a les propriétés suivantes :*

1) $f^*(0) = -\inf f(\mathbb{R}^n)$.

2) $-\infty \in f^*(\mathbb{R}^n) \Leftrightarrow f \equiv +\infty \Leftrightarrow f^* \equiv -\infty$

3) Si f^* est propre, alors f est propre.

4) Soit $u \in \mathbb{R}^n$, alors

$$f^*(u) = \sup_{x \in \mathrm{dom} f} \{\langle x, u\rangle - f(x)\} = \sup_{(x,\epsilon)\in \mathrm{epi}(f)} \{\langle x, u\rangle - \epsilon\}.$$

Preuve

On montre seulement la propriétés 4), soit $f(x) \le \epsilon$ alors

$$f^*(u) = \sup_{x\in\mathbb{R}^n} \{\langle x, u\rangle - f(x)\} \ge \sup_{(x,\epsilon)\in \mathrm{epi}(f)} \{\langle x, u\rangle - \epsilon\}. \tag{1.17}$$

$$\forall \epsilon > 0, \exists x_\epsilon \in \mathrm{dom}(f) : \quad f^*(u) - \epsilon < \langle x_\epsilon, u\rangle - f(x_\epsilon).$$

Or $(x_\epsilon, f(x_\epsilon)) \in \mathrm{epi}(f)$, on obtient alors

$$\langle x_\epsilon, u\rangle - f(x_\epsilon) \le \sup_{(x,r)\in \mathrm{epi}(f)} \{\langle x, u\rangle - r\}.$$

Par suite

$$f^*(u) < \sup_{(x,r)\in \mathrm{epi}(f)} \{\langle x, u\rangle - r\} + \epsilon \quad \forall \epsilon > 0.$$

En faisant tendre ϵ vers 0^+ on obtient

$$f^*(u) \le \sup_{(x,\epsilon)\in \mathrm{epi}(f)} \{\langle x, u\rangle - \epsilon\}. \tag{1.18}$$

Le résultat s'obtient en utilisant (1.17) et (1.18). ■

Proposition 1.12 (Inégalité de Fenchel-Young) *Soit $f : \mathbb{R}^n \to]-\infty, +\infty]$ une fonction convexe et propre, alors*

$$f(x) + f^*(u) \ge \langle x, u\rangle \quad \textit{pour tout } (x, u) \in \mathbb{R}^n \times \mathbb{R}^n. \tag{1.19}$$

Preuve

Soient x et u deux éléments de $\mathbb{R}^n$. Puisque f est propre, il existe $x_0 \in \mathbb{R}^n$ tel que

$$-\infty < f(x_0) < +\infty.$$

Par suite on a pour tout $u \in \mathbb{R}^n$

$$f^*(u) = \sup_{x\in\mathbb{R}^n} \{\langle u, x\rangle - f(x)\} \ge \langle x_0, u\rangle - f(x_0) > -\infty.$$

Soit $x \in \mathbb{R}^n$, si $f(x) = +\infty$ alors $f(x) + f^*(u) = +\infty$ et l'inégalité est vérifiée. Si $f(x) \in \mathbb{R}$, alors la relation (1.19) découle directement de la définition de f^*. ■

Remarque 1.4 *Si* $f : \mathbb{R}^n \to]-\infty, +\infty[$, *alors* $f = f^* \Leftrightarrow f = \frac{1}{2}\|\cdot\|^2$.

Supposons que f *est donnée par* $f(x) = \frac{1}{2}\|u\|^2$, *on a alors*

$$f^*(u) = \sup_{x\in\mathbb{R}^n}\{\langle x, u\rangle - \frac{1}{2}\|x\|^2\}$$

La fonction $g : x \mapsto -\frac{1}{2}\|x\|^2 + \langle x, u\rangle$ *est fortement concave,* $\nabla g(x) = -x + u = 0 \Rightarrow x = u$.

$$f^*(u) = \langle u, u\rangle - \frac{1}{2}\|u\|^2 = \|u\|^2 - \frac{1}{2}\|u\|^2 = \frac{1}{2}\|u\|^2.$$

Ensuite on montre que $f = f^* \Rightarrow f(x) = \frac{1}{2}\|x\|^2$.
Si $f : \mathbb{R}^n \to [-\infty, +\infty]$ *on a* $f \equiv +\infty \Leftrightarrow f^* \equiv -\infty$ *impossible. puisque* $f = f^*$ *et*

$$\textit{s'il existe } x_0 \textit{ tel que } f(x_0) = -\infty \textit{ alors } f^* = \sup_x\{\langle x, u\rangle - f(x)\} \geq \langle x_0, u\rangle - f(x_0) = +\infty,$$

Donc, si $f : \mathbb{R}^n \to [-\infty, +\infty]$ *est telle que* $f = f^*$ *alors la fonction* f *est propre* ($f \not\equiv +\infty$).
L'inégalité de Young-Fenchel donne

$$f(x) + f^*(u) \geq \langle x, u\rangle \quad \textit{pour tout } (x, u) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Par suite si on choisit $x = u$, *on obtient*

$$2f(x) \geq \langle x, x\rangle = \|x\|^2 \Rightarrow f(x) \geq \frac{1}{2}\|x\|^2 = \phi(x). \qquad (1)$$

D'après la Proposition (1.10), *on obtient*

$$f(x) = f^*(x) \leq \phi^*(x) = \frac{1}{2}\|x\|^2. \qquad (2)$$

Alors (1) *et* (2) *donnent* $f = \phi$.

1.1.6 Règles de calcul sous-différentiel

Proposition 1.13 *Soit* f *une fonction convexe différentiable sur son domaine . Alors*

$$\forall x \in \mathbf{int}(\mathrm{dom}(f)), \partial f(x) = \{\nabla f(x)\}.$$

Preuve
Pour tout $\alpha \in [0, 1]$ et $y \in \mathbb{R}^n$, on a

$$f(x + \alpha(y - x)) \leq (1 - \alpha)f(x) + \alpha f(y).$$

Alors,

$$\langle \nabla f(x), y - x\rangle = \lim_{\alpha\to 0}\frac{f(x + \alpha(y - x)) - f(x)}{\alpha} \leq f(y) - f(x).$$

D'où $\nabla f(x) \in \partial f(x)$.

Inversement, soit $u \in \partial f(x)$ et $y \in \mathbb{R}^n$, on a

$$f(x+\alpha y) \geq f(x) + \langle u, x+\alpha y - x\rangle,$$

alors, pour tout $\alpha \in [0, +\infty[$. En effet

$$\langle \nabla f(x), y\rangle = \lim_{\alpha \to 0} \frac{f(x+\alpha y) - f(x)}{\alpha} \geq \langle u, y\rangle.$$

D'où

$$\langle \nabla f(x) - u, y\rangle \geq 0 \quad \text{pour tout } u \in \mathbb{R}^n.$$

En prenant $y = u - \nabla f(x)$, on obtient $\|u - \nabla f(x)\|^2 \leq 0$. Donc, on déduit que

$$u = \nabla f(x).$$

■

Proposition 1.14
Soit f une fonction convexe et s.c.i, soit $A : \mathbb{R}^m \to \mathbb{R}^n$ un opérateur linéaire et $b \in \mathbb{R}^n$. La fonction $\phi(x) = f(Ax+b)$ est convexe s.c.i et pour $x \in \mathbf{int}(\mathrm{dom}(\phi))$,

$$\partial \phi(x) = A^\top \partial f(Ax+b).$$

Proposition 1.15
Soit $(f_i)_{1 \leq i \leq m}$ des fonctions convexes et s.c.i. Alors la fonction

$$f(x) = \max_{1 \leq i \leq m} f_i(x)$$

est aussi convexe et s.c.i de domaine $\mathrm{dom}(f) = \cap_{i=1}^m \mathrm{dom}(f_i)$ et $\forall x \in \mathbf{int}(\mathrm{dom}(f))$, on a

$$\partial f(x) = \mathrm{conv}\{\partial f_i(x) : f(x) = f_i(x)\}.$$

Proposition 1.16 *[4]*
Soit $(f_i)_{1 \leq i \leq p}$ une famille de fonctions convexes, s.c.i et propres. On a

$$\partial(\sum_{i=1}^p f_i)(x) \supset \sum_{i=1}^p \partial f_i(x) \quad \textit{pour tout } x \in \mathbb{R}^n.$$

De plus si $\cap_{1 \leq i \leq p} \mathbf{int}(\mathrm{dom}(f_i)) \neq \emptyset$ alors on a égalité.

1.1.7 Généralités sur l'expansivité

Définition 1.15 *Soit $C \subset \mathbb{R}^n$ non vide et $A : C \to \mathbb{R}^n$.*

- $\surd$ *A est non-expansive si $\|Ax - Ay\| \leq \|x - y\| \quad (\forall x, y \in C)$.*
- $\surd$ *A est fermement non-expansive si*

$$(\forall x \in C)(\forall y \in C) \quad \langle x - y, Ax - Ay\rangle \geq \|Ax - Ay\|^2,$$

L'inégalité ci-dessus est équivalente à

$$\|Ax-Ay\|^2+\|(Id-A)x-(Id-A)y\|^2\leq\|x-y\|^2 \quad (\forall x,y\in C).$$

Puisque

$$\begin{aligned}\|Ax-Ay\|^2+\|(Id-A)x-(Id-A)y\|^2 &= \|Ax-Ay\|^2+\|x-y\|^2+\|Ax-Ay\|^2-2\langle x-y,Ax-Ay\rangle \\ &= 2\|Ax-Ay\|^2-2\langle x-y,Ax-Ay\rangle+\|x-y\|^2\leq\|x-y\|^2.\end{aligned}$$

Donc, $\|Ax-Ay\|^2\leq\langle x-y,Ax-Ay\rangle$.
On note que tout opérateur fermement non-expansif est non-expansif.

Définition 1.16 *Soient* $C\subset\mathbb{R}^n$ *non vide,* $A:C\to\mathbb{R}^n$ *et* $\beta\in]0,+\infty[$. *Un opérateur* A *est* β*-cocoercif si* βA *est fermement non expansif,i.e ;*

$$(\forall x\in C)(\forall y\in C)\quad\langle x-y,Ax-Ay\rangle\geq\beta\|Ax-Ay\|^2.$$

Définition 1.17 *Soient* $C\subset\mathbb{R}^n$ *non-vide et* $\alpha\in]0,1[$. *Un opérateur* $A:C\to\mathbb{R}^n$ *est* α*-moyenné s'il existe un opérateur* $R:C\to\mathbb{R}^n$ *est non-expansive tel que*

$$A=(1-\alpha)Id+\alpha R. \tag{1.20}$$

Ou bien, si

$$(\forall(x,y)\in C^2)\quad\|Ax-Ay\|^2+\tfrac{1-\alpha}{\alpha}\|(Id-A)x-(Id-A)y\|^2\leq\|x-y\|^2. \tag{1.21}$$

On a l'équivalence entre (1.20) et (1.21). En effet,

$$\begin{aligned}\|Rx-Ry\|^2 &= \|(\tfrac{\alpha-1}{\alpha})Idx+\tfrac{1}{\alpha}Ax-(\tfrac{\alpha-1}{\alpha})Idy-\tfrac{1}{\alpha}Ay\|^2 \\ &= \|(\tfrac{1-\alpha}{\alpha})(x-y)+\tfrac{1}{\alpha}(Ax-Ay)\|^2 \\ &= (\tfrac{1-\alpha}{\alpha})\|x-y\|^2+\tfrac{1}{\alpha}\|Ax-Ay\|^2-\tfrac{1}{\alpha}(1-\tfrac{1}{\alpha})\|Ax-Ay-(x-y)\|^2.\end{aligned} \tag{1.22}$$

En multipliant l'équation (1.22) par α, on obtient

$$\alpha(\|x-y\|^2-\|Rx-Ry\|^2)=\|x-y\|^2-\|Ax-Ay\|^2-(\tfrac{1-\alpha}{\alpha})\|(Id-A)x-(Id-A)y\|^2.$$

Puisque $\alpha(\|x-y\|^2-\|Rx-Ry\|^2)\geq 0$, alors on conclut l'inégalité (2).

Proposition 1.17
- *Un opérateur* α*-moyenné est non-expansif.*
- *Un opérateur est fermement non-expansif si et seulement si il est* $\frac{1}{2}$*-moyenné.*
- *Soient* $\lambda\in]0,1/\alpha[$ *et* A *est* α*-moyenné alors* $(1-\lambda)Id+\lambda A$ *est* $\lambda\alpha$*-moyenné.*

Preuve
Voir[1] ■

Proposition 1.18 *Soient $C \subset \mathbb{R}^n$ non vide et m un entier strictement positive. L'ensemble $I = \{1, \ldots, m\}$, soient $(T_i)_{i \in I} : C \to C$ une famille d'opérateurs et (α_i) un réel dans $]0, 1[$, tel que pour tout $i \in I, (T_i)$ est α_i-moyenné et*

$$T = T_1 \cdots T_m \qquad \alpha = \frac{m}{m - 1 + \frac{1}{\max_{i \in I} \alpha_i}}$$

Alors, T est α-moyenné.

