UNIVERSITÉ DE TOULON

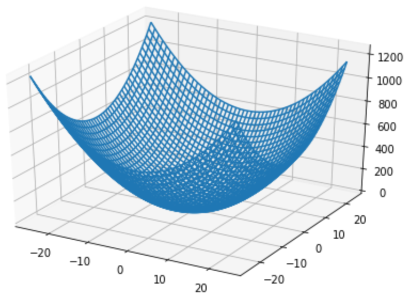
Master de mathématique appliquée

Option :

Mathématiques Appliquées : Analyse appliquée et Physique Mathématique

PROJET ACADEMIQUE

Minimisation de la somme d'une fonction différentiable et d'une fonction convexe : Variable Metric Forward-Backward algorithm



[**Source : image prise sur internet**]

Travail réalisé par :
HADAD Ahmed Ali
2021-2022

Sous l'encadrement de :
Mr **Champion THIERRY**

INTRODUCTION

Les problèmes d'optimisation interviennent dans différents domaines des mathématiques :

- Déterminer les dimensions optimales d'un objet,
- Déterminer la stratégie optimale dans un jeu,
- Déterminer un estimateur optimal pour une classe de fonction (seuillage pour un débruitage d'images par exemple), maximum de vraisemblance ou maximum a posteriori en statistiques,
- Déterminer le minimum d'un critère en restauration d'images,
- Transport optimal.

On peut écrire beaucoup de ces problèmes sous la forme :

$$\underset{x \in E}{\mathrm{Argmin}} f(x)$$

où f est une fonction définie sur un espace de Hilbert H ou plus souvent un espace euclidien, typiquement $\mathbb{R}^n$, à valeurs dans $\mathbb{R} \cup \{+\infty\}$ et E peut être l'espace H tout entier, un sous espace de ce dernier ou un sous ensemble convexe de H défini ou non par un ensemble de contraintes. Les points x de l'ensemble E sont dits admissibles.

Nous allons nous concentrer ici sur le problème de minimisation d'une fonction G,

$$\begin{aligned} G : \mathbb{R}^n &\longrightarrow (-\infty, +\infty] \\ x &\longmapsto G(x) = F(x) + R(x) \end{aligned}$$

qui est la somme d'une fonction différentiable F (pas nécessairement convexe) et d'une fonction convexe R (pas nécessairement différentiable). Un tel problème peut être résolu avec l'algorithme **Forward-Backward** (**FB**).

$$(\mathbf{FB}) \begin{cases} x_0 \in \mathbb{R}^n \quad \text{fixé} \\ \text{pour} \quad k \geqslant 0 : \\ \qquad y_k = \mathrm{prox}_{\gamma_k R}(x_k - \gamma_k \nabla F(x_k)) \\ \qquad x_{k+1} = x_k + \lambda_k (y_k - x_k) \end{cases}$$

où pour tout $k \in \mathbb{N}$, $(\gamma_k, \lambda_k) \in (0, +\infty)^2$ et prox est l'opérateur de proximité.
La convergence de l'algorithme **(FB)** repose principalement sur l'hypothèse que la fonction objectif G satisfait la propriété de **Kurdyka-Lojasiewicz**.
On peut voir qu'à chaque itération, on résout un problème de minimisation très compliqué dont le coût de calcul cause un ralentissement de convergence de l'algorithme. Une stratégie d'accélération est l'utilisation de la métrique variable : la métrique sous-jacente de l'algorithme (**FB**) est modifiée à chaque itération donnant lieu à ce qu'on appelle l'algorithme **Forward-Backward de métrique variable (MVFB)** :

$$(\mathbf{MVFB}) \begin{cases} x_0 \in \mathbb{R}^n \quad \text{fixé} \\ \text{pour} \quad k \geqslant 0 : \\ \qquad y_k = \mathrm{prox}_{\gamma_k^{-1} A_k, R}(x_k - \gamma_k A_k^{-1} \nabla F(x_k)) \\ \qquad x_{k+1} = x_k + \lambda_k (y_k - x_k) \end{cases}$$

où pour tout $k \in \mathbb{N}$, A_k est une matrice symétrique définie positive.
Ce travail est basé sur l'article ([1]), consacré à la minimisation de la somme d'une fonction différentielle et d'une fonction convexe, réalisée via l'algorithme **Forward-Backward de métrique variable**, expérimenté sur la reconstruction d'une image flouté.

Table des matières

Chapitre 1

Généralité et Algorithme inexact de Forward-Backward

Tout au long du travail, nous considérons le produit scalaire dans $\mathbb{R}^n$ avec n un entier,

$$\begin{aligned} \langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto \langle x, y \rangle = x_1 y_1 + x_2 y_2 + ... + x_n y_n, \end{aligned}$$

muni de la norme euclidienne définie par :

$$\forall x \in \mathbb{R}^n, \|x\|^2 = \langle x, x \rangle$$

et la norme pondérée par rapport à la matrice U définie par :

$$\forall x \in \mathbb{R}^n, \|x\|_U^2 = \langle Ux, x \rangle = x^T U x \tag{1.0.1}$$

où $U \in \mathbb{R}^{n \times n}$ une matrice symétrique définie positive et x un vecteur colonne de $\mathbb{R}^n$ de la forme :

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

ce qui fait x^T est un vecteur ligne de la forme :

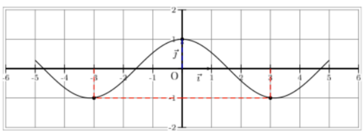
$$x^T = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}$$

qui donne sens au produit scalaire défini dans (1.0.1).

Un problème d'optimisation recherche une variable $x = (x_1, ..., x_n)$ qui minimise une fonction objectif f de $\mathbb{R}^n$ dans $\mathbb{R}$. Ce problème s'écrit :

$$\underset{x \in \mathbb{R}^n}{\text{Argmin}} f(x)$$

Le problème peut avoir plusieurs minima locaux. Le minimiseur global du problème, s'il existe, se note x^* et il n'est pas toujours unique. Par contre le minimum global, noté $f(x^*)$, s'il existe, est unique. C'est l'exemple de la fonction $f(x) = \cos(x)$ sur l'intervalle [-5,5] représentée ci-dessous :

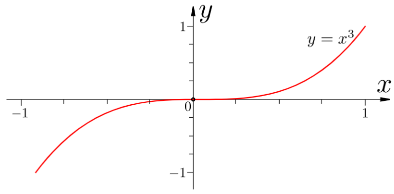


[**Source : image prise sur internet**]

En rouge, le minimum atteint deux fois dans cet intervalle, en $x^* = -\pi$ et $x^* = \pi$ qui vaut $f(-\pi) = f(\pi) = -1$.
Une condition nécessaire pour qu'un point soit un minimiseur est qu'il soit un point stationnaire :

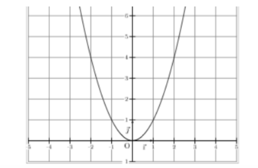
$$\nabla f(x^*) = 0 \tag{1.0.2}$$

Cependant, cette condition n'est pas suffisante, et un point stationnaire n'est pas forcément un minimiseur. Il peut s'agir d'un point d'inflexion ou un minimiseur local ou autre chose.
C'est l'exemple de la fonction $f(x) = x^3$ représenté si dessous :



[Source : image prise sur internet]

La fonction est à gradient nul au point $x^* = 0$ et pourtant $x^* = 0$ n'est pas le minimiseur, plutôt la fonction présente un point d'inflexion au point (0,0).
Si la fonction f est convexe et continue, alors la condition de recherche de minimum (1.0.2) est suffisante. De plus, dans ce cas, un minimiseur local est un minimiseur global et il en est de même pour un minimum. C'est l'exemple de la fonction $f(x) = x^2$ représentée ci-dessous :



[Source : image prise sur internet]

qui admet un minimum global (local) en $x^* = 0$ qui est $f(x) = 0$.
Il est en outre unique pour une fonction strictement convexe.

Les problèmes d'optimisation n'ont en général pas de solution analytique. Cependant, des méthodes de résolution itératives très efficaces sont à notre service, c'est le cas de l'algorithme **Forward-Backward** qui fait l'objet de notre travail, la minimisation de G donné par la somme des fonctions :

$$G = R + F$$

Cette première partie présente le passage de l'algorithme **Forward-Backward de métrique variable** à l'algorithme **inexact Forward-Backward de métrique variable** ainsi que les hypothèses nécessaires pour résoudre le problème considéré. L'algorithme **Forward-Backward** repose sur l'opérateur de proximité, ainsi les hypothèses demandent des notions d'optimisation comme la différentiabilité, le gradient, la sous-différentiablité, la convexité.... Nous allons commencer par énoncer quelques notions préliminaires.

1.1 Définitions

Soit Ω un ouvert de $\mathbb{R}^n$ et f une fonction de Ω dans $\mathbb{R}$:

Définition 1.1.1. *L'ensemble Ω est dite convexe si pour tout x, y éléments de Ω et pour tout $\lambda \in]0, 1[$, on a :*

$$(\lambda x + (1 - \lambda)y) \in \Omega$$

Définition 1.1.2. *On dit que la fonction f est* ***convexe*** *sur Ω, si Ω lui même est convexe et pour tout x, y dans Ω et pour tout λ dans]0,1[,*

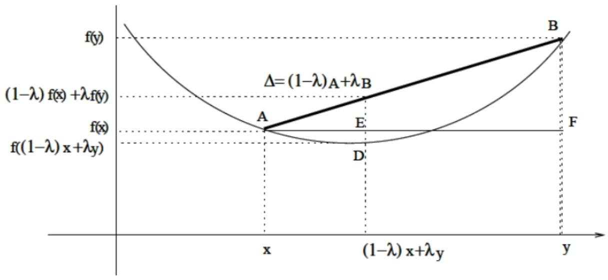
$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda)f(y). \tag{1.1.1}$$

La fonction f est dite fortement convexe sur Ω, de module de forte convexité $c > 0$, lorsque

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda)f(y) - \frac{1}{2}\ c\lambda(1-\lambda)\|x-y\|^2.$$

Lorsque l'inégalité (1.1.1) *est stricte alors on parle de la* ***stricte convexité****.*

la définition des fonctions convexes est naturelle, puisque l'interprétation graphique de la convexité d'une fonction est la suivante :



[**Source : image prise sur internet**]

pour tout couple de points A et B du graphe, le segment $[A, B]$ est situé au dessus du graphe.
Une propriété fondamentale est qu'une fonction strictement convexe admet au plus un minimiseur.

Exemple 1.1.3. *On peut montrer les exemples suivants :*

(a) Toute norme $\|\cdot\|$ sur $\mathbb{R}^n$ est convexe.

(b) Toute forme linéaire (même discontinue en dimension infinie) sur $\mathbb{R}^n$ est convexe.

(c) Les sous-niveaux $L_f(\beta) := \{x \in dom(f), f(x) \leqslant \beta\}$ d'une fonction convexe sont convexes .

(d) La composition des fonctions convexes n'est pas nécessairement convexe, Par exemple :

$$\begin{aligned} f, g : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto f(x) = exp(-x), \quad g(x) = x^2 \end{aligned}$$

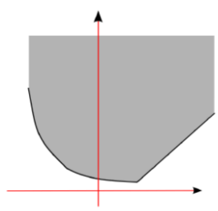
sont convexes alors que

$$\begin{aligned} f \circ g : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto exp(-x)^2 \end{aligned}$$

n'est pas convexe

Définition 1.1.4. *On appelle épigraphe de la fonction f et on le note epi(f) l'ensemble :*

$$epi(f) = \{(x,t) \in \Omega \times \mathbb{R}, \quad t \geqslant f(x)\}$$



[**Source : image prise sur internet**]

L'épigraphe de la fonction est la zone grisée au-dessus du graphe de la fonction (en noir).