Preuve
Voir[1] ■

Proposition 1.19 *Soient $C \subset \mathbb{R}^n$ non-vide et $\beta \in]0, +\infty[$. Un opérateur $A : C \to \mathbb{R}^n$ est β-cocoercif et $\gamma \in]0, 2\beta[$ alors,*

$$(Id - \gamma A) \textit{ est } (\tfrac{\gamma}{2\beta})\textit{moyenné.}$$

Preuve
On a βA est fermement non expansif (car A est β-cocoercif) alors βA est $\frac{1}{2}$-moyenné. Il existe donc un opérateur $R : C \to \mathbb{R}^n$ non expansif tel que $\beta A = \frac{(Id+R)}{2}$. Alors, on obtient

$$Id - \gamma A = (1 - \frac{\gamma}{2\beta})Id + \frac{\gamma}{2\beta}(-R).$$

$(-R)$ est une opérateur non expansif, en déduit que

$$(Id - \gamma A) \text{ est } (\tfrac{\gamma}{2\beta}) \text{ moyenné.}$$

■

1.2 Opérateurs monotones

Définition 1.18 (Opérateurs multivoques) *[8]*
Soient X et Y deux ensembles. Un opérateur multivoque T de X dans Y est un opérateur qui à chaque $x \in X$ associe un sous-ensemble $T(x)$ de Y. Un tel opérateur est aussi appelé multifonction, et on note : $T : X \rightrightarrows Y$ ou $T : X \to 2^Y$.
Les ensembles

$$\begin{aligned} \operatorname{graph}(T) &= \{(x, u) \in X \times Y | u \in Tx\} \\ \operatorname{dom}(T) &= \{x \in X | T(x) \neq \varnothing\} \\ \operatorname{Im}(T) &= \{y | \exists x : y \in T(x)\} = \cup_{x \in D(T)} T(x), \end{aligned}$$

sont appelés graphe, domaine et image de T respectivement.
L'inverse d'un opérateur multivoque est défini par

$$T^{-1}(y) = \{x \in X | y \in T(x)\}.$$

ou de façon équivalente

$$x \in T^{-1}(y) \Leftrightarrow (x, y) \in \operatorname{graph}(T).$$

On a les relations suivantes :

$$\operatorname{dom} T^{-1} = \operatorname{Im}(T), \quad \operatorname{Im}(T^{-1}) = \operatorname{dom} T, \quad (T^{-1})^{-1} = T.$$

1.2.1 Définition et Exemples

Définition 1.19
Soit $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ une multifonction (ou application multivoque). A est dite monotone si

$$(\forall (x_1, u_1) \in \operatorname{graph}(A))\, (\forall (x_2, u_2) \in \operatorname{graph}(A)) \qquad \langle u_1 - u_2, x_1 - x_2 \rangle \geq 0,$$

On note que pour tout $x \in \mathbb{R}^n$, $A(x)$ est une partie de $\mathbb{R}^n$.

Exemple 1.9 *Pour $n = 1$, A une application univoque est monotone si et seulement si elle est croissante car en posant $u_1 = f(x_1), u_2 = f(x_2)$*

$$\langle x_2 - x_1, u_2 - u_1 \rangle \geq 0 \Rightarrow \quad (x_2 - x_1)(f(x_2) - f(x_1)) \geq 0.$$

$$x_2 > x_1 \Rightarrow \quad f(x_2) - f(x_1) \geq 0 \Rightarrow \quad f(x_2) \geq f(x_1).$$

Proposition 1.20
Soit $f : \mathbb{R}^n \to]-\infty, +\infty]$ une fonction convexe et propre, alors ∂f est monotone.

Preuve
On prend (x, u) et (y, v) dans $\operatorname{graph}(\partial f)$. Alors, d'après la définition (1.15). on a

$$\langle u, x - y \rangle + f(y) \geq f(x). \tag{1.23}$$
$$\langle v, y - x \rangle + f(x) \geq f(y). \tag{1.24}$$

En additionnant les inégalités (1.23) et (1.24), on obtient $\langle x - y, u - v \rangle \geq 0$. ■

Proposition 1.21 *Soit $A : \mathbb{R}^n \to \mathbb{R}^n$ une application linéaire, :*

1) *A est monotone si et seulement si A est semi-défini positif (s.d.p)(i.e., si pour tout $(x \in \mathbb{R}^n) \quad \langle Ax, x \rangle \geq 0$).*
2) *A monotone $\Leftrightarrow A + A^\top$ monotone $\Leftrightarrow A^\top$ monotone.*
3) *$A^\top A, AA^\top, A - A^\top, A^\top - A$ sont monotone.*

Preuve

1) On montre que si A est monotone alors A est s.d.p.
   A est monotone donc $\langle Ax - Ay, x - y \rangle \geq 0$. En prenant $y = 0$, on obtient

$$\langle Ax, x \rangle \geq 0 \quad \forall x \in \mathbb{R} \Rightarrow A \text{ est s.d.p .}$$

Inversement, si A est s.d.p alors

$$\langle Ax - Ay, x - y\rangle = \langle A(x-y), x-y\rangle = \langle Av, v\rangle \geq 0 \quad \forall x, y \in \mathbb{R}^n,$$

A est monotone.

2)
$$\begin{aligned} A \text{ monotone} \quad &\Leftrightarrow \quad (\forall (x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^n) \qquad \langle x_1 - x_2, Ax_1 - Ax_2\rangle \geq 0. \\ &\Leftrightarrow \quad (\forall x \in \mathbb{R}^n) \qquad 2\langle x, Ax\rangle \geq 0. \\ &\Leftrightarrow \quad (\forall x \in \mathbb{R}^n) \qquad \langle x, Ax\rangle + \langle A^\top x, x\rangle \geq 0. \\ &\Leftrightarrow \quad (\forall x \in \mathbb{R}^n) \qquad \langle x, (A + A^\top)x\rangle \geq 0. \\ &\Leftrightarrow \quad A + A^\top \text{ monotone.} \end{aligned}$$

3) On montre que $A^\top A$ monotone $\forall A$;

$$\forall x \quad \langle x, (A^\top A)x\rangle = \langle Ax, Ax\rangle = \|Ax\|^2 \geq 0.$$

■

Définition 1.20 *Un opérateur $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ est strictement monotone si*

$$(\forall (x_1, u_1) \in \operatorname{graph}(A)) \quad (\forall (x_2, u_2) \in \operatorname{graph}(A)) \quad x_1 \neq x_2 \Rightarrow \langle u_1 - u_2, x_1 - x_2\rangle > 0.$$

- *Tout opérateur strictement monotone est monotone.*

Exemple 1.10

1) A est un opérateur strictement monotone si et seulement si A est définie positive.

2) Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction strictement croissante si et seulement si f est strictement monotone.

Définition 1.21 *Un opérateur $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ une multifonction est β-fortement monotone si*

$$(\forall (x_1, u_1) \in \operatorname{graph}(A))(\forall (x_2, u_2) \in \operatorname{graph}(A)) \quad \langle u_1 - u_2, x_1 - x_2\rangle \geq \beta \parallel x_1 - x_2 \parallel^2 .$$

où β un nombre réel strictement positif

Propriétés 1.2 *Soient $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ et $B : \mathbb{R}^m \to 2^{\mathbb{R}^m}$ deux opérateurs monotones. Les opérateurs suivants sont monotones :*

1) $A \times B : \mathbb{R}^n \times \mathbb{R}^m \to 2^{\mathbb{R}^n \times \mathbb{R}^m}$ $(x, y) \mapsto Ax \times Ay = \{(u, v) | u \in Ax, v \in Bx\}$.

2) $A + B : x \mapsto \{u + v \mid u \in Ax, v \in Bx\}$ si $\mathbb{R}^n = \mathbb{R}^m$.

3) $A^\top BA : x \mapsto \{A^\top v, v \in B(Ax)\}$ si $A \in L(\mathbb{R}^n, \mathbb{R}^m)$.

Proposition 1.22 *Soit $A, B : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ deux opérateurs monotones et $\gamma \in \mathbb{R}_+$.*

$$A^{-1}, \gamma A, \text{ et } A + A^\top BA$$

sont monotones.

1.3 Opérateur Monotone Maximal

Définition 1.22
L'opérateur $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ *un monotone maximal si l'une des deux conditions suivantes est vérifiée :*

(i) A est monotone et il n'existe pas d'opérateur monotone $B : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ *autre que A tell que* $\operatorname{graph}(A)$ *soit inclus dans* $\operatorname{graph}(B)$.

(ii) A est monotone et pour tout $(x_1, u_1) \in \mathbb{R}^n \times \mathbb{R}^n$,

$$(x_1, u_1) \in \operatorname{graph}(A) \quad \Leftrightarrow \quad (\forall (x_2, u_2) \in \operatorname{graph}(A)) \quad \langle x_1 - x_2 \mid u_1 - u_2 \rangle \geq 0.$$

L'équivalence des deux définitions :
$(i) \Rightarrow (ii)$
Soit $(x_1, u_1) \in \mathbb{R}^n \times \mathbb{R}^n$ tel que l'inégalité ci-dessus soit satisfaite. Soit B tel que $\operatorname{graph}(B) = \operatorname{graph}(A) \cup \{(x_1, u_1)\}$. si A est monotone, B est monotone tel que $\operatorname{graph}(A) \subset \operatorname{graph}(B)$ et d'après la condition (i), on a $B = A \Rightarrow (x_1, u_1) \in \operatorname{graph}(A)$.
$(ii) \Rightarrow (i)$
Si B monotone et $\operatorname{graph}(A) \subset \operatorname{graph}(B)$, alors

$$(\forall (x_1, u_1) \in \operatorname{graph}(B))(\forall (x_2, u_2) \in \operatorname{graph}(A)) \quad \langle x_1 - x_2 \mid u_1 - u_2 \rangle \geq 0.$$

la condition (ii) implique alors que $(x_1, u_1) \in \operatorname{graph}(A)$. D'où $B = A$.

Propriétés 1.3 *Si A et B deux opérateurs monotones maximaux de* $\mathbb{R}^n$ *dans* $2^{\mathbb{R}^n}$ *tel que l'une des propriétés suivantes soient vérifiées*

1. $\operatorname{dom} B = \mathbb{R}^n$
2. $\operatorname{dom} A \cap int(\operatorname{dom} B) \neq \varnothing$
3. $0 \in \mathbf{int}(\operatorname{dom} A - \operatorname{dom} B)$

alors $A + B$ *est monotone maximal.*

Proposition 1.23 *Si* $A : \mathbb{R}^n \to \mathbb{R}^n$ *un opérateur monotone et continue, alors A est monotone maximal.*

Preuve
Soit $(x_1, u_1) \in \mathbb{R}^2$, tel que$\langle x_1 - x_2, u_1 - Ax_2 \rangle \geq 0$ pour tout $x_2 \in \mathbb{R}^n$. En posant

$$x_2 = x_1 + \frac{1}{n}(u_1 - Ax_1).$$

On a

$$\langle u_1 - Ax_1, u_1 - Ax_2 \rangle = -n \langle x_1 - x_2, u_1 - Ax_2 \rangle \geq 0.$$

Donc,

$$\begin{aligned}\lim_{n\to+\infty}\langle u_1 - Ax_1, u_1 - Ax_2\rangle &= \langle u_1 - Ax_1, u_1 - Ax_1\rangle \\ &= \|u_1 - Ax_1\|^2 \leq 0.\end{aligned}$$

D'où $u_1 = Ax_1$. ■

Exemple 1.11 *Soient $T : \mathbb{R}^n \to \mathbb{R}^n$ une application non-expansive et $\alpha \in [-1, 1]$ alors $Id + \alpha T$ est monotone maximal.*

Preuve
Pour tout $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$, alors

$$\begin{aligned}\langle (Id + \alpha T)x - (Id + \alpha T)y, x - y\rangle &= \|x - y\|^2 - \alpha\langle Tx - Ty, x - y\rangle \\ &\geq \| x - y \|^2 - | \alpha | \| Tx - Ty \| \| x - y \| . \\ &\geq (1 - | \alpha |) \| x - y \|^2 \geq 0.\end{aligned}$$

$Id + \alpha T$ est un opérateur monotone continue sur $\mathbb{R}^n$, donc il est maximal d'après la proposition (1.23). ■

Proposition 1.24

1) *$A : \mathbb{R}^n \to \mathbb{R}^n$ est un opérateur α-moyenné avec $\alpha \in]0, \frac{1}{2}]$, alors A est monotone maximal.*
2) *Si $A : \mathbb{R}^n \to \mathbb{R}^n$ est β-cocoercif alors A est monotone maximal.*

Preuve

1) A est continue. De plus, pour tout $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$,

$$\|Ax - Ay\|^2 + \tfrac{1-\alpha}{\alpha}\|(Id - A)x - (Id - A)y\|^2 \leq \|x - y\|^2,$$

équivalent à
$$\|Ax - Ay\|^2 + \tfrac{1-\alpha}{\alpha}\|x - y - (Ax - Ay)\|^2 \leq \|x - y\|^2.$$

Alors, on obtient

$$0 \leq \|Ax - Ay\|^2 + (1 - 2\alpha)\|x - y\|^2 \leq 2(1 - \alpha)\langle x - y, Ax - Ay\rangle.$$

2) On a l'opérateur A est β-cocoercif si βA est fermement non expansif, il est monotone maximal d'après 1).

■

Proposition 1.25 *Si $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ un opérateur monotone maximal de domaine borné alors A est surjectif (i.e, $\forall u \in \mathbb{R}^n, \exists\, x \in \mathbb{R}^n \,| u \in Ax$).*

Théorème 1.8 (Minty)
Soit $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ un opérateur monotone. Alors A est monotone maximal si et seulement si

$$\mathrm{Im}(Id + A) = \mathbb{R}^n.$$

Preuve
Premièrement, on suppose que $\mathrm{Im}(Id + A) = \mathbb{R}^n$ et fixons $(x, u) \in \mathbb{R}^n \times \mathbb{R}^n$, tel que

$$(\forall (y, v) \in \mathrm{graph}(A)) \quad \langle x - y, u - v\rangle \geq 0. \tag{1.25}$$

Puisque $\mathrm{Im}(Id + A) = \mathbb{R}^n$, il existe $(y, v) \in \mathbb{R}^n$ tel que

$$v \in Ay \text{ et } x + u = y + v \in (Id + A)y. \tag{1.26}$$

D'après (1.25) et (1.26) on a,

$$0 \leq \langle y - x, v - u\rangle = \langle y - x, x - y\rangle = -\|y - x\|^2 \leq 0.$$

D'où $y = x$ et $v = u$ alors, $(x, u) = (y, v) \in \mathrm{graph}(A)$ et donc A est monotone maximal.
Inversement, d'après la propriété de la fonction de Fitzpatrick (i.e, $F_A(x, u) \geq \langle x, u\rangle$) d'un opérateur monotone définie par

$$\begin{array}{rcl} F_A : \mathbb{R}^n \times \mathbb{R}^n & \to & [-\infty, +\infty] \\ (x, u) & \longmapsto & \sup\limits_{(y,v)\in \mathrm{graph}(A)} (\langle y, u\rangle + \langle x, v\rangle - \langle y, v\rangle) \\ & = & \langle x, u\rangle - \inf\limits_{(y,v)\in \mathrm{graph}(A)} \langle x - y, u - v\rangle. \end{array}$$

Donc,

$$\begin{array}{rcl} 2F_A(x, u) + \|(x, u)\|^2 & = & 2F_A(x, u) + \|x\|^2 + \|u\|^2 \\ & \geq & 2\langle x, u\rangle + \|x\|^2 + \|u\|^2 \\ & = & \|x + u\|^2 \\ & \geq & 0. \end{array}$$

Par conséquent, il existe un vecteur $(v, y) \in \mathbb{R}^n \times \mathbb{R}^n$ tel que

$$(\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n), \quad F_A(x, u) + \frac{1}{2}\|(x, u)\|^2 \geq \frac{1}{2}\|(x, u) + (v, y)\|^2.$$

Alors

$$\begin{array}{rrcl} (\forall (x, u) \in \mathbb{R}^n \times \mathbb{R}^n) & F_A(x, u) & \geq & \frac{1}{2}\|v\|^2 + \langle x, v\rangle + \frac{1}{2}\|y\|^2 + \langle y, u\rangle \\ & & \geq & -\langle y, v\rangle + \langle x, v\rangle + \langle y, u\rangle \\ (\forall (x, u) \in \mathrm{graph}(A)) & \langle x, u\rangle & \geq & \frac{1}{2}\|v\|^2 + \langle x, v\rangle + \frac{1}{2}\|y\|^2 + \langle y, u\rangle \\ & & \geq & -\langle y, v\rangle + \langle x, v\rangle + \langle y, u\rangle. \end{array}$$

On a A est un opérateur monotone maximal donc en déduit que $v \in Ay$, alors on obtient

$$\begin{array}{rcl} 2\langle y, v\rangle & \geq & \|v\|^2 + 2\langle y, v\rangle + \|y\|^2 + 2\langle y, v\rangle \\ 0 & \geq & \|v\|^2 + 2\langle y, v\rangle + \|y\|^2 = \|y+v\|^2 \Rightarrow -v = y. \end{array}$$

On conclut que, $0 \in (Id + A)y \subset Im(Id + A)$. Maintenant, on fixe $w \in \mathbb{R}^n$ et on définit un opérateur monotone maximal $B : \mathbb{R}^n \to 2^{\mathbb{R}^n} : x \mapsto -w + Ax$. Alors, le raisonnement ci-dessus montre que $0 \in Im(Id + B)$ et par conséquent $w \in Im(Id + A)$. ■

Proposition 1.26 *Soit $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ un opérateur monotone maximal et pour tout $x \in \mathbb{R}^n$ alors, Ax est convexe et fermé.*

Preuve
On pose $x \in \mathrm{dom}A$,

$$Ax = \cap_{(y,v)\in \mathrm{graph}(A)}\{u \in \mathbb{R}^n | \langle x - y, u - v\rangle \geq 0\}.$$

Donc, Ax est l'intersection de convexes fermés. ■

Théorème 1.9 (Moreau) *Soit $f : \mathbb{R}^n \to]-\infty, +\infty]$ une fonction convexe, propre et semi-continu inférieurement, alors ∂f est monotone maximal.*

Preuve
Voir [1] ■

1.3.1 Résolvante d'un opérateur monotone

Définition 1.23 *Soit $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ une application et $\gamma \in \mathbb{R}_{++}$. La résolvante de A est l'application multivoque $J_A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ définie par*

$$J_A = (Id + A)^{-1}. \tag{1.27}$$

L'approximation de Yosida de A d'indice γ est définie par

$$^{\gamma}A = \frac{1}{\gamma}(Id - J_{\gamma A}).$$

Proposition 1.27 *Soit $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ un opérateur monotone. Alors,*

- *$J_{\gamma A}$ et $Id - J_{\gamma A}$ sont fermement non expansive et monotones maximales.*
- *$^{\gamma}A : \mathbb{R}^n \to \mathbb{R}^n$ est γ-cocoercive, monotone maximal et γ^{-1}-Lipschitz continuos.*

Preuve
Voir [1] ■

Corollaire 1 *La résolvante de A est $\frac{1}{2}$-moyenné d'après les Propositions (1.27) et (1.17).*

Proposition 1.28 *Soient* $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ *un opérateur monotone maximale et* $\gamma \in]0, +\infty[$. *Pour tout* $x \in \mathbb{R}^n$, *il existe un unique* $p \in \mathbb{R}^n$ *tel que* $x - p \in \gamma Ap$ *et on a* $p = J_{\gamma A} x$.

Proposition 1.29 *Soient* $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}, \gamma \in \mathbb{R}_{++}$, *et* $p \in \mathbb{R}^n$. *Alors*

a) $\mathrm{dom}(J_{\gamma A}) = \mathrm{dom}^{\gamma} A = \mathrm{Im}(Id + \gamma A)$ *et* $\mathrm{Im}(J_{\gamma A}) = \mathrm{dom}(A)$.

b) $p \in J_{\gamma A} x \Leftrightarrow x \in p + \gamma Ap \Leftrightarrow x - p \in \gamma Ap \Leftrightarrow (p, \gamma^{-1}(x - p)) \in \mathrm{graph}(A)$.

c) $p \in^{\gamma} Ax \Leftrightarrow p \in A(x - \gamma p) \Leftrightarrow (x - \gamma p, p) \in \mathrm{graph}(A)$.

Proposition 1.30 (Inverse d'un résolvant) *L'opérateur* $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ *est monotone maximal et* γ *un réel strictement positif, alors*

$$J_{\gamma A^{-1}} = (Id - \gamma J_{\gamma^{-1} A} \circ \gamma^{-1} Id).$$

En particulier,

$$J_{A^{-1}} = Id - J_A.$$

Preuve

Pour tout $x \in \mathbb{R}^n$,

$$\begin{aligned}
p = J_{\gamma A^{-1}} x \quad &\Leftrightarrow \quad x \in (Id + \gamma A^{-1})(p) \\
&\Leftrightarrow \quad \gamma^{-1}(x - p) \in A^{-1} p \\
&\Leftrightarrow \quad p \in A(\gamma^{-1}(x - p)) \\
&\Leftrightarrow \quad \gamma^{-1} p \in \gamma^{-1} A(\gamma^{-1}(x - p)) \\
&\Leftrightarrow \quad \gamma^{-1} x \in (Id + \gamma^{-1} A)(\gamma^{-1}(x - p)) \\
&\Leftrightarrow \quad \gamma^{-1}(x - p) = J_{\gamma^{-1} A}(\gamma^{-1} x) \\
&\Leftrightarrow \quad p = x - \gamma J_{\gamma^{-1} A}(\gamma^{-1} x).
\end{aligned}$$

■

Proposition 1.31 (Calcul de la résolvante)

Soit $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ *un opérateur monotone maximal et* $x \in \mathbb{R}^n$.

1) Soit $B : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ *l'opérateur définie par* $B(z) = A(x' - z)$ *pour tout* $z \in \mathbb{R}^n$, *on a alors* $J_B(z) = z + J_A(x' - z)$.

2) Soit $B(z) = x' + A(z)$. *On a* $J_B(z) = J_A(x' - z)$.

3) Soit $\alpha \in [0, +\infty[$ *et* $B = A + \alpha Id$. *On a* $J_B = J_{(1+\alpha)^{-1} A} \circ [(1 + \alpha)^{-1} Id]$.

Preuve

Soit x et p dans $\mathbb{R}^n$,

1) $p = J_B x \Leftrightarrow x - p \in A(p - z) \Leftrightarrow (x - z) - (p - z) \in A(p - z) \Leftrightarrow p - z = J_A(x - z)$.

2) $p = J_B x \Leftrightarrow x - p \in z + Ap \Leftrightarrow (x - z) - p \in Ap \Leftrightarrow p = J_A(x - z)$.

3) $p = J_B x \Leftrightarrow x - p \in Ap + \alpha p \Leftrightarrow x - (1 + \alpha) p \in Ap \Leftrightarrow (1 + \alpha)^{-1} x - p \in (1 + \alpha)^{-1} Ap \Leftrightarrow p = J_{(1+\alpha)^{-1} A} \circ [(1 + \alpha)^{-1} x]$.

■

Définition 1.24 *Soit $B : \mathbb{R}^n \to 2^{\mathbb{R}^n}$. L'ensemble des points fixes de B est défini*

$$\text{Fix}(B) = \{x \in \mathbb{R}^n \mid x \in Bx\}.$$

Définition 1.25 *Soit A un opérateur monotone et γ un réel strictement positif, on définit la résolvante réfléchie par*

$$R_A = 2J_A - Id.$$

Théorème 1.10

Soit D un sous ensemble non vide dans $\mathbb{R}^n, T : D \to \mathbb{R}^n$ une application. Notons $A = T^{-1} - Id$. Alors

1) T est fermement non expansive si et seulement si $R = (2T - Id)$ est non expansive.

2) $T = J_A$.

3) T est fermement non expansive si et seulement si A est monotone.

4) T est fermement non expansive et $D = \mathbb{R}^n$ si et seulement si A est monotone maximal.

Preuve

- On montre que T est fermement non expansive si et seulement si $R = (2T - Id)$ est non expansive. Il est facile de vérifie que

$$\|Rx - Ry\|^2 = \|2Tx - 2Ty\|^2 + \|x - y\|^2 - 2\langle Tx - Ty, x - y\rangle. \tag{1.28}$$

si T est fermement non expansive alors la relation (1.28) entraine,

$$\begin{aligned}\|Rx - Ry\|^2 &\leq 2\langle x - y, Tx - Ty\rangle + \|x - y\|^2 - 2\langle Tx - Ty, x - y\rangle \\ &\leq \|x - y\|^2.\end{aligned}$$

Si on suppose que R est non expansive alors la relation (1.28) nous donne,

$$2\|Tx - Ty\|^2 + \|x - y\|^2 - 2\langle Tx - Ty, x - y\rangle \leq \|x - y\|^2,$$

on obtient donc,

$$\|Tx - Ty\|^2 \leq \langle x - y, Tx - Ty\rangle.$$

- Pour montrer 2) on utilise (1.27), alors

$$J_A = (Id + A)^{-1} = (Id + T^{-1} - Id)^{-1} = T.$$

- Troisièmement, on suppose que T est fermement non expansive et on prend (x, u) et (y, v) deux éléments de graph(A), alors on obtient

$$x = T(x + u) \text{ et } y = T(y + v).$$

Par conséquent,

$$\langle x - y, u - v\rangle = \langle T(x + u) - T(y + v), (Id - T)(x + u) - (Id - T)(y + v)\rangle \geq 0.$$

qui prouve la monotonie de A.
Maintenant, si A est monotone et x et $y \in D$ alors $x - Tx \in A(Tx)$ et $y - Ty \in A(Ty)$ donc

$$\langle Tx - Ty, (x - Tx) - (y - Ty)\rangle \geq 0 \text{ implique que } T \text{ est fermement non expansive.}$$

- Dernièrement, d'après 3) et le Théorème 1.8 on a A est monotone maximal si et seulement si
$$\mathbb{R}^n = \operatorname{Im}(Id + A) = \operatorname{Im}(T^{-1}) = \operatorname{dom} T = D.$$

■

1.3.2 Zéros d'un opérateur monotone

Définition 1.26 *Soit $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ un opérateur monotone. L'ensemble des zéros de A, noté* $\operatorname{zer}(A)$, *est*
$$\operatorname{zer}(A) = A^{-1}0 = \{x \in \mathbb{R}^n \mid 0 \in Ax\}.$$

Proposition 1.32 *Soit $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ un opérateur maximal monotone. Si l'une des hypothèses suivantes est vérifiée*

a. A est surjectif.

b. $\operatorname{dom}(A)$ *est borné,*

alors $\operatorname{zer}(A) \neq \emptyset$.