Théoreme 1.1.5. *La fonction f est convexe si et seulement si son épigraphe est convexe.*

Preuve :
Supposons que f est convexe et soient (x_1, t_1) et (x_2, t_2) deux éléments de epi(f), on a alors pour tout $\alpha \in [0, 1]$:

$$f(x_1) \leqslant t_1 \quad \text{et} \quad f(x_2) \leqslant t_2, \quad \text{donc :}$$
$$\alpha f(x_1) + (1-\alpha)f(x_2) \leqslant \alpha t_1 + (1-\alpha)t_2$$

Puisque f est convexe alors :

$$f(\alpha x_1 + (1-\alpha)x_2) \leqslant \alpha f(x_1) + (1-\alpha)f(x_2)$$

donc on obtient :

$$f(\alpha x_1 + (1-\alpha)x_2) \leqslant \alpha t_1 + (1-\alpha)t_2$$

implique que :

$$(\alpha x_1 + (1-\alpha)x_2, \alpha t_1 + (1-\alpha)t_2) \in \text{epi}(f)$$

donc epi(f) est un ensemble convexe.
Supposons que epi(f) est convexe.
D'une part, pour x_1 et x_2 dans dom(f), on a :

$$(x_1, f(x_1)) \in \text{epi(f) et } (x_2, f(x_2)) \in \text{epi(f)},$$

alors pour tout $\alpha \in]0, 1[$, par la convexicité de l'épigraphe de f, on obtient :

$$(\alpha x_1 + (1-\alpha)x_2, \alpha f(x_1) + (1-\alpha)f(x_2)) \in \text{epi(f)}$$

qui est équivaut à :

$$f(\alpha x_1 + (1-\alpha)x_2) \leqslant \alpha f(x_1) + (1-\alpha)f(x_2)$$

donc f est convexe.

1.2 Différentiel et gradient

Définition 1.2.1. *La fonction f est* ***Fréchet différentiable*** *sur Ω dans $\mathbb{R}$ lorsque sa dérivée existe en tout point de Ω. Formellement, f est* ***différentiable*** *au sens de* ***Fréchet*** *au point x_0 de Ω, s'il existe un vecteur u de $\mathbb{R}^n$ tel que :*

$$f(x) = f(x_0) + \langle u, x - x_0 \rangle + o(\|x - x_0\|).$$

Le vecteur u, s'il existe, est unique et c'est le gradient de f en x_0 noté $\nabla f(x_0)$. L'égalité précédente est équivalente à :

$$\lim_{x \to x_0} \frac{f(x) - f(x_0) - \langle \nabla f(x_0), x - x_0 \rangle}{\|x - x_0\|} = 0$$

Le gradient de f au point x_0 est un vecteur de $\mathbb{R}^n$, dirigé dans la direction de la plus forte pente et qui est croissante si f est convexe. Les composantes du gradient ne sont rien d'autres que les dérivées partielles de f par rapport à chacune des variables $(x_i)_{1 \leqslant i \leqslant n}$:

$$\nabla f(x) = (\frac{\partial f}{\partial x_1}(x), ..., \frac{\partial f}{\partial x_n}(x))^T.$$

Définition 1.2.2. *La fonction f est de classe C^1 si toutes les dérivées partielles de f existent et sont continues en tout point de Ω.*

Soit l'application g définie par :

$$\begin{aligned} g : \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ x = (x_1, ..., x_n) &\longmapsto g(x) = \nabla f(x) \end{aligned}$$

Définition 1.2.3. *On dit que la fonction f est de classe C^2 en x_0 de Ω si l'application g est de classe C^1 en x_0 et s'il existe une matrice H de $\mathbb{R}^{n\times n}$ tel que :*

$$f(x) = f(x_0) + \langle \nabla f(x_0), x - x_0\rangle + \frac{1}{2}\langle H(x-x_0), x - x_0\rangle + o(\|x - x_0\|^2). \tag{1.2.1}$$

La matrice H, si elle existe, est unique et c'est la hessienne de f en x_0 noté $\nabla^2 f(x_0)$:

$$H : \mathbb{R}^n \longrightarrow \mathbb{M}_{n\times n}$$

$$x = (x_1, ..., x_n) \longmapsto H(x) = \nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) & \vdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}$$

L'égalité (1.2.1) est équivalente à :

$$\lim_{x\to x_0} \frac{f(x) - f(x_0) - \langle \nabla f(x_0), x - x_0\rangle - \frac{1}{2}\langle \nabla^2 f(x_0)(x - x_0), x - x_0\rangle}{\|x - x_0\|^2} = 0.$$

Une remarque fondamentale des fonctions différentiables est qu'elles sont localement approximées par des fonctions affines tangentes, c'est-à-dire si f est différentiable au point x_0, alors au voisinage de x_0, on trouve :

$$f(x) = \langle \nabla f(x_0), x - x_0\rangle + f(x_0) + \varepsilon(x)(x - x_0) \quad \text{où} \quad \lim_{x\to x_0} \varepsilon(x) = 0,$$

donc on a l'équivalence :

$$f(x) \underset{x_0}{\sim} \langle \nabla f(x_0), x - x_0\rangle + f(x_0)$$

Définition 1.2.4. *On dit que la fonction f est **coercive** sur Ω si :*

$$\forall x \in \Omega, \quad \lim_{\|x\|\to+\infty} f(x) = +\infty$$

*Soit α un réel, le **sous-niveau** de f de hauteur α est l'ensemble $L_f(\alpha) = \{x \in \mathbb{R}^n : f(x) \leqslant \alpha\}$.*

Autrement dit, la fonction f est **coercive** sur Ω si pour tout α dans $\mathbb{R}$ le sous-niveau $L_f(\alpha)$ est borné. Une propriété fondamentale de la **coercivité** est que le minimum d'une fonction continue **coercive** existe et il est atteint .
Minimiser f sur Ω équivaut à minimiser f sur n'importe quel sous-ensemble de niveau non vide de f dans Ω. Pour cette raison, les ensembles de niveau jouent un rôle important en optimisation.

Une fonction est continue en un point si et seulement si elle est semi-continue supérieurement et inférieurement en ce point. Nous allons rappeler une proposition sur la semi-continuité inférieure et semi-continuité supérieure :

Proposition 1.2.5. *La fonction f est **semi-continue inférieurement** si l'une des propriétés équivalentes suivantes est vérifiée :*

(i) Pour tout réel α, l'ensemble de sous-niveau $\{x \in \Omega \mid f(x) \leqslant \alpha\}$ est fermé.

(ii) L'épigraphe $\{(x,t) \in \Omega \times \mathbb{R} \mid f(x) \leqslant t\}$ est fermé.

*La fonction f est **semi-continue supérieurement** si l'une des propriétés équivalentes suivantes est vérifiée :*

(i) pour tout réel α, l'ensemble de sur-niveau $\{x \in \Omega \mid f(x) \geqslant \alpha\}$ est fermé.

(ii) l'hypographe $\{(x,t) \in \Omega \times \mathbb{R} \mid f(x) \geqslant t\}$ est fermé.

Théoreme 1.2.6. *(a) Soit f différentiable sur Ω, f est convexe sur Ω si et seulement si pour tout (x,y) dans Ω^2,*

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

(b) Soit f de classe C^2 sur Ω, f est convexe si et seulement si en tout point x de Ω, la matrice hessienne est semi-définie positive, c'est à dire que toutes ses valeurs propres sont positives.

(c) Soit f de classe C^2 sur Ω, f est strictement convexe si et seulement si la matrice hessienne est définie positive presque partout dans Ω, c'est à dire que toutes ses valeurs propres sont strictement positives presque partout. L'exemple de la fonction f définie :

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto x^4 \end{aligned}$$

qui est une fonction strictement convexe et que zéro est une valeur propre de la fonction f.

(d) f est fortement convexe sur Ω de module de forte convexité c si et seulement si la plus petite valeur propre de la matrice hessienne est minorée par c sur Ω, soit encore :

$$(\forall x \in \Omega), \quad c\|y\|^2 \leqslant \quad \langle \nabla^2 f(x) y, y \rangle$$

1.3 Sous-gradient et sous-différentiel

Définition 1.3.1. *Soit $f : \mathbb{R}^n \to (-\infty, +\infty]$, le domaine de f est l'ensemble noté dom(f), défini par :*

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

On dit que la fonction f est propre si $dom(f)$ est non vide.

Définition 1.3.2. *Soit $f : \mathbb{R}^n \to (-\infty, +\infty]$ une fonction convexe. Un vecteur $\eta \in \mathbb{R}^n$ est appelé **sous-gradient** de f au point $x_0 \in$ dom(f) si :*

$$\forall x \in dom(f), \quad f(x) - f(x_0) \geqslant \; < \eta, x - x_0 >$$

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

Lemme 1.3.3. *Soit $f : \mathbb{R}^n \to (-\infty, +\infty]$ une fonction convexe, le **sous-différentiel** de f en x_0 défini par :*

$$\partial f(x_0) = \{\eta \in \mathbb{R}^n, \quad \forall x \in dom(f) \quad f(x) - f(x_0) \geqslant \; < \eta, x - x_0 >\}$$

est un ensemble convexe fermé

Preuve :

Soit $(\eta_1, \eta_2) \in \partial f(x_0)$, donc pour tout x de $\mathbb{R}^n$ on a :

$$\begin{aligned} f(x) - f(x_0) &\geqslant \; < \eta_1, x - x_0 > \\ f(x) - f(x_0) &\geqslant \; < \eta_2, x - x_0 > \end{aligned}$$

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

$$\begin{aligned} f(x) - f(x_0) &\geqslant \alpha \langle \eta_1, x - x_0 \rangle + (1-\alpha) \langle \eta_2, x - x_0 \rangle \\ &\geqslant \langle \alpha \eta_1 + (1-\alpha) \eta_2, x - x_0 \rangle \end{aligned}$$

donc par définition,

$$\alpha\eta_1 + (1-\alpha)\eta_2 \in \partial f(x_0)$$

Alors $\partial f(x_0)$ est un ensemble convexe.
Soit $(\eta_n)_{n\in\mathbb{N}}$ une suite de $\partial f(x_0)$ qui converge vers η. Par définition, on a :

$$\forall x \in \mathrm{dom}(f), \quad f(x) - f(x_0) \geqslant \; < \eta_n, x - x_0 > \quad \text{pour tout} \quad n \in \mathbb{N}$$

En passant à la limite en n, on obtient facilement que :

$$\forall x \in \mathrm{dom}(f), \quad f(x) - f(x_0) \geqslant \; < \eta, x - x_0 >$$

ainsi la limite $\eta \in \partial f(x_0)$. Alors $\partial f(x_0)$ est un ensemble fermé.

L'interprétation géométrique du **sous-différentiel** au point x_0 est la suivante : Il est formé par toutes les directions des hyperplans qui passent par le point $(x_0, f(x_0))$ et restent **sous le graphe** de la fonction f. Ces hyperplans sont appelés hyperplans support ou hyperplans d'appui au graphe de f en x_0.

Exemple 1.3.4. *Calculons le* ***sous-différentiel*** *de la fonction f, en tout point x, définie par :*

$$f : \mathbb{R} \longrightarrow \mathbb{R}$$
$$x \longmapsto f(x) = |x|$$

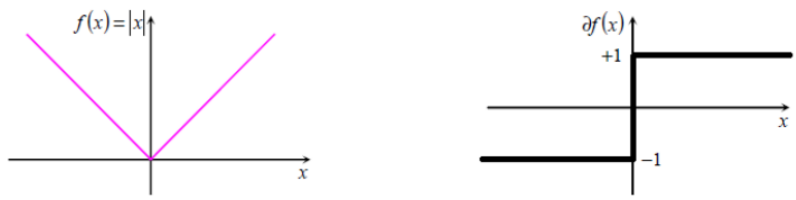
Le problème se pose au point 0 où la fonction n'est pas différentielle.

Analytiquement *:*

On a pour $x = 0$:

$$\begin{aligned}
\eta \in \partial f(0) &\Longleftrightarrow f(x) - f(0) \geqslant \langle \eta, x - 0 \rangle \quad \forall x \in \mathbb{R} \\
&\Longleftrightarrow |x| \geqslant \eta \cdot x \qquad \forall x \in \mathbb{R} \\
&\Longleftrightarrow \begin{cases} x \geqslant \eta \cdot x & si \;\; x > 0 \\ -x \geqslant \eta \cdot x & si \;\; x < 0 \end{cases} \\
&\Longleftrightarrow \eta \leqslant 1 \quad et \quad \eta \geqslant -1
\end{aligned}$$

ce qui donne en fin l'intervalle $\partial f(0) = [-1, 1]$.



[**Source : image prise sur internet**]

On obtient :

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

Géométriquement :
Dans $\mathbb{R}^2$, les hyperplans d'appui sont des droites et les sous-gradients associés leurs coefficients directeurs.

On peut montrer la proposition suivante :

Proposition 1.3.5. *Soit f de classe C^1 sur U ouvert de $\mathbb{R}^n$, alors pour tout $x_0 \in U$,*

$$\partial f(x_0) = \{\nabla f(x_0)\}$$

Remarque 1.3.6. *Le* ***sous-différentiel*** *peut ne pas exister sur le bord du domaine. Par exemple la fonction f définie par :*

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto f(x) = -\sqrt{x} \end{aligned}$$

est convexe, et le sous-différentiel n'existe pas en 0 car $\lim_{x \to 0^+} f'(x) = +\infty$.

1.4 Ensembles minimiseurs

Définition 1.4.1. *L'ensemble des minimiseurs de f sur un sous ensemble K de Ω est noté $\underset{K}{Argmin} f$, défini par :*

$$\underset{K}{Argmin} f = \{x \in K : f(x) \leq f(y), \quad \forall y \in K\}.$$

Théoreme 1.4.2. *Une condition nécessaire pour que x^* soit un minimiseur de f est que x^* soit un point critique de f, c'est à dire :*

a) $0 \in \partial f(x^*)$ si f n'est pas différentiable sur K et $x^* \in$ int K.

b) $\nabla f(x^*) = 0$ si f est différentiable sur K et $x^* \in$ int K.

Quand f est une fonction convexe alors cette condition est suffisante.

On peut démontrer le théorème fondamental de l'existence d'une solution d'un problème de minimisation suivant :

Théoreme 1.4.3. *Soit V un espace de Banach et K une partie de V. On considère le problème de minimisation suivant :*

$$(\boldsymbol{P}) \quad \underset{K}{min} \quad f$$

Si K est un ensemble fermé non vide, f semi-continue inférieurement sur V et coercive sur K, alors le problème $(\boldsymbol{P})$ admet au moins une solution.

1.5 L'opérateur de proximité :

L'opérateur de proximité est en lien étroit avec l'enveloppe de Moreau d'une fonction f de $\mathbb{R}^n$ dans $\bar{\mathbb{R}}$ propre, convexe, semi-continue inférieurement et d'indice λ de $]0, +\infty[$ défini par :

$$\begin{aligned} M_{\lambda f} : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ y &\longmapsto M_{\lambda f}(y) = \min_{x \in \mathbb{R}^n} \{f(x) + \frac{1}{2\lambda}\|y - x\|^2\} \end{aligned}$$

En fait, en rajoutant le terme carré, on obtient une fonction strictement convexe. L'enveloppe de Moreau fournit donc le minimum d'une version régularisée de f.
La version régularité est emportée par la coercivité de la forme quadratique (terme au carré), ce qui fait que cette version régularisée est coercive. La forme quadratique est strictement convexe et la fonction f est convexe, ce qui fait que la version régularisée est strictement convexe. Ainsi Cette version régularisée de f coercive, continue et strictement convexe, en conséquence le problème atteint sa borne inférieure en un unique point x noté $\mathbf{prox}_{\lambda f}(y)$, appelé point proximal défini par :

$$\begin{aligned} \text{prox}_{\lambda f} : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ y &\longmapsto \text{prox}_{\lambda f}(y) = \underset{x \in \mathbb{R}^n}{\text{Argmin}} \{f(x) + \frac{1}{2\lambda}\|y - x\|^2\} \end{aligned}$$

Pour une matrice U symétrique définie positive, l'opérateur de proximité de la fonction f par rapport à la matrice U est défini par :

$$\begin{aligned} \mathrm{prox}_{U,f} : \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ y &\longmapsto \mathrm{prox}_{U,f}(y) = \underset{x\in\mathbb{R}^n}{\mathrm{Argmin}}\{f(x) + \frac{1}{2}\|y-x\|_U^2\} \end{aligned}$$

où $\|.\|_U$ est la norme associée à la matrice U.
De cette façon, l'application $\mathbf{prox}_{\lambda f}$ peut-être vu comme l'application $\mathbf{prox}_{I,\lambda f}$ définie par :

$$\begin{aligned} \mathrm{prox}_{I,\lambda f} : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ y &\longmapsto \mathrm{prox}_{I,\lambda f}(y) = \underset{x\in\mathbb{R}^n}{\mathrm{Argmin}}\{f(x) + \frac{1}{2\lambda}\|y-x\|_I^2\} \end{aligned}$$

où I est la matrice identité de $\mathbb{R}^n$.

La résolution de l'opérateur de proximité est la suivante :

$$\begin{aligned} &Z = \mathrm{prox}_{U,f}(y) \\ \Longleftrightarrow\ &Z \in \underset{x\in\mathbb{R}^n}{\mathrm{Argmin}}\{f(x) + \frac{1}{2}\|y-x\|_U^2\} \\ \Longleftrightarrow\ &Z = \inf_{x\in\mathbb{R}^n}\{f(x) + \frac{1}{2}\|y-x\|_U^2\} \end{aligned}$$

Deux cas peuvent se présenter, cas où :
1) f est différentiable :

$$Z = \inf_{x\in\mathbb{R}^n}\{f(x) + \frac{1}{2}\|y-x\|_U^2\} \quad \text{si et seulement si} \quad \nabla f(Z) + \nabla(\frac{1}{2}\|y-Z\|_U^2) = 0$$
$$\text{U symétrique, implique que} \quad \boxed{\nabla f(Z) + U(Z-y) = 0}$$

2) f n'est pas différentiable :

$$Z = \inf_{x\in\mathbb{R}^n}\{f(x) + \frac{1}{2}\|y-x\|_U^2\} \quad \text{si et seulement si} \quad 0 \in \partial f(Z) + \nabla(\frac{1}{2}\|y-Z\|_U^2)$$
$$\text{équivaut à :} \quad \exists r(z) \in \partial f(Z) \quad \text{tel que} \quad 0 = r(Z) + \nabla(\frac{1}{2}\|y-Z\|_U^2)$$
$$\text{U symétrique, implique que} \quad \boxed{0 = r(Z) + U(Z-y)}$$

Une propriété fondamentale de l'opérateur de proximité est que ses points fixes sont les minimiseurs de la fonction f, en effet :
soit y un vecteur de $\mathbb{R}^n$ tel que :

$$\mathrm{prox}_{U,f}(y) = y \Longleftrightarrow y = \inf_{x\in\mathbb{R}^n}\{f(x) + \tfrac{1}{2}\|y-x\|_U^2\}$$

$$y = \inf_{x\in\mathbb{R}^n}\{f(x) + \frac{1}{2}\|y-x\|_U^2\} \quad \text{si et seulement si} \quad \nabla f(y) + U(y-y) = 0$$
$$\text{équivaut à} \quad \nabla f(y) = 0$$

Comme f est convexe alors cette dernière égalité montre que y est un minimiseur.

Une autre interprétation est de voir le point proximal comme un compromis entre la minimisation de la fonction f et la proximité à la valeur y. Cet équilibre est modulé par le paramètre λ. Plus celui-ci sera petit, plus le résultat se rapproche de y.

Vous pouvez voir l'opérateur de proximité sur ([[2], P.54, Sec.5.2.2])

1.6 Algorithme Forward-Backward

Le problème qui se pose tout au long du travail est le suivant :

$$\text{trouver} \quad y \in \text{Argmin} \quad G$$

où $G : \mathbb{R}^n \longrightarrow (-\infty, +\infty]$ est une fonction coercive. En plus, on impose que la fonction G soit une somme de :

$$G = F + R$$

Les fonctions F et R doivent satisfaire les hypothèses suivantes :

Hypothèse 1. Régularité :

(i) $R : \mathbb{R}^n \to (-\infty, +\infty]$ est propre, sémi-continue inférieurement, convexe et sa restriction à son domaine est continue.

(ii) $F : \mathbb{R}^n \to R$ est différentiable. En plus F est à gradient L-Lipschitz dans dom$R$, où $L > 0$, c'est à dire :

$$\forall (x, y) \in (\text{dom}R)^2, \quad \|\nabla F(x) - \nabla F(y)\| \leq L\|x - y\|.$$

(iii) G définie par $G = F + R$, est coercive.

Un tel problème peut être résolu avec l'algorithme **Forward-Backward** (**FB**) :

$$(\mathbf{FB}) \begin{cases} x_0 \in \mathbb{R}^n \quad \text{fixé} \\ \text{pour} \quad k \geqslant 0 : \\ \qquad y_k = \text{prox}_{\gamma_k R}(x_k - \gamma_k \nabla F(x_k)) \\ \qquad x_{k+1} = x_k + \lambda_k (y_k - x_k) \end{cases}$$

où pour tout $k \in \mathbb{N}$, on a $(\gamma_k, \lambda_k) \in (0, +\infty)^2$.
Pour un vecteur initial x_0 donné, à l'itération k, on a deux étapes :

1) La première étape qui évalue l'opérateur de proximité selon R, renvoie à une recherche d'un minimum de R suivant une direction de descente de gradient selon F, γ_k étant le pas de descente.
2) La seconde étape détermine le vecteur x_{k+1} suivant un processus dit **avant-arrière** puisque en partant de x_k, on fait un **un pas en avant (Forward)** de $\lambda_k y_k$ et un **pas en arrière (Backward)** de $(-\lambda_k x_k)$, le vecteur x_{k+1} est donné par :

$$x_{k+1} = x_k + \lambda_k y_k - \lambda_k x_k$$

d'où le nom algorithme **Forward-Backward**.