Propriétés 1.4 *Soit $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ un opérateur maximal monotone. alors* $\operatorname{zer}(A)$ *convexe et fermé.*

Preuve
On a si A un opérateur monotone maximal alors Ax est convexe et fermée et $\operatorname{zer}(A) = A^{-1}0$ est un convexe fermé (car A^{-1} est monotone maximal). ■

Proposition 1.33 *Soient $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ un opérateur monotone et $\gamma \in]0, +\infty[$. On a*

$$\operatorname{Fix}(J_{\gamma_A}) = \operatorname{zer}(A).$$

Preuve

$$\begin{aligned} (\forall x \in \mathbb{R}^n) \quad 0 \in Ax &\Leftrightarrow 0 \in \gamma Ax \\ &\Leftrightarrow x \in (Id + \gamma A)x \\ &\Leftrightarrow x \in (Id + \gamma A)^{-1}x \\ &\Leftrightarrow x = J_{\gamma_{\mathrm{A}}}x. \end{aligned}$$

■

Proposition 1.34 *Soit $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ un opérateur strictement monotone.* $\operatorname{zer}(A)$ *contient au plus un élément.*

Preuve
$\operatorname{zer}(A)$ contient deux point distinct qu'on note x et y. Alors, $0 \in Ax, 0 \in Ay$, puisque A est strictement monotone. On obtient,

$$0 = \langle x - y, 0 - 0 \rangle > 0.$$

Ce qui est impossible. ■

Proposition 1.35 *Soient $A, B : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ deux opérateurs et $\gamma \in \mathbb{R}_{++}$. Alors*
1) $\operatorname{zer}(A+B) = \operatorname{dom}(A \cap (-B))$.
2) On suppose que A est monotone et B est au plus un élément alors,

$$\operatorname{zer}(A+B) = \operatorname{Fix}(J_{\gamma_A} \circ (Id - \gamma B)).$$

Preuve

1)

$$\begin{aligned} 0 \in Ax + Bx \quad &\Leftrightarrow \quad \{(\exists u \in Bx) - u \in Ax\} \\ &\Leftrightarrow \quad Ax \cap (-Bx) \neq \varnothing. \end{aligned}$$

2)

$$\begin{aligned} 0 \in Ax + Bx \quad &\Leftrightarrow \quad -Bx \in Ax \\ &\Leftrightarrow \quad x - \gamma Bx \in x + \gamma Ax \\ &\Leftrightarrow \quad x \in ((Id + \gamma A)^{-1} \circ (Id - \gamma B))x \\ &\Leftrightarrow \quad x = (J_{\gamma_{\mathrm{A}}} \circ (Id - \gamma B))x. \end{aligned}$$

■

Définition 1.27
Une fonction différentiable $f : \mathbb{R}^n \to]-\infty, +\infty]$ est dite de gradient L-Lipschitz sur un sous-ensemble C de $\mathbb{R}^n$, pour $L > 0$, si

$$(\forall (x, y) \in C^2) \quad \|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|. \tag{1.29}$$

Proposition 1.36 *[7]*
Soit $f : \mathbb{R}^n \to]-\infty, +\infty]$ une fonction propre telle que, pour tout $x \in \mathbb{R}^n, f(x) = g(x) + h(x)$, où g est une fonction propre différentiable sur $\operatorname{dom} f$ *et h est une fonction propre. On a alors*

$$(\forall x \in \operatorname{dom} f) \quad \partial f(x) = \nabla g(x) + \partial h(x).$$

Lemme 1.1 (Lemme de descente) *Soit $f : \mathbb{R}^n \to]-\infty, +\infty]$ une fonction différentiable à gradient L-Lipschitz, elle satisfait*

$$(\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n) \quad f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|y - x\|^2.$$

Preuve
Pour tout $(x, y) \in \mathbb{R}^{n2}$ et $t \in \mathbb{R}$, soit $\phi(t) = f(x + t(y - x)).\phi$ est différentiable et

$$\phi'(t) = \langle \nabla f(x + t(y - x)), y - x \rangle.$$

On a alors

$$\phi(1) - \phi(0) = \int_0^1 \phi'(t)dt$$

.

$$\begin{aligned} f(y) &= f(x) + \int_{t=0}^1 \langle \nabla f(x + t(y - x)), y - x \rangle dt. \\ &= f(x) + \int_{t=0}^1 \langle \nabla f(x + t(y - x)) - \nabla f(x) + \nabla f(x), y - x \rangle dt. \\ &= f(x) + \langle \nabla f(x), y - x \rangle + \int_{t=0}^1 \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x \rangle dt. \end{aligned}$$

d'après l'inégalité de Cauchy Schwartz et (1.29),

$$\begin{aligned} f(y) &\leq f(x) + \langle \nabla f(x), y - x \rangle + \int_{t=0}^1 tL \parallel y - x \parallel^2 dt. \\ &= f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \parallel y - x \parallel^2 . \end{aligned}$$

■

Définition 1.28 *L'opérateur $T : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ (un opérateur multivoque) est dit Lipschitz continu sur V (ensemble ouvert $V \subset \mathbb{R}^n$) s'il est continu et si de plus, il existe une constante $L > 0$ telle que pour $x_1, x_2 \in V$ et $y_1 \in T(x_1)$, il existe $y_2 \in T(x_2)$ satisfait*

$$\parallel y_1 - y_2 \parallel \leq L \parallel x_1 - x_2 \parallel .$$

1.3.3 Suites Fejér-monotones

Définition 1.29 *Soit (x_n) une suite d'éléments de $\mathbb{R}^n$ et $D \subset \mathbb{R}^n$. (x_n) est une suite Fejér-monotone par rapport à D si*

$$(\forall x \in D)(\forall n \in \mathbb{N}) \quad \|x_{n+1} - x\| \leq \|x_n - x\|.$$

$$(\forall x \in D), \ \textit{la suite } (\|x_n - x\|) \textit{ est décroissante .}$$

Exemple 1.12

1) Toute suite (x_n) croissante et majorée dans $\mathbb{R}$ de limite ℓ est Fejér-monotone par rapport $[\ell, +\infty[$ (voir Figure 1.4).

2) Soit D un sous ensemble non vide de $\mathbb{R}^n$ et $T : D \to D$ un opérateur non-expansif tel que $\mathrm{Fix}T \neq \varnothing$. *Soit $x_0 \in D$, on définit la suite (x_n) par $x_{n+1} = Tx_n$. Alors, (x_n) est Fejér-monotone par rapport à* $\mathrm{Fix}\, T$.

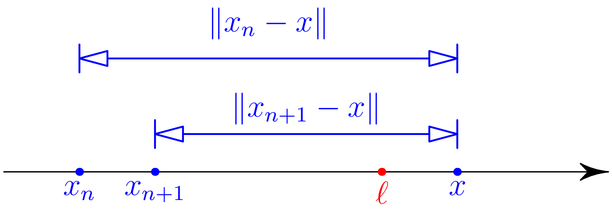


FIGURE 1.4 – Preuve de l'exemple (1.11).

Propriétés 1.5 *Si $(x_n)_{n\in\mathbb{N}}$ est une suite Fejér-monotone par rapport à $D \subset \mathbb{R}^n$ alors*

a) (x_n) est bornée.

b) Pour tout $x \in D, (\|x_n - x\|)$ converge.

c) $(d_D(x_n))_{n\in\mathbb{N}}$ est décroissante et convergente.

Preuve
Voir [1] ∎

Chapitre 2

Algorithmes Proximaux

2.1 Opérateur proximal

L'opérateur proximal en un point $x \in \mathbb{R}^n$ a été défini par Moreau en 1965 [3]. Pour une fonction f convexe, s.c.i et propre et $\gamma \in]0, +\infty[$, il existe un unique $p \in \mathbb{R}^n$ tel que

$$f(p) + \frac{1}{2\gamma}\|p - x\|^2 = \inf_{y \in \mathbb{R}^n} \Big(f(y) + \frac{1}{2\gamma}\|y - x\|^2 \Big).$$

Définition 2.1

- *Soit* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *une fonction convexe, s.c.i et propre. On appelle opérateur proximal de* f*, qu'on note* prox_f*. L'opérateur défini par*

$$\operatorname{prox}_f(u) = \underset{x \in \mathbb{R}^n}{\mathit{argmin}} \Big(f(x) + \frac{1}{2}\|x - u\|^2 \Big).$$

- *Soit* $f \in \Gamma_0(\mathbb{R}^n)$*. L'enveloppe de Moreau de paramètre* $\gamma \in]0, +\infty[$ *de* f *est*

$$^{\gamma}f : \mathbb{R}^n \to \mathbb{R} : x \mapsto \inf_{y \in \mathbb{R}^n} \Big(f(y) + \frac{1}{2\gamma}\|y - x\|^2 \Big).$$

2.1.1 Propriétés élémentaires

Proposition 2.1
Si $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *une fonction convexe, fermée et propre.* $\operatorname{prox}_f(x)$ *est bien défini (existe et unique) pour tout* $x \in \mathbb{R}^n$ *car la fonction* $(f(x) + \frac{1}{2}\|x - u\|^2)$ *est strictement convexe et coercive.*

De plus, il est caractérisé par

$$\begin{aligned} u &= \text{prox}_f(x). \\ \Leftrightarrow 0 &\in \partial f(u) + u - x. \\ \Leftrightarrow u &= (Id + \partial f)^{-1}(x). \\ \Leftrightarrow u &= J_{\partial f}(x). \end{aligned}$$

Proposition 2.2 (Caractérisation)
Soit f une fonction convexe sur $\mathbb{R}^n$, $(x, u) \in (\mathbb{R}^n)^2$,

$$u = \text{prox}_f(x) \Leftrightarrow \forall y \in \mathbb{R}^n, \langle y - u, x - u\rangle + f(u) \leq f(y).$$

Preuve
($\Rightarrow$) Supposons que $u = \text{prox}_f(x)$. Soit $p_\alpha = \alpha y + (1-\alpha)u$ où $y \in \mathbb{R}^n$ et $\alpha \in]0, 1]$ alors en utilisant la définition de l'opérateur proximal puis la convexité de f, on obtient

$$\begin{aligned} f(u) + \frac{1}{2}\|x - u\|^2 &\leq f(p_\alpha) + 1/2\|x - p_\alpha\|^2 \\ &\leq \alpha f(y) + (1-\alpha)f(u) + 1/2\|x - u + \alpha(u - y)\|^2. \end{aligned}$$

D'où

$$f(u) \leq \alpha f(y) + (1-\alpha)f(u) - \alpha\langle y - u, x - u\rangle + \frac{\alpha^2}{2}\|u - y\|^2.$$

Alors,

$$\langle y - u, x - u\rangle + f(u) \leq f(y) + \frac{\alpha}{2}\|y - u\|^2.$$

Le résultat s'obtient en faisant tendre α vers 0.
Réciproquement, si on suppose que $\langle y - u, x - u\rangle + f(u) \leq f(y)$ pour tout $y \in \mathbb{R}^n$, on a alors

$$\begin{aligned} f(u) + 1/2\|x - u\|^2 &\leq f(y) + 1/2\|x - u\|^2 + \langle u - y, x - u\rangle \\ &\leq f(y) + 1/2\|x - u\|^2 + \langle u - y, x - u\rangle + 1/2\|u - y\|^2 \\ &= f(y) + 1/2\|x - y\|^2. \end{aligned}$$

donc, d'après la définition de l'opérateur proximal on déduit que $u = \text{prox}_f(x)$. ■

Une propriété importante des points fixes de l'opérateur proximal est la suivante.

Proposition 2.3
Soit $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe et propre définie sur $\mathbb{R}^n$. Alors

$$\text{Fix}(\text{prox}_f) = \mathit{argmin}\, f.$$

i.e., $x^* = \text{prox}_f(x^*)$ si et seulement si $x^* \in \mathit{argmin}\, f$.

Preuve
$\Rightarrow$) Si x^* est un minimiseur de la fonction f, i.e., $f(x) \geq f(x^*)$ pour tout $x \in \mathbb{R}^n$, alors

$$f(x) + (1/2)\|x - x^*\|_2^2 \quad \geq f(x^*) = f(x^*) + (1/2)\|x^* - x^*\|_2^2$$

pour tout $x \in \mathbb{R}^n$ donc, x^* minimise la fonction $x \mapsto f(x) + (1/2)\|x - x^*\|_2^2$. On obtient

$$x^* = \operatorname{prox}_f(x^*).$$

$\Leftarrow$) Soit x^* tel que $\operatorname{prox}_f(x^*) = x^*$, donc x^* est un minimiseur de la fonction g définie par

$$g(x) = f(x) + \frac{1}{2}\|x - x^*\|^2.$$

D'après le Théorème 1.5 ou $0 \in \partial g(x^*)$. Or

$$\partial g(x) = \partial f(x) + (x - x^*).$$

En particulier $\partial g(x^*) = \partial f(x^*)$, on obtient $0 \in \partial f(x^*)$. Alors $x^* \in \operatorname{argmin} f$. ∎

Proposition 2.4 (Contraction et expansivité)
Soit $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ une fonction convexe, propre et fermée. Alors les applications prox_f et $Id - \operatorname{prox}_f$ sont fermement non-expansives, c'est à dire

$$(\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n) \quad \|\operatorname{prox}_f(x) - \operatorname{prox}_f(y)\|^2 + \|(x - \operatorname{prox}_f(x)) - (y - \operatorname{prox}_f(y))\|^2 \leq \|x - y\|^2.$$

En particulier, ces deux opérateurs sont non-expansif (1-Lipschitz).

Preuve
Soit $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$, si on note $p = \operatorname{prox}_f(x)$ et $q = \operatorname{prox}_f(y)$ alors d'après la Proposition 2.2 on a

$$\langle q - p, x - p\rangle + f(p) \leq f(q) \text{ et } \langle p - q, y - q\rangle + f(q) \leq f(p).$$

En additionnant ces deux inégalités on obtient

$$0 \leq \langle p - q, (x - p) - (y - q)\rangle.$$

On conclut

$$\begin{aligned}
\|x - y\|^2 &= \|p - q + (x - p) - (y - q)\|^2 \\
&= \|p - q\|^2 + \|(x - p) - (y - q)\|^2 + 2\langle p - q, (x - p) - (y - q)\rangle \\
&\geq \|p - q\|^2 + \|(x - p) - (y - q)\|^2 \\
&= \|\operatorname{prox}_f(x) - \operatorname{prox}_f(y)\|^2 + \|(x - \operatorname{prox}_f(x)) - (y - \operatorname{prox}_f(y))\|^2.
\end{aligned}$$

∎

Proposition 2.5 (Identité de Moreau)
Soit $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe et s.c.i. Alors

$$x = \operatorname{prox}_f(x) + \operatorname{prox}_{f^*}(x).$$

Preuve

On note $u = \text{prox}_f(x)$ et $v = x - u$. Le vecteur u satisfait

$$0 \in \partial f(u) + u - x \text{ alors } x - u \in \partial f(u) \text{ soit encore } v \in \partial f(u).$$

On a donc $u \in \partial f^*(v)$ soit encore $x - v \in \partial f^*(v)$ ou $v = \text{prox}_{f^*}(x)$.

$$\text{Alors, on conclut } x = u + v \Rightarrow x = \text{prox}_f(x) + \text{prox}_{f^*}(x).$$

■

Propriétés 2.1
Pour tout $i \in \{1, \dots, k\}$. soit $f_i \in \Gamma_0(\mathbb{R}^{n_i})$, pour tout $(x_1, \dots, x_k) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$. Si

$$f(x_1, \dots, x_k) = \sum_{i=1}^{k} f_i(x_i).$$

Alors

$$\text{prox}_f(x_1, \dots, x_k) = (\text{prox}_{f_1}(x_1), \dots, \text{prox}_{f_k}(x_k)) = (\text{prox}_{f_i}(x_i))_{1 \leq i \leq k}.$$

2.1.2 Calcul d'opérateurs proximaux

- ▷ Si $f(x) = \alpha$, alors $\text{prox}_f = Id_{\mathbb{R}^n}$.
  car, on a $\text{prox}_f(u) = \text{argmin}_{x \in \mathbb{R}^n}(f(x) + \frac{1}{2}\|x - u\|^2) = u$.
- ▷ Si $f(x) = \alpha\|x\|_1$.

$$(\text{prox}_f(x))_i = \begin{cases} u_i - \alpha & \text{si} \quad u_i < \alpha. \\ 0 & \text{si} \quad | u_i | \leq \alpha. \\ u_i + \alpha & \text{si} \quad u_i > -\alpha. \end{cases}$$

Pour $f(x) = \alpha|x|$ on a $\text{prox}_f(u) = \text{argmin}_{x \in \mathbb{R}^n}(\alpha|x| + \frac{1}{2}(x - u)^2)$. En posant que $F(x) = \alpha|x| + \frac{1}{2}(x - u)^2$ donc,

$$0 \in \partial F(x) = (x - u) + \alpha\partial|x| = \begin{cases} (x - u) - \alpha & \text{si} \quad x < 0. \\ -u + \alpha[-1, 1] & \text{si} \quad x = 0. \\ (x - u) + \alpha & \text{si} \quad x > 0. \end{cases}$$

- $x - u - \alpha = 0 \Rightarrow x = u + \alpha \Rightarrow u + \alpha < 0 \Rightarrow u < -\alpha$. donc, si $u < -\alpha$ alors $prox_f(u) = u + \alpha$.
- $[-\alpha, \alpha] - u = [-\alpha - u, \alpha - u]$ pour tout $x = 0$.
- $x - u + \alpha = 0 \Rightarrow x = u - \alpha > 0$. Donc, si $u > \alpha \Rightarrow prox_f(u) = u - \alpha$.

$$\text{prox}_f(u) = \begin{cases} u + \alpha & \text{si} \quad u < -\alpha. \\ 0 & \text{si} \quad -\alpha \leq u \leq \alpha. \\ u - \alpha & \text{si} \quad u > \alpha. \end{cases}$$

on a $(\text{prox}_f(u))_i = (\text{prox}_{f_i}(v_i))$.

▷ Si f est une fonction indicatrice d'un convexe fermé C dans $\mathbb{R}^n$ défini par

$$(\forall x \in \mathbb{R}^n) \quad \iota_c(x) = \begin{cases} 0 & \text{si } x \in C. \\ +\infty & \text{sinon .} \end{cases}$$

Son opérateur proximal est la projection euclidienne de x sur C. En effet,

$$\text{prox}_f(x) = \underset{u \in C}{\text{argmin}} \|u - x\|_2^2 = P_C(x)$$

.

▷ En particulier, si f est une fonction convexe et différentiable alors l'opérateur proximal de f au point x est caractérisé par l'équation implicite suivante

$$p = prox_f(x) \Leftrightarrow \nabla f(p) + p = x.$$

En effet, $p = \text{argmin}_y f(y) + \frac{1}{2}\|y - x\|^2$ si et seulement si $\nabla f(p) + p - x = 0$.
Par exemple si $f(x) = \frac{1}{2}\|Ax - b\|^2$ où A est une matrice $m \times n$ et $b \in \mathbb{R}^m$

$$p = (Id + A^\top A)^{-1}(x + A^\top b).$$

Selon les propriétés de A il est plus ou moins facile de calculer p. Il existe des situations où ce calcul peut être fait de manière rapide, c'est le cas par exemple si on dispose d'une transformée rapide pour diagonaliser $A^\top A$. C'est évidemment le cas si A est une matrice orthogonale. Si on ne dispose pas d'algorithmes rapides, on peut utiliser des algorithmes itératifs qui donneront des valeurs approchées du prox en un nombre fini d'étapes.

▷ Si $f(x_1, x_2) = f_1(x_1) + f_2(x_2)$ l'opérateur proximal de f s'exprime simplement à partir de ceux de f_1 et de f_2 dans la mesure où les minimisations sur x_1 et x_2 sont indépendantes, ainsi

$$\text{prox}_f(x_1, x_2) = (\text{prox}_{f_1}(x_1), \text{prox}_{f_2}(x_2)).$$

Propriétés	$g(x)$	$\text{prox}_g(x)$
Translation	$f(x - z), \quad z \in \mathbb{R}^n$	$z + \text{prox}_f(x - z)$
Perturbation quadratique	$f(x) + \alpha\|x\|^2/2 + \langle z, x\rangle + \gamma \; z \in \mathbb{R}^n, \; \alpha > 0, \; \gamma \in \mathbb{R}$	$\text{prox}_{\frac{f}{\alpha+1}}(\frac{x-z}{\alpha+1})$
Changement d'échelle	$f(\rho x), \; \rho \in \mathbb{R}^*$	$\frac{1}{\rho}\text{prox}_{\rho^2 f}(\rho x)$
Réflexion	$f(-x)$	$-\text{prox}_f(-x)$
Enveloppe de Moreau	${}^\gamma f(x) = \inf_{y \in \mathbb{R}^n} f(y) + \frac{1}{2\gamma}\|x - y\|^2 \; \gamma > 0$	$\frac{1}{1+\gamma}[\gamma x + \text{prox}_{(1+\gamma)f}(x)]$

2.2 Les algorithmes proximaux pour l'optimisation non-différentiable

L'opérateur proximal est central pour la résolution des problèmes d'optimisation non différentiable. Les méthodes proximales sont utilisées pour la résolution de quelques classes de problèmes telle que :

1. Algorithme du point proximal.
2. Somme d'une fonction convexe différentiable et d'une fonction convexe non-différentiable.
3. Somme de deux fonction convexe, propre et s.c.i.
4. Algorithme des directions alternées.

2.2.1 Algorithme du point proximal

Algorithme de point fixe

Dans la section des opérateurs proximaux, on a montré que x^* est un minimum de la fonction f si et seulement si x^* est un point fixe de l'opérateur proximal. Ce dernier est un opérateur non-expansif. L'algorithme du point fixe appliqué à un opérateur non expansif ne converge pas en général. Si on l'applique à l'opérateur $-Id$ qui est non expansif on obtient la suite (x_n) telle que $x_{n+1} = -x_n$ qui oscille entre x_0 et $-x_0$. Mais, si on prend N un opérateur non-expansif et l'opérateur $T = (1-\alpha)Id + \alpha N$ où $\alpha \in]0,1[$ alors, les points fixes de N et T sont les mêmes, i.e.,

$$x_{n+1} = (1-\alpha)x_n + \alpha N(x_n)$$

c'est à dire

$$\begin{aligned} Tx = x &= (1-\alpha)x + \alpha Nx \\ x &= x - \alpha x + \alpha Nx \\ \alpha x &= \alpha Nx \Leftrightarrow Nx = x. \end{aligned}$$

L'opérateur T est un opérateur α-moyenné alors cet opérateur est non expansif. Donc

$$\begin{aligned} \|Tx - Ty\| &= \|(1-\alpha)x + \alpha Nx - (1-\alpha)y + \alpha Ny\| \\ &= \|(1-\alpha)(x-y) + \alpha(Nx - Ny)\| \\ &\leq (1-\alpha)\|x-y\| + \alpha\|Nx - Ny\| \\ &\leq (1-\alpha)\|x-y\| + \alpha\|x-y\| = \|x-y\|. \quad (car N \text{ est non expansif }) \end{aligned}$$

Théorème 2.1

Soit $C \subset \mathbb{R}^n$ un convexe, fermé et non vide. Soit $T : C \to C$ un opérateur non expansif tel que $\mathrm{Fix}T \neq \varnothing$. *Soit $x_0 \in C$,*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = T(x_n).$$

Si $\|T(x_n) - x_n\| \to 0$ *alors la suite* $(x_n)_{n\in\mathbb{N}}$ *converge vers un point fixe de* T.

Preuve
Pour tout $n \in \mathbb{N}$ et y un point fixe de l'opérateur T alors

$$\|x_{n+1} - y\| \leq \|Tx_n - Ty\| \leq \|x_n - y\|.$$

Alors, la suite $(x_n)_{n\in\mathbb{N}}$ est Fejér-monotone par rapport à Fix$T$ et $(\|x_n - y\|)$ est décroissante donc bornée, ainsi la suite (x_n) est bornée. Soit $(x_{n_k})_{k\in\mathbb{N}}$ une sous-suite de $(x_n)_{n\in\mathbb{N}}$ telle que x_{n_k} converge vers x'. comme $\|Tx_{n_k} - x_{n_k}\| \to 0$ alors, $Tx' = x'$ et $x' \in$ Fix$T$. Puisque la suite (x_n) est Fejér-monotone par rapport à Fix$T$, on obtient que la suite $(\|x_n - x'\|)$ est décroissante, comme la sous suite $(\|x_{n_k} - x'\|)$ converge vers 0 alors la suite $(\|x_n - x'\|)$ converge aussi vers 0. D'où, la convergence de la suite (x_n) vers un élément de Fix$T$. ■

Ceci suggère immédiatement la méthode proximal la plus simple :

L'algorithme du point proximal est aussi appelé algorithme d'itération proximal où algorithme de minimisation proximal comme son nom l'indique est l'algorithme qui consiste à appliquer de manière récursive l'opérateur proximal de la fonction λf pour minimiser f sur $\mathbb{R}^n$.

Soit $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe, propre et s.c.i, alors l'algorithme est défini par

$$(x_0 \in \mathbb{R}^n) \quad x_{n+1} = \text{prox}_{\lambda f}(x_n).$$

où λ est un réel strictement positif. Une telle suite converge vers un minimiseur de f, il suffit pour cela d'appliquer le Théorème 2.1. En effet, $x_{n+1} = \text{argmin}\, f(x) + \frac{1}{2\lambda}\|x - x_n\|^2$ alors,

$$f(x_{n+1}) + \frac{1}{2\lambda}\|x_{n+1} - x_n\|^2 \leq f(x) + \frac{1}{2\lambda}\|x - x_n\|^2 \qquad \text{pour tout } x \in \mathbb{R}^n.$$

De même, on obtient

$$f(x_{n+1}) + \frac{1}{2\lambda}\|x_{n+1} - x_n\|^2 \leq f(x_n) \Rightarrow f(x_{n+1}) \leq f(x_n) - \frac{1}{2\lambda}\|x_{n+1} - x_n\|^2 \leq f(x_n).$$

Alors, la suite $(f(x_n))$ est décroissante et elle est minorée par $\inf f$, donc elle converge car

$$\frac{1}{2\lambda}\|x_{n+1} - x_n\|^2 \leq |f(x_n) - f(x_{n+1})|.$$

Par passage à la limite, on obtient $|f(x_n) - f(x_{n+1})|$ converge vers 0.

Théorème 2.2 *Soit* $f \in \Gamma_0(\mathbb{R}^n)$ *telle que* $\text{argmin}\, f \neq \varnothing$. *Soit* $(\lambda_n)_{n\in\mathbb{N}^n}$ *une suite de* $]0, +\infty[$ *telle que*

$$\sum_{n=0}^{+\infty} \lambda_n = +\infty \text{ et } x_0 \in \mathbb{R}^n.$$

Si, pour tout $n \in \mathbb{N}$,

$$x_{n+1} = \text{prox}_{\lambda_n f}(x_n).$$

alors,

a) $f(x_n) \to \inf f(\mathbb{R}^n)$.

b) $(x_n)_{n\in\mathbb{N}}$ converge vers un point de $\operatorname{argmin} f$.