Remarque 1.6.1. *a) Comme $G = F + R$ alors dom$G$ = dom$F \cap$ dom$R$ et selon l'hypothèse (ii) de **régularité**, dom$G$ = dom$R$. Selon l'hypothèse (i) de **régularité**, R est propre et convexe, alors dom$G$ est un ensemble convexe non vide.*

b) La fonction F est à gradient L-Lipschitz donc continue a priori semi-continue inférieurement et R est semi-continue inférieurement alors G est semi-continue inférieurement et sa restriction à son domaine est continue.

L'hypothèse qui suit jouera un rôle central dans la deuxième partie sur la convergence de l'algorithme.

Hypothèse 2. inégalité KL :
La fonction G satisfait l'inégalité **Kurdyka-Lojasiewicz**, c'est à dire pour tout $\xi \in \mathbb{R}$ et pour tout sous-ensemble E fermé de $\mathbb{R}^n$, il existe trois constantes $\kappa > 0$, $\zeta > 0$ et $\theta \in [0, 1)$ telles que

$$\forall x \in E \quad \text{tel que} \quad |G(x) - \xi| \leq \zeta, \quad (\forall t(x) \in \partial G(x)), \quad \kappa |G(x) - \xi|^\theta \leqslant \|t(x)\|$$

Exemple 1.6.2. *Soit G la fonction définie par :*

$$\begin{aligned} G :& \mathbb{R} \longrightarrow \mathbb{R} \\ & x \longmapsto G(x) = x^2 \end{aligned}$$

Pour tout $\xi \in \mathbb{R}$, On cherche des constant $\kappa > 0$, $\zeta > 0$ et $\theta \in [0,1)$ telles que pour tout $x \in \mathbb{R}$:

$$|x^2 - \xi| \leq \zeta \implies \kappa |x^2 - \xi|^{\theta} \leqslant 2|x|$$

1. pour $\xi = 0$, on a :

$$|x^2| \leq \zeta, \quad \kappa |x|^{2\theta} \leqslant 2|x|$$

   on peut prendre $\boxed{\zeta = 1}$ et $\boxed{\theta = \frac{1}{2}}$, donc pour κ :

$$\kappa |x| \leqslant 2|x| \implies \kappa \leqslant 2$$

   donc soit $\boxed{\kappa = 2}$.

1.7 Métrique de Majoration-Minoration

Des matrices servant à définir une métrique variable joueront un rôle central dans l'algorithme **Forward-Backward**. En fait, avec l'algorithme **Forward-Backward** on résout un problème de minimisation assez compliqué à chaque itération dont le coût de calcul fait souffrir l'algorithme d'une convergence lente. Une stratégie d'accélération est l'utilisation de la métrique variable. Plus précisément, soit $(x_k)_{k \in \mathbb{N}}$ une suite donnée dans dom$R$ et soit $(A_k)_{k \in \mathbb{N}}$ une suite des matrices symétriques de $\mathbb{R}^{n \times n}$ et on considère l'algorithme **Forward-Backward de métrique variable (MVFB)** :

$$\textbf{(MVFB)} \begin{cases} x_0 \in \mathbb{R}^n \quad \text{fixé} \\ \text{pour} \quad k \geqslant 0 : \\ \qquad y_k = \operatorname{prox}_{\gamma_k^{-1} A_k, R}(x_k - \gamma_k A_k^{-1} \nabla F(x_k)) \\ \qquad x_{k+1} = x_k + \lambda_k (y_k - x_k) \end{cases}$$

La suite des matrices $(A_k)_{k \in \mathbb{N}}$ remplissent les **conditions de majoration** suivantes :

Hypothèse 3. Majoration :

(i) Pour tout $k \in \mathbb{N}$, la fonction quadratique définie par :

$$\forall x \in \mathbb{R}^n, \quad Q(x, x_k) = F(x_k) + \langle \nabla F(x_k), x - x_k \rangle + \frac{1}{2} \|x - x_k\|_{A_k}^2$$

est un majorant de F dans dom$R$, c'est à dire :

$$\forall x \in \mathrm{dom} R, \quad F(x) \leq Q(x, x_k)$$

(ii) Il existe $(\underline{\nu}, \bar{\nu}) \in (0, +\infty)^2$ telles que,

$$\forall k \in \mathbb{N}, \quad \underline{\nu} I_n \preceq A_k \preceq \bar{\nu} I_n$$

De plus $(\gamma_k)_{k \in \mathbb{N}}$ et $(\lambda_k)_{k \in \mathbb{N}}$ sont des suites réelles positives qui satisfont les hypothèses suivantes :

Hypothèse 4. Paramètres :

(i) Il existe $(\underline{\eta}, \bar{\eta}) \in (0, +\infty)^2$ tel que, pour tout $k \in \mathbb{N}$, $\underline{\eta} \leqslant \gamma_k \lambda_k \leqslant 2 - \bar{\eta}$.

(ii) Il existe $\underline{\lambda} \in (0, +\infty)$ tel que, pour tout $k \in \mathbb{N}$, $\underline{\lambda} \leqslant \lambda_k \leqslant 1$.

Le lemme suivant montre l'existence d'une telle suite des matrices :

Lemme 1.7.1. *On suppose que l'hypothèse de **Régularité** est vraie et pour tout k de $\mathbb{N}$, A_k égale à LI_n, avec $L > 0$ la constante de Lipshitz du gradient de F. Alors $(A_k)_{k\in\mathbb{N}}$ satisfait l'hypothèse de **majoration** avec $\underline{\nu} = \bar{\nu} = L$.*

Preuve :
Soit g une fonction dérivable définie sur $\mathbb{R}$ telle que g' est K-Lipschitz,

$$
\begin{aligned}
g(1) &= g(0) + \int_0^1 g'(t)dt \\
&= g(0) + g'(0) + \int_0^1 (g'(t) - g'(0))dt \\
&\leqslant g(0) + g'(0) + \int_0^1 |g'(t) - g'(0)|dt \\
&\leqslant g(0) + g'(0) + \int_0^1 K|t-0|dt \\
g(1) &\leqslant g(0) + g'(0) + \frac{K}{2}
\end{aligned}
$$

Soit $(y; z) \in (\mathbb{R}^n)^2$, pour tout $t \in [0; 1]$ on pose :

$$v = x - y, \quad y_t = y + t(x-y) \quad \text{et} \quad g(t) = F(y_t)$$

La fonction g est dérivable et $g'(t) = \langle \nabla F(y_t), v\rangle$.

$$
\begin{aligned}
|g'(t_1) - g'(t_2)| &= |\langle \nabla F(y_{t_1}), v\rangle - \langle \nabla F(y_{t_2}), v\rangle| \\
&= |\langle \nabla F(y_{t_1}) - \nabla F(y_{t_2}), v\rangle| \\
&\leqslant \|v\| \|\nabla F(y_{t_1}) - \nabla F(y_{t_2})\| \\
&\leqslant \|v\| L \|y_{t_1} - y_{t_2}\| \\
&= \|v\| L \|(t_1 - t_2)(x-y)\| \\
&= \|v\|^2 L |t_1 - t_2|
\end{aligned}
$$

Alors g' est à gradient K-Lipschitz avec $K = L\|v\|^2$, donc on obtient :

$$
\begin{aligned}
g(1) &\leqslant g(0) + g'(0) + \frac{K}{2} \\
F(y_1) &\leqslant F(y_0) + \langle \nabla F(y_0), x-y\rangle + \frac{L}{2}\|x-y\|^2 \\
F(x) &\leqslant F(y) + \langle \nabla F(y), x-y\rangle + \frac{1}{2}\|x-y\|^2_{LI_n}
\end{aligned}
$$

Posons $A_k = LI_n$, l'hypothèse tient bel et bien.

1.8 Algorithme inexact Forward-Backward de métrique variable

En général, l'opérateur de proximité par rapport à une métrique arbitraire n'a pas une expression fermée (explicite) pour calculer sa valeur, c'est le cas de l'algorithme **(MVFB)** qui, par l'opérateur **prox**, on a :

$$y_k = \operatorname{prox}_{\gamma_k^{-1}A_k, R}(x_k - \gamma_k A_k^{-1} \nabla F(x_k))$$

qui est équivalent à résoudre le problème :

$$\inf_{x\in\mathbb{R}^n} V(x) \tag{1.8.1}$$

où V est la fonction définie par :

$$
\begin{aligned}
V :& \mathbb{R}^n \longrightarrow \mathbb{R} \\
& x \longmapsto V(x) = R(x) + \frac{1}{2}\|x - (x_k - \gamma_k A_k^{-1}\nabla F(x_k))\|^2_{\gamma_k^{-1}A_k}
\end{aligned}
$$

Minimiser la fonction V revient à minimiser une forme régularisée de la fonction R. Un tel problème n'est généralement pas facile à résoudre. Pour contourner cette difficulté, on propose une autre façon de voir le problème :
La fonction R n'est pas nécessairement différentiable donc V en est ainsi, alors y_k est solution du problème (1.8.1) si et seulement si $0 \in \partial V(y_k)$, c'est-à-dire :

$$\begin{aligned}
&0 \in \{\partial R(y_k) + \frac{1}{2} \times 2 \times \gamma_k^{-1} A_k(y_k - x_k + \gamma_k A_k^{-1} \nabla F(x_k))\} \\
&\iff 0 \in \{\partial R(y_k) + \gamma_k^{-1} A_k(y_k + x_k) + \nabla F(x_k)\} \\
&\iff \quad \exists r(y_k) \in \partial R(y_k), \quad \text{tel que} \quad r(y_k) + \gamma_k^{-1} A_k(y_k - x_k) + \nabla F(x_k) = 0 \\
&\iff \quad \nabla F(x_k) + r(y_k) = -\gamma_k^{-1} A_k(y_k - x_k)
\end{aligned}$$

donc on obtient :

$$\boxed{\|\nabla F(x_k) + r(y_k)\| = \gamma_k^{-1} \|A_k(y_k - x_k)\|} \tag{1.8.2}$$

Selon l'hypothèse (ii) de **majoration** :

$$\begin{aligned}
\exists \quad (\underline{\nu}, \bar{\nu}) \in (0, +\infty), &\quad \underline{\nu} I_n \preceq A_k \preceq \bar{\nu} I_n \\
\text{donc :} &\quad \frac{1}{\bar{\nu}} I_n \preceq A_k^{-1} \preceq \frac{1}{\underline{\nu}} I_n \\
\text{en particulier :} &\quad I_n \preceq \bar{\nu} A_k^{-1} \\
\text{donc} \quad (\forall x \in \mathbb{R}^n), &\quad \langle I_n x, x \rangle \leqslant \langle \bar{\nu} A_k^{-1} x, x \rangle
\end{aligned}$$

donc on obtient :

$$\begin{aligned}
\|A_k(y_k - x_k)\|^2 &= \langle A_k(y_k - x_k), A_k(y_k - x_k) \rangle \\
&\leqslant \bar{\nu} \langle A_k^{-1} A_k(y_k - x_k), A_k(y_k - x_k) \rangle \\
&= \bar{\nu} \langle A_k(y_k - x_k), y_k - x_k \rangle \\
&= \bar{\nu} \|y_k - x_k\|_{A_k}^2
\end{aligned}$$

donc on trouve :

$$\|A_k(y_k - x_k)\| \leqslant \sqrt{\bar{\nu}} \|y_k - x_k\|_{A_k}$$

Alors à partir de l'égalité (1.8.2), on obtient :

$$\|\nabla F(x_k) + r(y_k)\| \leqslant \gamma_k^{-1} \sqrt{\bar{\nu}} \|y_k - x_k\|_{A_k} \tag{1.8.3}$$

A partir de l'hypothèse des **paramètres**, on trouve :

$$\begin{aligned}
\exists \quad (\underline{\eta}, \underline{\lambda}) \in (0, +\infty)^2, &\quad \underline{\eta} \leqslant \gamma_k \lambda_k \quad \text{et} \quad \underline{\lambda} \leqslant \lambda_k \leqslant 1 \\
\text{donc :} &\quad \underline{\eta} \leqslant \gamma_k \\
\text{ainsi :} &\quad \gamma_k^{-1} \leqslant \underline{\eta}^{-1}
\end{aligned}$$

Donc l'inégalité (1.8.3) devient :

$$\|\nabla F(x_k) + r(y_k)\| \leqslant \underline{\eta}^{-1} \sqrt{\bar{\nu}} \|y_k - x_k\|_{A_k}$$

On pose $\tau = \underline{\eta}^{-1} \sqrt{\bar{\nu}}$ où $\bar{\nu}$ est la plus grande valeur propre de la matrice A_k et on obtient :

$$\boxed{\|\nabla F(x_k) + r(y_k)\| \leqslant \tau \|y_k - x_k\|_{A_k}}$$

Bilan :
En fait, on cherche y_k solution du problème (1.8.1), $r(y_k)$ élément du sous-différentielle de $R(y_k)$, c'est-à-dire :

$$V(y_k) \leqslant V(x_k)$$

qui implique que :

$$\boxed{\|\nabla F(x_k) + r(y_k)\| \leqslant \tau \|y_k - x_k\|_{A_k}} \tag{1.8.4}$$

Ainsi on peut voir l'algorithme **(MVFB)** comme la résolution du problème (1.8.1) tel que l'inégalité (1.8.4) est vraie. L'inégalité (1.8.4) est une condition d'arrêt.

On obtient alors une version équivalente dite la **version inexacte** de l'algorithme **Forward-Backward de métrique Variable** suivante :

$$\begin{cases} \tau = \underline{\eta}^{-1}\sqrt{\nu}, \quad x_0 \in \mathrm{dom} R \quad \text{fixés} \\ \text{pour} \quad k \geqslant 0 : \\ \quad \text{trouver} \quad y_k, \quad r(y_k) \in \partial R(y_k), \quad V(y_k) \leqslant V(x_k) \quad \text{tel que} \\ \qquad \|\nabla F(x_k) + r(y_k)\| \leqslant \tau \|y_k - x_k\|_{A_k} \\ \quad x_{k+1} = (1 - \lambda_k)x_k + \lambda_k y_k \end{cases}$$

La question se pose sur la recherche de y_k solution du problème :

$$\inf_{x \in \mathbb{R}^n} V(x)$$

On applique la méthode de **la plus grande pente avec recherche linéaire** au problème (1.8.1), voir ([Chp.2 discrétisation de la méthode de la plus grande pente, [6]]), c'est-à-dire :

$$\begin{cases} x_k \in \mathbb{R}^n \quad \text{fixé selon} \quad k \\ \text{pour} \quad n \geqslant 0 : \\ \qquad z_0 \quad \text{fixé,} \quad \text{on pose} \quad z_0 = x_k \\ \qquad r(z_n) \in \partial R(z_n), \quad v_n \in \partial V(z_n) : \\ \qquad\quad v_n = r(z_n) + \gamma_k^{-1} A_k (z_n - x_k) + \nabla F(x_k) \\ \qquad z_{n+1} = z_n - t_n v_n \end{cases}$$

où t_n est donné par la **méthode d'Armijo** suivante :
On pose la fonction :

$$\begin{aligned} j_n : \mathbb{R} &\longrightarrow \mathbb{R} \\ t &\longmapsto j_n(t) = V(z_n - t v_n) \end{aligned}$$

donc on a :

$$j_n^{'}(t) = -v_n \cdot \partial V(z_n - t v_n)$$

ainsi on trouve :

$$j_n(0) = V(z_n) \quad \text{et} \quad j_n^{'}(0) = -\langle v_n, v_n \rangle$$

Soit $c_1 \in]0,1[$ et $\beta > 0$, **le pas d'Armijo** est donné par :

$$t_n = \frac{\beta}{2^n}$$

le plus grand possible tel que :

$$j_n(t_n) \leqslant j_n(0) + c_1 t_n j_n^{'}(0)$$

En suite on pose $y_k = z_n$ et $r(y_k) = r(z_n)$.

Finalement, on considère la **version inexacte** de l'algorithme **Forward-Backward de métrique Variable (VIFBMV)** donné par :

$$
\textbf{(VIFBMV)}\begin{cases}
\tau = \underline{\eta}^{-1}\sqrt{\bar{\nu}}, \quad x_0 \in \mathrm{dom}R \ \text{ fixés} \\
\text{pour } \ k \geqslant 0 : \\
\qquad z_0 \ \text{ fixé : soit } \ z_0 = x_k \\
\qquad \text{pour } \ n \geqslant 0 : \\
\qquad\qquad r(z_n) \in \partial R(z_n) \ \text{ tel que :} \\
\qquad\qquad\qquad \|\nabla F(x_k) + r(z_n)\| \leqslant \tau \|z_n - x_k\|_{A_k} \\
\qquad\qquad z_{n+1} = z_n - t_n v_n \\
\qquad \text{on pose } \ y_k = z_n \ \text{ et } \ r(y_k) = r(z_n) \\
\qquad x_{k+1} = (1-\lambda_k)x_k + \lambda_k y_k
\end{cases}
$$

Dans la pratique, on peut voir l'algorithme **(VIFBMV)** comme :

$$
\begin{cases}
\tau = \underline{\eta}^{-1}\sqrt{\bar{\nu}}, \\
A = LI_n, \\
\beta > 0 \ \text{ et } \ c_1 \in]0,1[\ \text{ fixés}, \\
x_0 \ \text{ fixé}, \\
\text{def armijo}(y, x, A) : \\
\qquad j(0) = v(y, x, A), \\
\qquad j'(0) = -\text{np.vdot}(\partial V(y, x, A), \partial V(y, x, A)) \\
\qquad t = \beta \\
\quad \text{while } \ V(y - t * \partial V(y, x, A)) > (j(0) + t * c_1 * j'(0)) : \\
\qquad\quad t = t/2 \\
\quad \text{return } \ t \\
\text{norme} = 1, \\
x = x_0 \\
\text{while } \ \text{norme} > 10^{-6} : \\
\qquad z_0 = x \\
\qquad \text{while } \ \|\nabla F(z_n) - r_n\| > \tau \|z_n - x_k\| : \\
\qquad\qquad\quad z_{n+1} = z_n - \text{armijo(z,x,A)} * \partial V(z_n, x, A) \\
\quad y_k = z \\
\quad p = \lambda_k(y_k - x_k) \\
\quad x_{k+1} = x_k + p \\
\quad \text{norme} = \text{la.norme(p)}
\end{cases}
$$

Remarque 1.8.1. *Pour tout $k \in \mathbb{N}$, x_k et y_k appartiennent tout les deux à dom$R$.*

En effet, selon l'hypothèse des **paramètre** $\lambda_k \in (0,1]$, on a aussi :

$$\forall k \in \mathbb{N}, \quad x_{k+1} = (1-\lambda_k)x_k + \lambda_k y_k.$$

D'après l'hypothèse de **régularité** R est convexe, donc on obtient :

$$R(x_{k+1}) = R((1-\lambda_k)x_k + \lambda_k y_k) \leqslant (1-\lambda_k)R(x_k) + \lambda_k R(y_k),$$

qui implique que x_k et y_k appartiennent tout les deux à dom$R$ pour tout k dans $\mathbb{N}$.

Hypothèse 5. Faible convexité :
Il existe $\underline{\alpha} \in (0,1]$ tel que, pour tout $k \in \mathbb{N}$,

$$G(x_{k+1}) \leqslant (1-\underline{\alpha})G(x_k) + \underline{\alpha}G(y_k)$$

Corollaire 1.8.2. *Selon les hypothèses de* ***régularité****,* ***majoration*** *et des* ***paramètre****, si G est convexe sur $[x_k, y_k]$ pour tout $k \in \mathbb{N}$, alors l'hypothèse de la* ***Faible convexité*** *est satisfaite.*

Preuve :

Le vecteur x_{k+1} est donné par :

$$x_{k+1} = (1 - \lambda_k)x_k + \lambda_k y_k$$

donc si G est convexe sur $[x_k, y_k]$, alors on obtient :

$$G(x_{k+1}) \leqslant (1 - \lambda_k)G(x_k) + \lambda_k G(y_k). \tag{1.8.5}$$

Selon l'hypothèse sur les **paramètres** :

$$\exists \ \underline{\lambda} \in (0, +\infty), \quad \lambda_k \in [\underline{\lambda}, 1] \implies \lambda_k \in (0, 1]$$

On pose $\underline{\alpha} = \lambda_k \in (0, 1]$ pour tout $k \in \mathbb{N}$ et selon l'inégalité (1.8.5), on trouve :

$$G(x_{k+1}) \leqslant (1 - \underline{\alpha})G(x_k) + \underline{\alpha}G(y_k).$$

Lorsque G satisfait à une propriété de convexité, l'hypothèse de la **faible convexité** est vrai en considérant $\underline{\alpha}$ le paramètre de relaxation λ_k de l'algorithme.

Chapitre 2

Convergence de l'algorithme Forward-Backward

Cette partie présente l'étude de la convergence de l'algorithme. Après avoir vu les hypothèses nécessaires, nous commençons par énoncer des lemmes qui seront utiles sur la convergence de l'algorithme.

2.1 Variation de la fonction G en x_k et y_k

La présente section rassemble quelques résultats concernant le comportement des suites $(G(x_k))_{k\in\mathbb{N}}$ et $(G(y_k))_{k\in\mathbb{N}}$ générées par l'algorithme inexact, qui seront utiles pour prouver la convergence de l'algorithme. Nous commençons par le lemme suivant :

Lemme 2.1.1. *Sous les hypothèses de* ***régularité****, de* ***majoration*** *et des* ***paramètres****, il existe $\mu_1 \in (0, +\infty)$ tel que pour tout $k \in \mathbb{N}$:*

$$\begin{aligned} G(x_{k+1}) &\leqslant G(x_k) - \frac{\mu_1}{2}\|x_{k+1} - x_k\|^2 \quad (1) \\ &\leqslant G(x_k) - \underline{\lambda}^2\frac{\mu_1}{2}\|y_k - x_k\|^2 \quad (2) \end{aligned}$$

Preuve :
Le vecteur x_{k+1} donné par l'algorithme **inexact Forward-Backward de métrique variable** est défini par :

$$x_{k+1} = (1 - \lambda_k)x_k + \lambda_k y_k, \tag{2.1.1}$$

d'après l'hypothèse sur les **paramètres** :

$$\exists \;\; \underline{\lambda} \in (0, +\infty), \quad k \in \mathbb{N}, \quad \underline{\lambda} \leqslant \lambda_k \leqslant 1 \quad \Longrightarrow \quad \lambda_k \in (0, 1],$$

donc en utilisant la convexité de R fournie par l'hypothèse de **régularité**, on a :

$$\begin{aligned} G(x_{k+1}) &= F(x_{k+1}) + R(x_{k+1}) \\ &= F(x_{k+1}) + R((1 - \lambda_k)x_k + \lambda_k y_k) \\ &\leqslant F(x_{k+1}) + (1 - \lambda_k)R(x_k) + \lambda_k R(y_k). \end{aligned}$$

donc on obtient :

$$G(x_{k+1}) \leqslant F(x_{k+1}) + (1 - \lambda_k)R(x_k) + \lambda_k R(y_k) \tag{2.1.2}$$