Preuve

a) Soit $z \in \operatorname{argmin} f, x_{n+1} = \operatorname{prox}_{\lambda_n f}(x_n) \Leftrightarrow x_{n+1} = (Id + \lambda_n \partial f)^{-1}(x_n)$ donc

$$(\forall n \in \mathbb{N}) x_n - x_{n+1} \in \lambda_n \partial f(x_{n+1}).$$

D'après la définition du sous différentiel, on a

$$f(z) \geq f(x_{n+1}) + \frac{\langle z - x_{n+1}, x_n - x_{n+1}\rangle}{\lambda_n}.$$

De même, on obtient

$$f(x_n) - f(x_{n+1}) \geq \frac{\langle x_n - x_{n+1}, x_n - x_{n+1}\rangle}{\lambda_n}.$$

Donc,

$$\begin{aligned}
\|x_{n+1} - z\|^2 &= \|x_{n+1} - x_n + x_n - z\|^2 \\
&= \|x_{n+1} - x_n\|^2 + \|x_n - z\|^2 + 2\langle x_n - z, x_{n+1} - x_n\rangle \\
&= \|x_n - z\|^2 + \|x_{n+1} - x_n\|^2 + 2\langle x_n + x_{n+1} - x_{n+1} - z, x_{n+1} - x_n\rangle \\
&= \|x_n - z\|^2 + \|x_{n+1} - x_n\|^2 - 2\|x_{n+1} - x_n\|^2 + 2\langle z - x_{n+1}, x_n - x_{n+1}\rangle \\
&\leq \|x_n - z\|^2 - \|x_{n+1} - x_n\|^2 + 2\lambda_n[f(z) - f(x_{n+1})] \\
&\leq \|x_n - z\|^2 - 2\lambda_n[f(x_{n+1}) - f(z)] \\
&\leq \|x_n - z\|^2 - 2\lambda_n[f(x_{n+1}) - \inf f] \qquad (2.1) \\
&\leq \|x_n - z\|^2.
\end{aligned}$$

La suite (x_n) est Fejér-monotone par rapport à $\operatorname{argmin} f$ et d'après l'inégalité (2.1) on a

$$\begin{aligned}
\sum_{n=0}^{k} 2\lambda_n\Big[f(x_{n+1}) - \inf f\Big] &\leq \sum_{n=0}^{k} \|x_n - z\|^2 - \|x_{n+1} - z\|^2 \\
&\leq \|x_0 - z\|^2 - \|x_{k+1} - z\|^2.
\end{aligned}$$

On obtient,

$$\sum_{n=0}^{\infty} 2\lambda_n\Big[f(x_{n+1}) - \inf f\Big] < +\infty.$$

Alors, $f(x_n) \geq \inf f$ et la suite $(f(x_n))$ est décroissante et minorée donc elle converge vers ℓ.

- Si $\ell > \inf f$ alors $f(x_{n+1}) - \ell < f(x_{n+1}) - \inf f$, en effet

$$\begin{aligned}\sum_{n=0}^{\infty} 2\lambda_n\Big[f(x_{n+1}) - \inf f\Big] &\geq \sum_{n=0}^{\infty} 2\lambda_n(\ell - \inf f) \quad \text{pour tout } n \\ &\geq 2(\ell - \inf f)\Big(\sum_{n=0}^{\infty} \lambda_n\Big) \\ &\geq +\infty.\end{aligned}$$

On en déduit que $\ell = \inf f$ et $f(x_n) \to \inf f(\mathbb{R}^n)$.

b) La suite (x_n) admet une seule valeur d'adhérence. En effet, si x et y sont deux valeurs de la suite (x_n), alors il existe deux sous-suites (x_{n_k}) et (x_{n_l}) qui convergent vers x et y respectivement. Le sous suite (x_{n_k}) converge vers x et la fonction f est s.c.i en x alors $\liminf f(x_{n_k}) \geq f(x)$ comme $f(x_n) \to \inf f$ alors $\liminf f(x_{n_k}) = \inf f$. Par suite, $f(x) \leq \inf f$. Ce qui montre que $x \in \operatorname{argmin} f$. Pour montre $y \in \operatorname{argmin} f$ on a le sous suite (x_{n_l}) converge vers y et la fonction f est s.c.i en y alors $\liminf f(x_{n_l}) \geq f(y)$ comme $f(x_n) \to \inf f$ alors $\liminf f(x_{n_l}) = \inf f$. Par suite, $f(y) \leq \inf f$. Ce qui montre que $y \in \operatorname{argmin} f$.

$$\langle x_n, x - y\rangle = \|x_n - x\|^2 - \|x_n - y\|^2 + \|x\|^2 - \|y\|^2.$$

Alors, $\langle x_n, x - y\rangle$ converge vers une limite ℓ car les suites $(\|x_n - x\|)$ et $(\|x_n - y\|)$ convergent. En effet,

$$\begin{aligned}\langle x_{n_k}, x - y\rangle \quad &\text{converge vers} \quad \langle x, x - y\rangle = \ell \\ \langle x_{n_l}, x - y\rangle \quad &\text{converge vers} \quad \langle y, x - y\rangle = \ell.\end{aligned}$$

Donc, on obtient $\langle x, x - y\rangle = \langle y, x - y\rangle$ alors $\|x - y\|^2 = 0 \Rightarrow x = y$. Par conséquent, $(x_n)_{n\in\mathbb{N}}$ converge vers un point de $\operatorname{argmin} f$.

■

2.2.2 Algorithme d'optimisation : Forward-Backward

L'algorithme Forward-Backward (FB) ou bien méthode de gradient proximal est un algorithme qui permet de résoudre le problème d'optimisation suivant :

$$\text{trouver } x^* \in \underset{x\in\mathbb{R}^n}{\operatorname{argmin}} \Big(h(x) = f(x) + g(x)\Big) \tag{2.2}$$

où

- $f : \mathbb{R}^n \to \mathbb{R}$ est convexe, différentiable et à gradient L-Lipschitz i.e.,

$$(\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n) \quad \|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|.$$

- $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ est convexe et s.c.i qu'on peut calculer son opérateur proximal.

Cet algorithme consiste à alterner une descente de gradient explicite sur f et un opérateur proximal sur g. On suppose que ce problème admet au moins une solution. Les conditions d'optimalité de ce problème sont : $0 \in \partial g(x) + \nabla f(x) \Leftrightarrow -\nabla f(x) \in \partial g(x)$.

Proposition 2.6 *Soient $x^* \in \mathbb{R}^n$ et $\lambda \in]0, +\infty[$. Les assertions suivantes sont équivalentes :*

a) x^* est une solution du problème (2.2).

b) $x^* = \mathrm{prox}_{\lambda g}(x^* - \lambda \nabla f(x^*))$. Cette equation est appelé l'équation de point fixe.

Preuve

$$
\begin{aligned}
0 \in \nabla f(x^*) + \partial g(x^*) &\Leftrightarrow -\nabla f(x^*) \in \partial g(x^*). \\
&\Leftrightarrow x^* - \lambda \nabla f(x^*) \in x^* + \lambda \partial g(x^*). \\
&\Leftrightarrow x^* = (I + \lambda \partial g)^{-1}(I - \lambda \nabla f)(x^*). \\
&\Leftrightarrow x^* = prox_{\lambda g}(x^* - \lambda \nabla f(x^*)).
\end{aligned}
$$

■

D'après la caractérisation de point fixe précédente, on définit un algorithme itératif permettant de trouver une solution du problème (2.2) comme suit

$$\text{(Descente de gradient proximale)} \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = \mathrm{prox}_{\lambda g}(x_n - \lambda \nabla f(x_n)).$$

Théorème 2.3 (Algorithme de Krasnosel'skii-Mann)
Soit C un ensemble convexe, fermé et non vide de $\mathbb{R}^n$. Soit $T : C \to C$ un opérateur non-expansif tel que $\mathrm{Fix}T \neq \varnothing$. $(\lambda_n)_{n \in \mathbb{N}}$ une suite dans $[0, 1]$ telle que

$$\sum_{n \in \mathbb{N}} \lambda_n (1 - \lambda_n) = +\infty.$$

On a

$$x_0 \in C \ \textit{et} \ (\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n (T x_n - x_n).$$

Les propriétés suivantes sont satisfaites :

- *(x_n) est Fejér-monotone par rapport à $\mathrm{Fix}T$.*
- *$(T x_n - x_n)$ converge vers 0.*
- *(x_n) converge vers un point de $\mathrm{Fix}T$.*

Preuve

- On montre la Fejér-monotonicité de la suite (x_n) par rapport à $\mathrm{Fix}T$: $(\forall n \in \mathbb{N})(\forall x \in$

Fix$T$) on a

$$\begin{aligned}
\|x_{n+1}-x\|^2 &= \|x_n+\lambda_n(Tx_n-x_n)-x\|^2 \qquad (2.3)\\
&= \|(1-\lambda_n)(x_n-x)+\lambda_n(Tx_n-x)\|^2\\
&= (1-\lambda_n)\|x_n-x\|^2+\lambda_n\|Tx_n-x\|^2-\lambda_n(1-\lambda_n)\|Tx_n-x+x-x_n\|^2\\
&= (1-\lambda_n)\|x_n-x\|^2+\lambda_n\|Tx_n-Tx\|^2-\lambda_n(1-\lambda_n)\|Tx_n-x_n\|^2\\
&\le (1-\lambda_n)\|x_n-x\|^2+\lambda_n\|x_n-x\|^2-\lambda_n(1-\lambda_n)\|Tx_n-x_n\|^2\\
&\le \|x_n-x\|^2-\lambda_n(1-\lambda_n)\|Tx_n-x_n\|^2 \qquad (2.4)\\
&\le \|x_n-x\|^2.
\end{aligned}$$

- On montre que $Tx_n-x_n\to 0$. D'après l'inégalité (2.4) on obtient

$$\sum_{n\in\mathbb{N}}\lambda_n(1-\lambda_n)\|Tx_n-x_n\|^2\le\|x_0-x\|^2-\|x_{k+1}-x\|^2\le\|x_0-x\|^2. \qquad (2.5)$$

De plus, en utilisant l'hypothèse que T est non-expansif on obtient

$$\begin{aligned}
\|Tx_{n+1}-x_{n+1}\| &= \|Tx_{n+1}-Tx_n+(1-\lambda_n)(Tx_n-x_n)\|\\
&\le \|x_{n+1}-x_n\|+(1-\lambda_n)\|Tx_n-x_n\|\\
&= \|Tx_n-x_n\|.
\end{aligned}$$

Comme $\sum_{n\in\mathbb{N}}\lambda_n(1-\lambda_n)=+\infty$ et que la suite $(\|Tx_n-x_n\|)_{n\in\mathbb{N}}$ est décroissante on déduit de (2.5) que $(Tx_n-x_n)_{n\in\mathbb{N}}$ tend vers 0.

- Soit (x_{n_k}) une sous suite de (x_n) telle que $(x_{n_k})_{k\in\mathbb{N}}$ converge vers x', comme Tx_n-x_n tend vers 0, la suite $Tx_{n_k}-x_{n_k}\to 0$, en déduit que $Tx'=x'\Rightarrow x'\in\mathrm{Fix}T$. Puisque la suite (x_n) est Fejér-monotone par rapport à Fix$T$, on obtient que la suite $(\|x_n-x'\|)$ est décroissante, comme la sous suite $(\|x_{n_k}-x'\|)$ converge vers 0 alors la suite $(\|x_n-x'\|)$ converge aussi vers 0. D'où, la convergence de la suite (x_n) vers un élément de Fix$T$.

■

Proposition 2.7
Soient $T:\mathbb{R}^n\to\mathbb{R}^n$ un opérateur α-moyenné avec $\alpha\in]0,1[$ tel que Fix$T\neq\varnothing$ *et $(\lambda_n)_{n\in\mathbb{N}}$ une suite dans $[0,\frac{1}{\alpha}]$ telle que*

$$\sum_{n\in\mathbb{N}}\lambda_n(1-\alpha\lambda_n)=+\infty.$$

On a

$$(x_0\in\mathbb{R}^n)\ \text{et}\ (\forall n\in\mathbb{N})\quad x_{n+1}=x_n+\lambda_n(Tx_n-x_n).$$

Les propriétés suivantes sont satisfaites :

- *(x_n) est Fejér-monotone par rapport à* Fix$T$.
- *(Tx_n-x_n) converge vers* 0.
- *(x_n) converge vers un point de* Fix$T$.

Preuve
L'opérateur T étant α-moyenné alors il existe un opérateur R non expansif telle que

$$T = (1-\alpha)Id + \alpha R.$$

On note que, $\mathrm{Fix}T = \mathrm{Fix}R$. En posant, $\mu_n = \alpha\lambda_n$, on s'aperçoit que $\mu_n \in [0,1]$ pour tout $n \in \mathbb{N}$.