Selon l'hypothèse de **majoration** :

$$\forall x \in \mathrm{domR}, \quad F(x) \leqslant F(x_k) + \langle \nabla F(x_k), x - x_k\rangle + \tfrac{1}{2}\|x - x_k\|^2_{A_k},$$

en remplaçant x par x_{k+1} on obtient :

$$F(x_{k+1}) \leqslant F(x_k) + \langle \nabla F(x_k), x_{k+1} - x_k\rangle + \tfrac{1}{2}\|x_{k+1} - x_k\|^2_{A_k}$$

ainsi l'inégalité (2.1.2) devient :

$$G(x_{k+1}) \leqslant F(x_k) + \langle \nabla F(x_k), x_{k+1} - x_k \rangle + \frac{1}{2}\|x_{k+1} - x_k\|^2_{A_k} + (1-\lambda_k)R(x_k) + \lambda_k R(y_k). \tag{2.1.3}$$

On peut vérifier facilement que y_k solution du problème (1.8.1) implique l'inégalité suivante :

$$R(y_k) + \langle \nabla F(x_k), y_k - x_k \rangle + \gamma_k^{-1}\|y_k - x_k\|^2_{A_k} \leqslant R(x_k), \tag{2.1.4}$$

Selon l'égalité (2.1.1) on a :

$$x_{k+1} - x_k = \lambda_k(y_k - x_k) \quad \Longrightarrow \quad y_k - x_k = \lambda_k^{-1}(x_{k+1} - x_k),$$

à partir de l'inégalité (2.1.4) on obtient :

$$R(y_k) + \lambda_k^{-1}\langle \nabla F(x_k), x_{k+1} - x_k \rangle + \gamma_k^{-1}\lambda_k^{-2}\|x_{k+1} - x_k\|^2_{A_k} \leqslant R(x_k),$$

donc on trouve :

$$\langle \nabla F(x_k), x_{k+1} - x_k \rangle \leqslant -\gamma_k^{-1}\lambda_k^{-1}\|x_{k+1} - x_k\|^2_{A_k} + \lambda_k(R(x_k) - R(y_k)). \tag{2.1.5}$$

En combinant (2.1.3) et (2.1.5), on trouve :

$$\begin{aligned} G(x_{k+1}) &\leqslant F(x_k) + R(x_k) + \frac{1}{2}\|x_{k+1} - x_k\|^2_{A_k} - \gamma_k^{-1}\lambda_k^{-1}\|x_{k+1} - x_k\|^2_{A_k} \\ &\leqslant G(x_k) - (\gamma_k^{-1}\lambda_k^{-1} - \frac{1}{2})\|x_{k+1} - x_k\|^2_{A_k}. \end{aligned}$$

A partir de l'hypothèse sur les **paramètres**, on trouve facilement que :

$$-\gamma_k^{-1}\lambda_k^{-1} + \tfrac{1}{2} \leqslant -\tfrac{1}{2}\tfrac{\bar{\eta}}{2-\bar{\eta}},$$

alors on obtient :

$$\begin{aligned} G(x_{k+1}) &\leqslant G(x_k) - \frac{1}{2}\frac{\bar{\eta}}{2-\bar{\eta}}\|x_{k+1} - x_k\|^2_{A_k} \\ &= G(x_k) - \frac{1}{2}\frac{\bar{\eta}}{2-\bar{\eta}}(x_{k+1} - x_k)^T A_k (x_{k+1} - x_k) \\ &= G(x_k) + \frac{1}{2}\frac{\bar{\eta}}{2-\bar{\eta}}(x_{k+1} - x_k)^T (-A_k)(x_{k+1} - x_k). \end{aligned}$$

donc on a :

$$G(x_{k+1}) \leqslant G(x_k) + \frac{1}{2}\frac{\bar{\eta}}{2-\bar{\eta}}(x_{k+1} - x_k)^T (-A_k)(x_{k+1} - x_k)$$

Selon de l'hypothèse de **majoration** :

$$\exists \quad (\underline{\nu}, \bar{\nu}) \in (0, +\infty)^2, \quad \forall k \in \mathbb{N}, \quad \underline{\nu} I_n \preceq A_k \preceq \bar{\nu} I_n \quad \Longrightarrow \quad -A_k \preceq -\underline{\nu} I_n$$

donc on obtient :

$$G(x_{k+1}) \leqslant G(x_k) - \frac{1}{2}\frac{\underline{\nu}\bar{\eta}}{2-\bar{\eta}}(x_{k+1} - x_k)^T I_n (x_{k+1} - x_k)$$

On pose $\boxed{\mu_1 = \dfrac{\underline{\nu}\bar{\eta}}{2-\bar{\eta}}}$, d'où le résultat :

$$\boxed{G(x_{k+1}) \leqslant G(x_k) - \frac{\mu_1}{2}\|x_{k+1} - x_k\|^2}$$

Puisque $x_{k+1} - x_k = \lambda_k(y_k - x_k)$, alors

$$G(x_{k+1}) \leqslant G(x_k) - \frac{\mu_1}{2}\lambda_k^2\|y_k - x_k\|^2,$$

selon l'hypothèse sur les **paramètres** il existe $\underline{\lambda}$ dans $(0, +\infty)$ tel que λ_k est dans $[\underline{\lambda}, 1]$ donc on trouve que $(-\lambda_k^2)$ est inférieur à $(-\underline{\lambda}^2)$, ainsi on obtient :

$$\boxed{G(x_{k+1}) \leqslant G(x_k) - \underline{\lambda}^2 \frac{\mu_1}{2} \|y_k - x_k\|^2}$$

Remarque 2.1.2. *Le lemme ci-dessus, montre que la suite* $(G(x_k))_{k \in N}$ *est décroissante.*

Par conséquent de ce lemme, on peut exprimer l'hypothèse de la **faible convexité** de G sous une autre forme :

Corollaire 2.1.3. *On suppose que les hypothèses de **régularité**, de **majoration** et sur les **paramètres** sont vraies. L'hypothèse de la **faible convexité** de G est satisfaite si et seulement si pour tout $k \in \mathbb{N}$, il existe $\alpha_k \in [\underline{\alpha}, 1]$ tel que :*

$$G(x_{k+1}) \leqslant (1 - \alpha_k) G(x_k) + \alpha_k G(y_k) \tag{2.1.6}$$

Preuve :
On peut voir que l'inégalité (2.1.6) est équivalente à l'inégalité suivante :

$$\alpha_k (G(x_k) - G(y_k)) \leqslant G(x_k) - G(x_{k+1}). \tag{2.1.7}$$

Soit $\underline{\alpha} \in (0, 1]$ et $k \in \mathbb{N}$:
Supposons qu'il existe $\alpha_k \in [\underline{\alpha}, 1]$ tel que l'inégalité (2.1.7) est vraie.
Selon l'inégalité (2.1.7), deux cas peuvent se présenter :

1) Cas où $G(x_k) \leqslant G(y_k)$ qui implique que $\underline{\alpha}(G(x_k) - G(y_k)) \leqslant 0$:
   selon le lemme (2.1.1) de la variation de la suite $G(x_k)_{k \in \mathbb{N}}$, cette dernière est une suite décroissante donc $0 \leqslant G(x_k) - G(x_{k+1})$, alors on obtient :

$$\underline{\alpha}(G(x_k) - G(y_k)) \leqslant G(x_k) - G(x_{k+1})$$

   d'où l'hypothèse de la **faible convexité** sur G :

$$G(x_{k+1}) \leqslant (1 - \underline{\alpha}) G(x_k) + \underline{\alpha} G(y_k).$$

2) Cas où $G(x_k) \geqslant G(y_k)$ qui implique $0 \leqslant G(x_k) - G(y_k)$:
   Puisque $0 \leqslant \underline{\alpha} \leqslant \alpha_k$ pour tout $k \in \mathbb{N}$ alors à partir de l'inégalité (2.1.7), on obtient :

$$\underline{\alpha}(G(x_k) - G(y_k)) \leqslant \alpha_k (G(x_k) - G(y_k)) \leqslant G(x_k) - G(x_{k+1}).$$

   ce qui donne alors :

$$\underline{\alpha}(G(x_k) - G(y_k)) \leqslant G(x_k) - G(x_{k+1})$$

   d'où l'hypothèse de la **faible convexité** est vraie,

$$G(x_{k+1}) \leqslant (1 - \underline{\alpha}) G(x_k) + \underline{\alpha} G(y_k).$$

Lemme 2.1.4. *Selon les hypothèses de **régularité**, de **majoration** et des **paramètres** il existe une constante $\mu_2 \in \mathbb{R}$ telle que,*

$$(\forall k \in \mathbb{N}), G(y_k) \leqslant G(x_k) - \mu_2 \|y_k - x_k\|^2$$

Preuve :
A partir de l'hypothèse de **majoration** on remplace x par $y_k \in \mathbb{R}^n$, on obtient :

$$F(y_k) \leqslant F(x_k) + \langle \nabla F(x_k), y_k - x_k \rangle + \frac{1}{2} \|y_k - x_k\|^2_{A_k} \tag{2.1.8}$$

et selon l'inégalité (2.1.4) :

$$R(y_k) + \langle \nabla F(x_k), y_k - x_k \rangle + \gamma_k^{-1} \|y_k - x_k\|^2_{A_k} \leqslant R(x_k),$$

on trouve :

$$\langle \nabla F(x_k), y_k - x_k \rangle \leqslant R(x_k) - R(y_k) - \gamma_k^{-1} \|y_k - x_k\|^2_{A_k}$$

donc l'inégalité (2.1.8) devient :

$$F(y_k) \leqslant F(x_k) + R(x_k) - R(y_k) - \gamma_k^{-1}\|y_k - x_k\|_{A_k}^2 + \tfrac{1}{2}\|y_k - x_k\|_{A_k}^2,$$

on trouve :

$$F(y_k) + R(y_k) \leqslant F(x_k) + R(x_k) - (\gamma_k^{-1} - \tfrac{1}{2})\|y_k - x_k\|_{A_k}^2$$

ainsi on obtient :

$$G(y_k) \leqslant G(x_k) - (\gamma_k^{-1} - \frac{1}{2})(y_k - x_k)^T A_k (y_k - x_k). \tag{2.1.9}$$

Selon l'hypothèse sur les **paramètres** :

$$\exists \ (\underline{\eta}, \bar{\eta}, \underline{\lambda}) \in (0, +\infty)^3, \quad \begin{cases} \underline{\eta} \leqslant \gamma_k \lambda_k \leqslant 2 - \bar{\eta} \\ \underline{\lambda} \leqslant \gamma_k \leqslant 1 \end{cases} \Longrightarrow \begin{cases} \frac{1}{2-\bar{\eta}} \leqslant \frac{1}{\gamma_k \lambda_k} \leqslant \frac{1}{\underline{\eta}} \\ \underline{\lambda} \leqslant \lambda_k \leqslant 1 \end{cases} \Longrightarrow -\gamma_k^{-1} + \frac{1}{2} \leqslant \frac{\underline{\lambda}}{2 - \bar{\eta}} - \frac{1}{2}$$

donc (2.1.9) devient :

$$\begin{aligned} G(y_k) &\leqslant G(x_k) - (\frac{\underline{\lambda}}{2 - \bar{\eta}} - \frac{1}{2})(y_k - x_k)^T A_k (y_k - x_k) \\ &\leqslant G(x_k) + (\frac{\underline{\lambda}}{2 - \bar{\eta}} - \frac{1}{2})(y_k - x_k)^T (-A_k)(y_k - x_k). \end{aligned}$$