Les itérations peuvent se réécrire

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad x_{n+1} &= x_n + \lambda_n(Tx_n - x_n) \\ &= x_n + \mu_n(Rx_n - x_n).\end{aligned}$$

Les résultats découlent alors du Théorème 2.3. ∎

Proposition 2.8 *Soit $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ un opérateur monotone maximal et $B : \mathbb{R}^n \to \mathbb{R}^n$ un opérateur β-cocoercif où $\beta \in]0,+\infty[$. Soient $\gamma \in]0, 2\beta[$ et $\delta = \min\{1, \beta/\gamma\} + 1/2$ et $(\lambda_n)_{n\in\mathbb{N}}$ une suite dans $[0, \delta[$ telle que $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.*
Supposons que $zer(A+B) \neq \varnothing$. Soit $x_0 \in \mathbb{R}^n$ et

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = x_n - \gamma B x_n \\ x_{n+1} = x_n + \lambda_n(J_{\gamma A} y_n - x_n). \end{cases}$$

Alors, la suite $(x_n)_{n\in\mathbb{N}}$ converge vers un point de $zer(A+B)$.

Preuve
Soit $T = J_{\gamma A} \circ (Id - \gamma B)$. D'un coté, on a $J_{\gamma A}$ est $\frac{1}{2}$- moyenné d'après le Corollaire 1. Et d'un autre coté, la Proposition 1.19 implique que $Id - \gamma B$ est $\frac{\gamma}{(2\beta)}$-moyenné. Pour tout $x \in \mathbb{R}^n$,

$$x \in \mathrm{Fix}\ T \Leftrightarrow x - \gamma Bx \in x + \gamma A \Leftrightarrow 0 \in \gamma(Ax + Bx) \Leftrightarrow 0 \in Ax + Bx.$$

D'où $FixT = zer(A+B) \neq \varnothing$. De plus, pour tout $n \in \mathbb{N}$

$$x_{n+1} = x_n + \lambda_n(Tx_n - x_n).$$

On en déduit que T est α-moyenné d'après la Proposition 1.18 avec

$$\alpha = \frac{2}{1 + \frac{1}{\max\{\frac{1}{2}, \frac{\gamma}{2\beta}\}}} = \frac{2}{1 + \frac{2}{\max\{1, \frac{\gamma}{\beta}\}}} = \frac{2}{1 + 2\min\{1, \frac{\beta}{\gamma}\}} = \frac{2}{2\delta} = \frac{1}{\delta}.$$

Le résultat découle alors du précédent. ∎

Remarque 2.1 *Quand $B = 0$ et $\lambda_n \equiv 1$, on retrouve l'algorithme du point proximal.*

Corollaire 2 *Soit $f \in \Gamma_0(\mathbb{R}^n)$ et g une fonction convexe différentiable de gradient $1/\beta$-lipschitzien où $\beta \in]0,+\infty[$. Soient $\gamma_n \in]0, 2\beta[$ et $\delta = \min\{1, \frac{\beta}{\gamma}\} + \frac{1}{2}$ et $(\lambda_n)_{n\in\mathbb{N}}$ une suite dans $[0, \delta[$ telle que*

$$\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty.$$

Supposons que argmin$(f+g) \neq \varnothing$. Soit $x_0 \in \mathbb{R}^n$ et

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = x_n - \gamma_n \nabla g(x_n) \\ x_{n+1} = x_n + \lambda_n \Big[\text{prox}_{\gamma_n f}(y_n) - x_n\Big]. \end{cases}$$

Alors, la suite $(x_n)_{n\in\mathbb{N}}$ converge vers un minimiseur de $f+g$.

Preuve
Le résultat s'obtient en utilisant la Proposition 2.8 avec $A = \partial f$ et $B = \nabla g$ et en remarquant que $\text{zer}(\partial f + \nabla g) = \text{argmin}(f+g)$. ■

Interprétation Majoration-Minimisation (MM)

On interpréter la méthode de gradient proximal par l'algorithme Majoration-Minimisation, alors pour tout $k \in \mathbb{N}$ et d'après le Lemme 1.1, on obtient

$$\begin{aligned} g(x) &\leq g(x_k) + \langle \nabla g(x_k), x - x_k \rangle + \frac{1}{2\beta}\|x - x_k\|^2 \\ &\leq g(x_k) + \langle \nabla g(x_k), x - x_k \rangle + \frac{1}{\gamma_k}\|x - x_k\|^2 = \widehat{f}_{\gamma_k}(x, x_k). \end{aligned}$$

On en déduit

$$f(x) + g(x) \leq \widehat{f}_{\gamma_k}(x, x_k) + f(x) = q_{\gamma_k}(x, x_k) \qquad \text{pour tout } x.$$

On minimise une majoration de $f+g$ qui est $q_{\gamma_k}(x, x_k)$.

$$x_{k+1} = \text{argmin}\, q_{\gamma_k}(x, x_k).$$

Alors, on obtient

$$\begin{aligned} 0 \in \partial f(x_{k+1}) + \nabla g(x_k) + \frac{2}{\gamma_k}(x_{k+1} - x_k) &\Leftrightarrow 0 \in \frac{\gamma_k}{2}\partial f(x_{k+1}) + \frac{\gamma_k}{2}\nabla g(x_k) + (x_{k+1} - x_k) \\ &\Leftrightarrow x_k - \frac{\gamma_k}{2}\nabla g(x_k) \in (Id + \frac{\gamma_k}{2}\partial f)(x_{k+1}) \\ &\Leftrightarrow x_{k+1} = (Id + \frac{\gamma_k}{2}\partial f)^{-1}(x_k - \frac{\gamma_k}{2}\nabla g(x_k)) \\ &\Leftrightarrow x_{k+1} = prox_{\frac{\gamma_k}{2}f}(x_k - \frac{\gamma_k}{2}\nabla g(x_k)) \\ &\Leftrightarrow x_{k+1} = prox_{\beta_k f}(x_k - \beta_k \nabla g(x_k)). \end{aligned}$$

Remarque 2.2 *Si les suites (γ_n) et (λ_n) vérifiant la relation $\gamma_n \lambda_n \leq \beta$ pour tout n, alors la suite $(f(x_n) + g(x_n))$ est décroissante. En effet, pour tout $n \in \mathbb{N}$, définissons $p_n = \text{prox}_{\gamma f}(x_n - \gamma_n \nabla g(x_n))$, on a en choisissant $0 \leq \lambda_n \leq 1$*

$$f(x_{n+1}) = f((1-\lambda_n)x_n + \lambda_n p_n) \leq (1-\lambda_n)f(x_n) + \lambda_n f(p_n). \tag{2.6}$$

Par ailleurs, d'après le lemme de descente,

$$g(x_{n+1}) \leq g(x_n) + \langle \nabla g(x_n), x_{n+1} - x_n \rangle + \frac{1}{2\beta}\|x_{n+1} - x_n\|^2. \tag{2.7}$$

Enfin, d'après la définition de l'opérateur proximal,

$$\gamma_n f(p_n) + \frac{1}{2}\|p_n - x_n + \gamma_n \nabla g(x_n)\|^2 \leq \gamma_n f(x_n) + \frac{1}{2}\gamma_n^2 \|\nabla g(x_n)\|^2.$$

On en déduit

$$f(p_n) + \langle \nabla g(x_n), p_n - x_n \rangle + \frac{1}{2}\gamma_n^{-1}\|p_n - x_n\|^2 \leq f(x_n).$$

Puisque $x_{n+1} - x_n = \lambda_n(p_n - x_n)$*, on a*

$$\lambda_n f(p_n) + \langle \nabla g(x_n), x_{n+1} - x_n \rangle + \frac{1}{2}\gamma_n^{-1}\lambda_n^{-1}\|x_{n+1} - x_n\|^2 \leq \lambda_n f(x_n) \tag{2.8}$$

Par l'addition des équations (2.6), (2.7) *et* (2.8)*, on obtient ;*

$$f(x_{n+1}) + g(x_{n+1}) + \frac{1}{2}(\gamma_n^{-1}\lambda_n^{-1} - \beta^{-1})\|x_{n+1} - x_n\|^2 \leq f(x_n) + g(x_n).$$

Donc, si $\gamma_n^{-1}\lambda_n^{-1} - \beta^{-1} \geq 0$ *ou* $\gamma_n\lambda_n \leq \beta$*, la suite* $(f(x_n) + g(x_n))_{n\in\mathbb{N}}$ *est décroissante.*

2.2.3 Algorithme d'optimisation : Douglas-Rachford

Dans cette partie on souhaite résoudre le problème

$$\min_{x\in\mathbb{R}^n} F(x) = \min_{x\in\mathbb{R}^n} \Big(f(x) + g(x)\Big).$$

où les fonctions f et g sont convexes, propres et s.c.i. Ici on ne fait pas d'hypothèses de différentiabilité sur f. On remarque dans le cas de l'algorithme Forward-Backward les points fixes d'un certain opérateur sont les minimiseurs de F et pour l'algorithme Douglas-Rachford, les minimiseurs sont les images par un opérateurs de ces points fixes.

Proposition 2.9 *Soient* $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ *et* $B : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ *deux opérateurs monotones et* $\gamma \in]0, +\infty[$*. On a*

$$\mathrm{zer}(A + B) = J_{\gamma B}(\mathrm{Fix}(R_{\gamma A}R_{\gamma B})).$$

Preuve

$$
\begin{aligned}
0 \in \gamma(Ax+Bx) &\Leftrightarrow (\exists z \in \mathbb{R}^n) \quad z \in \gamma Ax \text{ et } -z \in \gamma Bx \\
&\Leftrightarrow (\exists y \in \mathbb{R}^n) \quad x-y \in \gamma Ax \text{ et } y-x \in \gamma Bx \\
&\Leftrightarrow 2x-y \in (Id+\gamma A)x \text{ et } x = J_{\gamma \mathrm{B}} y \\
&\Leftrightarrow x = J_{\gamma \mathrm{A}}(R_{\gamma \mathrm{B}} y) \text{ et } x = J_{\gamma \mathrm{B}} y \\
&\Leftrightarrow R_{\gamma \mathrm{A}}(R_{\gamma \mathrm{B}} y) = 2x - R_{\gamma \mathrm{B}} y = y \text{ et } x = J_{\gamma \mathrm{B}} y \\
&\Leftrightarrow (\exists y \in \mathrm{Fix} R_{\gamma \mathrm{A}} R_{\gamma \mathrm{B}}) \quad x = J_{\gamma \mathrm{B}} y.
\end{aligned}
$$

■

Proposition 2.10
Soient $A, B : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ deux opérateurs monotones maximaux. Soit $\gamma \in]0, +\infty[$ et soit $(\beta_n)_{n\in\mathbb{N}}$ une suite dans $[0,2]$ telle que $\sum_{n\in\mathbb{N}} \beta_n(2-\beta_n) = +\infty$. Supposons que $\mathrm{zer}(A+B) \neq \varnothing$.

Soit $x_0 \in \mathbb{R}^n$ et

$$
(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = J_{\gamma B} x_n \\ z_n = J_{\gamma A}(2y_n - x_n) \\ x_{n+1} = x_n + \beta_n(z_n - y_n). \end{cases}
$$

Les propriétés suivantes sont satisfaites :

1) $x_n \to x' \in \mathbb{R}^n$.

2) (y_n) et (z_n) convergent vers $J_{\gamma B} x' \in \mathrm{zer}(A+B)$.

Preuve
L'opérateur $T = R_{\gamma \mathrm{A}} R_{\gamma \mathrm{B}}$ est non expansif puisque $R_{\gamma \mathrm{A}}$ et $R_{\gamma \mathrm{B}}$ le sont. Aussi, d'après Proposition 2.9 on a $\mathrm{zer}(A+B) = J_{\gamma \mathrm{B}}(\mathrm{Fix} T)$ qui est non vide. Par suite l'ensemble $\mathrm{Fix} T$ est non vide. De plus, on a pour tout $n \in \mathbb{N}$.

$$
\begin{aligned}
x_{n+1} &= x_n + \beta_n[J_{\gamma \mathrm{A}}(2J_{\gamma \mathrm{B}} x_n - x_n) - J_{\gamma \mathrm{B}} x_n] \\
&= x_n + \beta_n[J_{\gamma \mathrm{A}}(R_{\gamma \mathrm{B}} x_n) - J_{\gamma \mathrm{B}} x_n] \\
&= x_n + \tfrac{\beta_n}{2}[2J_{\gamma \mathrm{A}}(R_{\gamma \mathrm{B}} x_n) - R_{\gamma \mathrm{B}} x_n - x_n] \\
&= x_n + \tfrac{\beta_n}{2}(T x_n - x_n) \\
&= x_n + \lambda_n(T x_n - x_n).
\end{aligned}
$$

D'après le Théorème 2.3 avec $\lambda_n = \frac{\beta_n}{2}$, on obtient que $Tx_n - x_n \to 0$ et $x_n \to x' \in \mathrm{Fix} T$. Montrons maintenant que les suites (y_n) et (z_n) convergent vers $J_{\gamma \mathrm{B}} x' \in \mathrm{zer}(A+B)$. Pour tout $n \in \mathbb{N}$,

$$
\begin{aligned}
z_n - y_n &= J_{\gamma \mathrm{A}}(2J_{\gamma \mathrm{B}} x_n - x_n) - J_{\gamma \mathrm{B}} x_n \\
&= \tfrac{1}{2}[2J_{\gamma \mathrm{A}}(R_{\gamma \mathrm{B}} x_n) - R_{\gamma \mathrm{B}} x_n - x_n] \\
&= \tfrac{1}{2}(T x_n - x_n) \to 0.
\end{aligned}
$$

$J_{\gamma \mathrm{B}}$ est continue, on a $y_n \to J_{\gamma \mathrm{B}}x' \in \operatorname{zer}(A+B)$ et puisque $z_n - y_n \to 0$, on a aussi que $z_n \to J_{\gamma \mathrm{B}}x' \in \operatorname{zer}(A+B)$. ■

Théorème 2.4 *Soient f et g deux fonctions dans $\Gamma_0(\mathbb{R}^n)$ telles que argmin$(f+g) \neq \varnothing$. Etant donné $\gamma \in]0,+\infty[$ et $(\lambda_n)_{n\in\mathbb{N}}$ une suite d'éléments de $[0,2]$ telle que $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$. Considérons la suite (x_n) définie par :*

Soit $x_0 \in \mathbb{R}$ et

$$\begin{cases} y_n = \operatorname{prox}_{\gamma g} x_n \\ z_n = \operatorname{prox}_{\gamma f}(2y_n - x_n) \\ x_{n+1} = x_n + \lambda_n(z_n - y_n). \end{cases}$$

On suppose que l'une de ces trois hypothèses est vérifiée :

a) $(\mathbf{ri}\operatorname{dom}(f)) \cap (\mathbf{ri}\operatorname{dom}(g)) \neq \varnothing$.

b) g est polyédrique et $\operatorname{dom}(g) \cap \mathbf{ri}\operatorname{dom}(f) \neq \varnothing$.

c) f et g sont polyédrique et $\operatorname{dom}(f) \cap \operatorname{dom}(g) \neq \varnothing$.

Alors, il existe $x \in \mathbb{R}^n$ tel que

1) $\operatorname{prox}_{\gamma g} x \in$ *argmin$(f+g)$.*

2) (x_n) converge vers x.

3) (y_n) et (z_n) convergent vers $\operatorname{prox}_{\gamma g} x$.

Preuve

Le résultat s'obtient en utilisant la Proposition 2.10 avec $A = \partial f$ et $B = \partial g$ qui sont des opérateurs monotones maximaux d'après le Théorème 1.9. De plus, on a $J_{\gamma \partial f} = \operatorname{prox}_{\gamma f}$ d'après la Proposition 2.1 et $\operatorname{argmin}(f+g) = \operatorname{zer}(A+B)$. ■

Remarque 2.3 *L'algorithme de Douglas-Rachford peut être exprimé de plusieurs manières dans la littérature. Il est fréquent par exemple que les paramètres λ_n soient fixés à 1, l'algorithme s'exprime alors de la manière suivante*

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = \operatorname{prox}_{\gamma g} x_n \\ z_n = \operatorname{prox}_{\gamma f}(2y_n - x_n) \\ x_{n+1} = x_n + z_n - y_n. \end{cases}$$

Mais en se passant de la variable z_n on obtient alors la description des suites $(x_n)_{n\in\mathbb{N}}$ et $(y_n)_{n\in\mathbb{N}}$ de la manière suivante

$$(\forall n \geq 1) \quad \begin{cases} x_n = x_{n-1} + \operatorname{prox}_{\gamma f}(2y_{n-1} - x_{n-1}) - y_{n-1}, \\ y_n = \operatorname{prox}_{\gamma g}(x_n). \end{cases}$$

On peut aussi introduire comme variable auxiliaire $u_n = \operatorname{prox}_{\gamma f}(2y_{n-1} - x_{n-1})$, changer l'ordre de mise à jour des variables et réécrire l'algorithme

$$(\forall n \geq 1) \quad \begin{cases} u_n = \operatorname{prox}_{\gamma f}(2y_{n-1} - x_{n-1}) \\ x_n = x_{n-1} + u_n - y_{n-1}, \\ y_n = \operatorname{prox}_{\gamma g}(x_{n-1} + u_n - y_{n-1}). \end{cases}$$

Beaucoup de choix sont possibles, on peut citer pour terminer un changement de variables qui apparait quelque fois dans la littérature : $w_n = y_n - x_n$, *on a alors*

$$(\forall n \geq 1) \quad \begin{cases} u_n = \mathrm{prox}_{\gamma f}(y_{n-1} + w_{n-1}) \\ y_n = \mathrm{prox}_{\gamma g}(u_n - w_{n-1}), \\ w_n = w_{n-1} + y_n - u_n. \end{cases}$$

Toutes ces formulations sont ainsi équivalentes.

2.2.4 Algorithme des directions alternées

L'algorithme des directions alternées (en englais Alternating Direction Method of Multipliers ou ADMM) est un algorithme pour résoudre un problème d'optimisation de la forme

$$(1) \quad \begin{cases} \underset{(x_1,x_2)}{\text{Minimiser}} \Big(f_1(x_1) + f_2(x_2)\Big) \\ \text{s.c.} \\ A_1x_1 + A_2x_2 = b, \end{cases}$$

où A_1 et A_2 sont deux applications linéaires à valeurs dans $\mathbb{R}^m, b$ un vecteur de $\mathbb{R}^m$, f_1 et f_2 deux fonctions convexes, propres et s.c.i. C'est un cadre général qui contient le cas $x_2 = x_1$ et même $x_2 = Ax_1$. Nous présentons ici l'algorithme sans en donner de preuve de convergence dans un cadre général. Nous verrons que dans le cas particulier où $A_1 = Id$, $A_2 = -Id$ et $b = 0$, c'est à dire le cas $x_1 = x_2$, cet algorithme se réduit à l'algorithme de Douglas-Rachford.

- Algorithme ADMM : Soit $(x_1^0, x_2^0) \in \mathbb{R}^n$, soit $\gamma > 0$ et $z^0 \in \mathbb{R}^m$. On définit les suites $(x_1^n)_{n\in\mathbb{N}}, (x_2^n)_{n\in\mathbb{N}}$ et $(z^n)_{n\in\mathbb{N}}$ de la manière suivante :

$$(\forall n \geq 1) \begin{cases} x_1^n = \mathrm{argmin}_x\, f_1(x) + \langle z^{n-1}, A_1x\rangle + \frac{1}{2\gamma} \parallel A_1x + A_2x_2^{n-1} - b \parallel^2 \\ x_2^n = \mathrm{argmin}_y\, f_2(y) + \langle z^{n-1}, A_2y\rangle + \frac{1}{2\gamma} \parallel A_1x_1^n + A_2y - b \parallel^2 \\ z^n = z^{n-1} + \gamma(A_1x_1^n + A_2x_2^n - b). \end{cases}$$

Cet algorithme permet de résoudre le problème au sens suivant :

Théorème 2.5 *Si* f_1 *et* f_2 *sont deux fonctions propres, s.c.i, convexes et coercives, alors*

1. *La suite* $(f_1(x_1^n) + f_2(x_2^n))$ *converge vers la valeur minimale de* $f_1 + f_2$.
2. *Les suites* (x_1^n) *et* (x_2^n) *convergent.*
3. *La suite* $(Ax_1^n + Ax_2^n - b)$ *tend vers* 0.

La démonstration de cette théorème dont la preuve la plus simple consiste à montrer que résoudre (1) est équivalent à résoudre un problème dual par un algorithme de Douglas-Rachford.

On peut cependant faire plusieurs remarques sur cet algorithme :

- Son nom vient du fait qu'il peut être vu comme une variante d'un algorithme connu sous le nom de méthode du Lagrangien augmenté. Si on remplace les mises à jours de x_1 et x_2 par une mise à jour conjointe

$$(x_1^n, x_2^n) = \underset{(x,y)\in\mathbb{R}^n}{\mathrm{argmin}}\, f_1(x) + f_2(y) + \langle z^{n-1}, A_1x + A_2y\rangle + \frac{1}{2\gamma}\|A_1x + A_2y - b\|^2.$$

On retrouve la méthode du Lagrangien augmenté qui consiste à pénaliser les contraintes avec d'une part un multiplicateur de Lagrange z et un terme quadratique. Un des problème de cette méthode et qu'une telle minimisation conjointe est souvent très difficile à réaliser. L'ADMM découple ce problème en optimisant d'abord sur la première variable x_1 puis sur la variable x_2.

- La variable z peut être interprétée comme un multiplicateur de Lagrange qui est mis à jour à chaque étape.
- Dans le cas particulier où $A_1 = Id, A_2 = -Id, b = 0$ et $\gamma = 1$, l'algorithme peut s'exprimer à l'aide d'opérateurs proximaux, plus précisément il devient,

$$\begin{cases} x_1^n = \text{prox}_{f_1}(x_2^{n-1} - z^{n-1}) \\ x_2^n = \text{prox}_{f_2}(x_1^n + z^{n-1}) \\ z^n = z^{n-1} + (x_1^n - x_2^n). \end{cases}$$

  ce qui est bien une des formes du Douglas-Rachford.
- Dans un cadre général, aucune des deux premières mises à jour n'est simple à effectuer. Si $A_1 = Id$ la mise à jour de x_1 revient à un calcul de l'opérateur proximal de f_1, même chose si $A_2 = Id$ pour la seconde mise à jour. Si l'un des deux opérateurs n'est pas l'identité, il faut souvent utiliser un algorithme itératif pour effectuer la minimisation. On a alors des boucles internes. Chaque situation doit être étudiée au cas par cas, en fonction des fonctions f_1 et f_2 et des opérateurs A_1 et A_2. les performances de l'ADMM dépendront alors des choix qui seront faits pour réaliser ces 2 minimisations. On peut noter toutefois que les deux problèmes de minimisation peuvent être traités par FB si on dispose des opérateurs proximaux de f_1 et f_2. En effet le terme quadratique est différentiable et son gradient est explicite.

ADMM linéarisé

Si on considère un problème de la forme

$$\underset{x \in \mathbb{R}^n}{\text{argmin}}\, f(x) + g(Ax),$$

où A est un opérateur linéaire quelconque et que g n'est pas différentiable, on ne peut a priori pas utiliser ni FB, ni Douglas-Rachford où ADMM directement même si on dispose de forme explicite des opérateurs proximaux de f et g. En revanche on peut réécrire ce problème de la manière suivante pour utiliser un ADMM.

$$\underset{x_2 - Ax_1 = 0}{\text{argmin}}\, f(x_1) + g(x_2).$$

Dans ce cas, $A_2 = Id$ et une mise à jour de l'ADMM se calcule via un opérateur proximal et l'autre doit être traitée à part, soit en utilisant des propriétés de l'opérateur A, soit en utilisant FB. On peut également réécrire ce problème de la manière suivante

$$\underset{x_1, x_2}{\text{argmin}}\, f(x_1) + g(x_2) + i_{x_2 = Ax_1}(x_1, x_2).$$

Si on considère $f_1(x_1, x_2) = f(x_1) + g(x_2)$ et $f_2(x_1, x_2) = i_{x_2=Ax_1}(x_1, x_2)$, on peut appliquer un Douglas-Rachford, en effet l'opérateur proximal de f_1 s'exprime simplement à partir de celui de f et de g (voir la section sur le calcul des opérateurs proximaux) et l'opérateur proximal de l'indicatrice est une projection sur un espace vectoriel. Cette projection nécessite la résolution d'un système de la forme prox $= (Id + A^\top A)^{-1}z$ qui peut être plus ou moins facile selon l'opérateur A mais il est toujours possible de le résoudre au moins de manière approchée. On peut noter que cette manière de réécrire le problème est également possible même si $A = Id$, on peut alors appliquer un Douglas-Rachford non pas sur f et g mais sur $f(x_1) + g(x_2)$ d'un coté et sur une indicatrice de l'autre.

Conclusion

Ce mémoire a pour objectif de présenter de nouvelles méthodes pour la résolution des problèmes d'optimisation non différentiable, parmi ces méthodes les méthodes proximaux. Ces dernières sont présentés en détail par Combettes et Pesquet en 2011 et plus récemment par Parikh et Boyd en 2013.

On a étudié les méthodes proximaux par l'aide l'opérateur proximal. La premier méthode est l'algorithme du point proximal qui consiste à appliquer l'opérateur proximal d'une fonction de manière récursive pour minimiser une fonction. L'algorithme du gradient proximal permet de trouver les minimiseurs d'une somme de deux fonctions convexes, propres et fermés l'une étant supposée différentiable. L'algorithme de Douglas-Rachford consiste à minimiser deux fonctions convexes, propres et semi continues inférieurement, ainsi que l'algorithme des directions alternée dans un cadre général.

Enfin, les problèmes de convergence et la théorie des opérateurs proximaux (et monotone) sont bien détaillés par Bauschke et Combettes en 2011.

Malheureusement, on a pas le temps de faire une étude numérique de ces méthodes afin de voir leur efficacité.

Bibliographie

[1] H. Bauschke and P. Combettes. *"Convex Analysis and Monotone Operator Theory in Hilbert Spaces."*, Springer-Verlag, 2011.

[2] P. L. Combettes. *"Monotone Operator Theory in Convex Optimization."*, North Carolina State University, Departement of Mathematics, Raleigh, NC 27695-8205, USA, arXiv :1802.02694v3 [math.OC] 1 Jun 2018.

[3] P. Combettes and J. C. Pesquet. *"Proximal splitting methods in signal processing"*., Fixed Point Algorithms for Inverse Problems in Science and Engineering, pp. 185-212, 2011.

[4] J.J. Moreau. *"Proximité et dualité dans un espace hilbertien."*, Bull Soc. Math. France 93 :273-299, 1965.

[5] N. Parikh and S. Boyd. *"Proximal Algorithms."*, Foundations and Trends in Optimization, Vol. 1, No. 3, 123-231, 2013.

[6] R. T. Rockafellar. *"Convex Analysis."*, Princeton University Press, 1970.

[7] R. T. Rockafellar et R.J.B. Wets. *"Variational Analysis."*, Springer-Verlag, 1997.

[8] S. Stephen. *"From Hahn-Banach to Monotonicity."*, Lecture Notes in Mathematics, 2nd, expanded edition, Springer, 2008.