D'une part, selon l'hypothèse de **majoration** :

$$\exists \quad \underline{\nu} \in (0, +\infty), \quad -A_k \preceq -\underline{\nu} I_n$$

on obtient :

$$\begin{aligned} G(y_k) &\leqslant G(x_k) - \underline{\nu}(\frac{\underline{\lambda}}{2 - \bar{\eta}} - \frac{1}{2})(y_k - x_k)^T I_n (y_k - x_k) \\ &\leqslant G(x_k) - \underline{\nu}(\frac{\underline{\lambda}}{2 - \bar{\eta}} - \frac{1}{2})\langle I_n(y_k - x_k), y_k - x_k \rangle \\ &\leqslant G(x_k) - \underline{\nu}(\frac{\underline{\lambda}}{2 - \bar{\eta}} - \frac{1}{2})\|y_k - x_k\|^2 \end{aligned}$$

selon l'hypothèse sur les **paramètres**, $2 - \bar{\eta} > 0$, alors :

$$\tfrac{2\underline{\lambda} - 2 + \bar{\eta}}{2(2 - \bar{\eta})} = \tfrac{\underline{\lambda}}{2 - \bar{\eta}} - \tfrac{1}{2} \geqslant 0 \quad \text{si} \quad 2\underline{\lambda} + \bar{\eta} \geqslant 2$$

On pose alors :

$$\mu_2 = \underline{\nu}(\tfrac{\underline{\lambda}}{2 - \bar{\eta}} - \tfrac{1}{2}) \geqslant 0 \quad \text{si} \quad 2\underline{\lambda} + \bar{\eta} \geqslant 2$$

D'autre part, selon l'hypothèse **majoration** $A_k \preceq \bar{\nu} I_n$, on obtient :

$$\begin{aligned} G(y_k) &\leqslant G(x_k) + (-(\frac{\underline{\lambda}}{2 - \bar{\eta}} - \frac{1}{2}))(y_k - x_k)^T \bar{\nu} I_n (y_k - x_k) \\ &= G(x_k) - \bar{\nu}(\frac{\underline{\lambda}}{2 - \bar{\eta}} - \frac{1}{2})(y_k - x_k)^T I_n (y_k - x_k) \\ &= G(x_k) - \bar{\nu}(\frac{\underline{\lambda}}{2 - \bar{\eta}} - \frac{1}{2})\langle I_n(y_k - x_k), y_k - x_k \rangle \\ &= G(x_k) - \bar{\nu}(\frac{\underline{\lambda}}{2 - \bar{\eta}} - \frac{1}{2})\|y_k - x_k\|^2. \end{aligned}$$

On a :

$$\frac{\underline{\lambda}}{2 - \bar{\eta}} - \frac{1}{2} < 0 \quad \text{si} \quad 2\lambda + \bar{\eta} < 2 \quad \text{donc} \quad \bar{\nu}(\frac{\underline{\lambda}}{2 - \bar{\eta}} - \frac{1}{2}) < 0$$

On pose alors :

$$\mu_2 = \bar{\nu}(\frac{\lambda}{2-\bar{\eta}} - \frac{1}{2}) < 0 \quad \text{si} \quad 2\lambda + \bar{\eta} < 2$$

donc on obtient :

$$\mu_2 = \begin{cases} \underline{\nu}(\frac{\lambda}{2-\bar{\eta}} - \frac{1}{2}) & \text{si} \quad 2\underline{\lambda} + \bar{\eta} \geqslant 2 \\ \bar{\nu}(\frac{\lambda}{2-\bar{\eta}} - \frac{1}{2}) & \text{sinon} \end{cases}$$

donc il existe $\mu_2 \in \mathbb{R}$ tel que,

$$\boxed{G(y_k) \leqslant G(x_k) - \mu_2 \|y_k - x_k\|^2}$$

Remarque 2.1.5. *Le lemme ci-dessus montre la variation de G lors du passage au point x_k et y_k à chaque itération de l'algorithme :*

(i) Dans le cas où $\mu_2 \in \mathbb{R}_+$, alors $G(y_k) \leqslant G(x_k)$ et donc G augmente de plus en plus en x_k par rapport à y_k à chaque itération k.

(ii) Dans le cas où $\mu_2 \in \mathbb{R}_-$, on n'est peut rien conclure.

2.2 Le Résultat de convergence

Nous sommes sur le point de conclure quant à la convergence de l'algorithme mais avant ça, le lemme suivant jouera un rôle central pour prouver la convergence de l'algorithme :

Lemme 2.2.1. ***suite Sommable :***
Soit $(u_k)_{k\in\mathbb{N}}$, $(g_k)_{k\in\mathbb{N}}$, $(g_k^{'})_{k\in\mathbb{N}}$ et $(\Delta_k)_{k\in\mathbb{N}}$ des suites réelles positives et soit $\theta \in (0,1)$. Supposons que :

(i) Pour tout $k \in \mathbb{N}$, $u_k^2 \leqslant g_k^\theta \Delta_k$.

(ii) $(\Delta_k)_{k\in\mathbb{N}}$ est sommable.

(iii) Pour tout $k \in \mathbb{N}$, $g_{k+1} \leqslant (1-\underline{\alpha})g_k + g_k^{'}$ où $\underline{\alpha} \in (0,1]$.

(iv) Pour tout $k \geqslant k^*$, $(g_k^{'})^\theta \leqslant \beta u_k$ ou $\beta > 0$ et $k^* \in \mathbb{N}$.

Alors $(u_k)_{k\in\mathbb{N}}$ est une suite sommable.

Preuve :
A partir de (iii), par la formule de binome Newton, on a :

$$\begin{aligned} g_{k+1}^\theta &= ((1-\underline{\alpha})g_k + g_k^{'})^\theta \\ &= (1-\underline{\alpha})^\theta g_k^\theta + (g_k^{'})^\theta + (1-\underline{\alpha})g_k^{\theta-1}g_k^{'} \\ &= (1-\underline{\alpha})^\theta g_k^\theta + (g_k^{'})^\theta + \underbrace{g_k^{\theta-1}g_k^{'} - \alpha g_k^{\theta-1}g_k^{'}}_{\leqslant 0} \end{aligned}$$

donc on obtient :

$$\forall k \in \mathbb{N}, \quad g_{k+1}^\theta = (1-\underline{\alpha})^\theta g_k^\theta + (g_k^{'})^\theta$$

à partir de (iv), on obtient : $\quad g_{k+1}^\theta \leqslant (1-\underline{\alpha})^\theta g_k^\theta + \beta u_k$.
Si on somme en k selon k^*, pour tout $K > k^*$, on obtient :

$$\sum_{k=k^*+1}^{K} g_{k+1}^\theta \leqslant (1-\underline{\alpha})^\theta \sum_{k=k^*+1}^{K} g_k^\theta + \beta \sum_{k=k^*+1}^{K} u_k,$$

on trouve :

$$\sum_{k=k^*}^{K-1} g_{k+1}^\theta - (1-\underline{\alpha})^\theta \sum_{k=k^*}^{K-1} g_k^\theta \leqslant \beta \sum_{k=k^*}^{K-1} u_k \tag{2.2.1}$$

On remarque que :

$$\begin{aligned} \sum_{k=k^*}^{K-1} g_{k+1}^\theta - (1-\underline{\alpha})^\theta \sum_{k=k^*}^{K-1} g_k^\theta &= -g_{k^*}^\theta + g_K^\theta + \sum_{k=k^*}^{K-1} g_k^\theta - (1-\underline{\alpha})^\theta \sum_{k=k^*}^{K-1} g_k^\theta \\ &= -g_{k^*}^\theta + g_K^\theta + (1-(1-\underline{\alpha})^\theta) \sum_{k=k^*}^{K-1} g_k^\theta. \end{aligned}$$

On obtient via l'inégalité (2.2.1) que :

$$(1-(1-\underline{\alpha})^\theta)\sum_{k=k^*}^{K-1} g_k^\theta \leqslant g_{k^*}^\theta - g_K^\theta + \beta\sum_{k=k^*}^{K-1} u_k. \tag{2.2.2}$$

D'autre part, l'hypothèse (i) est équivalente à :

$$u_k^2 \leqslant (\beta^{-1}(1-(1-\underline{\alpha})^\theta)g_k^\theta)(\beta(1-(1-\underline{\alpha})^\theta)^{-1}\Delta_k),$$

on obtient :

$$u_k \leqslant \sqrt{(\beta^{-1}(1-(1-\underline{\alpha})^\theta)g_k^\theta)(\beta(1-(1-\underline{\alpha})^\theta)^{-1}\Delta_k)}.$$

On a l'identité remarquable suivante :

$$(v,v')\in[0,+\infty)^2, \quad \sqrt{vv'} \leqslant \tfrac{1}{2}(v+v').$$

Alors on pose :

$$\begin{cases} v = (\beta^{-1}(1-(1-\underline{\alpha})^\theta)g_k^\theta) \geqslant 0 \\ v' = (\beta(1-(1-\underline{\alpha})^\theta)^{-1}\Delta_k) \geqslant 0 \end{cases}$$

donc on obtient :

$$u_k \leqslant \tfrac{1}{2}[(\beta^{-1}(1-(1-\underline{\alpha})^\theta)g_k^\theta) + (\beta(1-(1-\underline{\alpha})^\theta)^{-1}\Delta_k)]$$

On somme de k^* à $K-1$ et on a :

$$\sum_{k=k^*}^{K-1} u_k \leqslant \frac{1}{2}(\beta^{-1}(1-(1-\underline{\alpha})^\theta)\sum_{k=k^*}^{K-1} g_k^\theta + \frac{1}{2}(\beta(1-(1-\underline{\alpha})^\theta)^{-1}\sum_{k=k^*}^{K-1}\Delta_k.$$

A partir de l'inégalité (2.2.2), on obtient :

$$\sum_{k=k^*}^{K-1} u_k \leqslant \frac{1}{2}\beta^{-1}g_{k^*}^\theta - \frac{1}{2}\beta^{-1}g_k^\theta + \frac{1}{2}\beta^{-1}\beta\sum_{k=k^*}^{K-1} u_k + \frac{1}{2}\beta(1-(1-\underline{\alpha})^\theta)^{-1}\sum_{k=k^*}^{K-1}\Delta_k,$$

en simplifiant, on obtient :

$$\sum_{k=k^*}^{K-1} u_k \leqslant \beta^{-1}g_{k^*}^\theta + \beta(1-(1-\underline{\alpha})^\theta)^{-1}\sum_{k=k^*}^{K-1}\Delta_k.$$

Du fait que $(\Delta_k)_{k\in\mathbb{N}}$ est sommable, alors $(u_k)_{k\in\mathbb{N}}$ l'est aussi.

Théoreme 2.2.2. *Selon les hypothèses de* ***régularité****, l'****inégalité KL****, de* ***majoration****, des* ***paramètre*** *et de la* ***faible convexité****, les assertions suivantes sont vraies :*

(i) les suites $(x_k)_{k\in\mathbb{N}}$ et $(y_k)_{k\in\mathbb{N}}$ définies dans l'algorithme inexact convergent toutes les deux vers un point crique $\hat{x}$ de G.

(ii) Ces suites ont une longueur finie dans le sens où

$$\sum_{k=0}^{+\infty}\|x_{k+1}-x_k\| < +\infty \quad \textit{et} \quad \sum_{k=0}^{+\infty}\|y_{k+1}-y_k\| < +\infty$$

(iii) $G(x_k)_{k\in\mathbb{N}}$ et $G(y_k)_{k\in\mathbb{N}}$ convergent vers $G(\hat{x})$. De plus, $G(x_k)_{k\in\mathbb{N}}$ est une suite décroissante.

Preuve :

(1) On cherche d'abord à montrer que les suites $(x_k)_{k\in\mathbb{N}}$ et $(y_k)_{k\in\mathbb{N}}$ convergent :
D'après le lemme sur la variation de la suite $(G(x_k))_{k\in\mathbb{N}}$, $(G(x_k))_{k\in\mathbb{N}}$ est une suite décroissante. D'après la remarque sur l'hypothèse de **régularité**, dom$G$ est non vide, G semi-continue supérieur, continue dans son domaine et pour tout $x \in \mathrm{dom}G = \mathrm{dom}R$, le sous-niveau $\{G \leqslant G(x)\}$ est compact et selon la remarque (1.8.1), pour tout k dans $\mathbb{N}$, x_k et y_k appartiennent toutes les deux dans dom$R$. Donc, $(x_k)_{k\in\mathbb{N}}$ appartient à un sous-ensemble compact E de sous-niveau de G inclu dans dom$R$. En conséquence, la suite $(G(x_k))_{k\in\mathbb{N}}$ converge vers un réel ϵ, donc on obtient :

$$\boxed{(G(x_k) - \epsilon) \xrightarrow[k\to+\infty]{} 0}$$

De plus, en évoquant à nouveau le lemme de la variation de $G(x_k)_{k\in\mathbb{N}}$, pour tout k dans $\mathbb{N}$, on trouve :

$$\begin{aligned}\underline{\lambda}^2\frac{\mu_1}{2}\|y_k - x_k\|^2 &\leqslant G(x_k) - G(x_{x+1})\\ &\leqslant G(x_k) - G(x_{x+1}) - \epsilon + \epsilon\end{aligned}$$

alors on trouve :

$$\underline{\lambda}^2\frac{\mu_1}{2}\|y_k - x_k\|^2 \leqslant (G(x_k) - \epsilon) - (G(x+1) - \epsilon) \tag{2.2.3}$$

donc on obtient :

$$\|y_k - x_k\|^2 \leqslant \frac{2}{\underline{\lambda}^2\mu_1}[\underbrace{(G(x_k) - \epsilon)}_{\to 0} - \underbrace{(G(x_{k+1}) - \epsilon)}_{\to 0}]$$

alors :

$$\boxed{(y_k - x_k) \xrightarrow[k\to+\infty]{} 0}.$$

D'une part, selon l'hypothèse de la **faible convexité** :

$$\begin{aligned}G(x_{k+1}) &\leqslant (1-\underline{\alpha})G(x_k) + \underline{\alpha}G(y_k)\\ G(x_{k+1}) - \epsilon &\leqslant (1-\underline{\alpha})G(x_k) + \underline{\alpha}G(y_k) - \epsilon\\ &\leqslant G(x_k) - \underline{\alpha}G(x_k) + \underline{\alpha}G(y_k) - \epsilon\\ &\leqslant G(x_k) - \epsilon - \underline{\alpha}G(x_k) + \underline{\alpha}G(y_k) - \underline{\alpha}\epsilon + \underline{\alpha}\epsilon\\ &\leqslant G(x_k) - \epsilon - \underline{\alpha}(G(x_k) - \epsilon) + \underline{\alpha}(G(y_k) - \epsilon)\end{aligned}$$

donc obtient :

$$G(x_{k+1}) - \epsilon \leqslant (1-\underline{\alpha})(G(x_k) - \epsilon) + \underline{\alpha}(G(y_k) - \epsilon) \tag{2.2.4}$$

ainsi on trouve :

$$\frac{1}{\underline{\alpha}}[G(x_{k+1}) - \epsilon) - (1-\underline{\alpha})(G(x_k) - \epsilon)] \leqslant G(y_k) - \epsilon. \tag{2.2.5}$$

Selon le lemme (2.1.4) sur la Variation de G en y_k et x_k, pour tout $k \in \mathbb{N}$:

$$\exists \mu_2 \in \mathbb{R}, \quad G(y_k) \leqslant G(x_k) - \mu_2\|y_k - x_k\|^2$$

donc on trouve :

$$G(y_k) - \epsilon \leqslant G(x_k) - \epsilon - \mu_2\|y_k - x_k\|^2. \tag{2.2.6}$$

Par combinaison de (2.2.5) et (2.2.6) on obtient :

$$\frac{1}{\underline{\alpha}}[\underbrace{G(x_{k+1}) - \epsilon}_{\to 0} - (1-\underline{\alpha})\underbrace{(G(x_k) - \epsilon)}_{\to 0}] \leqslant G(y_k) - \epsilon \leqslant \underbrace{(G(x_k) - \epsilon)}_{\to 0} - \mu_2\underbrace{\|y_k - x_k\|^2}_{\to 0},$$

donc on obtient :

$$\boxed{G(y_k) - \epsilon \xrightarrow[k\to+\infty]{} 0}.$$

Soit la fonction g définie par :,

$$\begin{aligned} g : [0, +\infty) &\longrightarrow [0, +\infty) \\ u &\longmapsto g(u) = u^{\frac{1}{1-\theta}} \qquad \text{où} \quad \theta \in [0, 1) \end{aligned}$$

On a :

$$\begin{aligned} g' : [0, +\infty) &\longrightarrow [0, +\infty) \\ u &\longmapsto g'(u) = \frac{(2\theta - 1)u^{\frac{2\theta-1}{1-\theta}}}{1-\theta} \end{aligned}$$

pour $\theta \in [0, 1)$, $g'(u) \geqslant 0$, donc la fonction g est convexe.
La fonction g est dérivable et convexe alors :

$$\begin{aligned} \forall (u, v) \in [0, +\infty)^2, \quad g(u) &\leqslant g'(u)(u - v) + g(v) \\ g(u) - g(v) &\leqslant g'(u)(u - v) \end{aligned} \tag{2.2.7}$$

Par changement de variables, on pose :

$$\begin{cases} M = g(u) \\ N = g(v) \end{cases} \Longrightarrow \begin{cases} u = g^{-1}(M) \\ v = g^{-1}(N) \end{cases},$$

l'inégalité (2.2.7) devient :

$$M - N \leqslant g'(g^{-1}(M))(g^{-1}(M) - g^{-1}(N)).$$

On remarque que :

$$g(x) = y \quad \Longrightarrow \quad x = y^{1-\theta}, \quad \text{alors} \quad g^{-1}(M) = M^{1-\theta} \quad \text{et} \quad g'(M^{1-\theta}) = \tfrac{1}{1-\theta} M^{\theta},$$

donc on obtient :

$$M - N \leqslant \frac{1}{1-\theta} M^{\theta}(M^{1-\theta} - N^{1-\theta}).$$

Autrement dit, on obtient :

$$(\forall (u, v) \in [0, +\infty)^2), \quad u - v \leqslant (1 - \theta)^{-1} u^{\theta}(u^{1-\theta} - v^{1-\theta})$$

Soit $u = G(x_k) - \epsilon$ et $v = G(x_{k+1}) - \epsilon$, on a :

$$(G(x_k) - \epsilon) - (G(x_{k+1}) - \epsilon) \leqslant (1 - \theta)^{-1}(G(x_k) - \epsilon)^{\theta} \Delta_k^{'} \tag{2.2.8}$$

où pour tout $k \in \mathbb{N}$, $\Delta_k^{'} = (G(x_k) - \epsilon)^{1-\theta} - (G(x_{k+1}) - \epsilon)^{1-\theta}$.
En combinant l'inégalité (2.2.3) et (2.2.8), on obtient :

$$\underline{\lambda}^2 \frac{\mu_1}{2} \|y_k - x_k\|^2 \leqslant (1 - \theta)^{-1}(G(x_k) - \epsilon)^{\theta} \Delta_k^{'},$$

on obtient :

$$\|y_k - x_k\|^2 \leqslant 2\underline{\lambda}^{-2} \mu^{-1}(1 - \theta)^{-1}(G(x_k) - \epsilon)^{\theta} \Delta_k^{'} \tag{2.2.9}$$

En outre, comme E est un borné, alors selon l'hypothèse de l'**inégalité KL** : $\forall \xi \in \mathbb{R}$, $\exists \quad \kappa > 0$, $\zeta > 0$ et $\theta \in [0, 1)$ telles que $\forall x \in E$, $|G(x) - \xi| \leq \zeta$, on a :

$$(\forall t(x) \in \partial G(x)), \quad \kappa |G(x) - \xi|^{\theta} \leqslant \|t(x)\|.$$

$G = F + R$, selon l'hypothèse de **régularité**, F est différentiable et R est non nécessairement différentiable, donc on a :

$$\forall x \in E \ \text{ tel que } \ |G(x) - \xi| \leq \zeta, \ \ \forall r(x) \in \partial R(x) \ \ \kappa |G(x) - \xi|^{\theta} \leqslant \|\nabla F(x) + r(x)\|.$$

De ce qui précède on a trouvé qu'il existe un $\epsilon \in \mathbb{R}$ tel que :

$$(G(y_k) - \epsilon) \xrightarrow[k\to+\infty]{} 0.$$

On pose alors $\boxed{\xi = \epsilon}$, donc on obtient :

$$G(y_k) \xrightarrow[k\to+\infty]{} \xi,$$

On rappelle que la suite $(y_k)_{k\in\mathbb{N}}$ appartient dans le sous-ensemble E. Puisque la suite $(G(y_k))_{\in\mathbb{N}}$ converge vers ξ pour tout $k \in \mathbb{N}$, alors $\exists\ k^* \in \mathbb{N}$, $\exists\ \zeta > 0$ tel que $\forall k \geqslant k^*$ $|G(y_k) - \xi| \leq \zeta$, donc l'**inégalité KL** est vraie en y_k, c'est-à-dire :

$$\forall \quad r(y_k) \in \partial R(x), \quad \kappa|G(y_k) - \xi|^\theta \leqslant \|\nabla F(y_k) + r(y_k)\|.$$

Soit $r(y_k)$ définie dans l'algorithme **inexact FB**, alors nous avons,

$$\begin{aligned}
\kappa|G(y_k) - \xi|^\theta &\leqslant \|\nabla F(y_k) - \nabla F(x_k) + \nabla F(x_k) + r(y_k)\| \\
&\leqslant \|\nabla F(y_k) - \nabla F(x_k)\| + \|\nabla F(x_k) + r(y_k)\| \\
&\leqslant \|\nabla F(y_k) - \nabla F(x_k)\| + \tau\|y_k - x_k\|_{A_k}
\end{aligned}$$

on trouve :

$$\kappa|G(y_k) - \xi|^\theta \leqslant \|\nabla F(y_k) - \nabla F(x_k)\| + \tau\|y_k - x_k\|_{A_k}. \tag{2.2.10}$$

Selon l'hypothèse (ii) de **régularité** et l'hypothèse (ii) de **majoration**, F est différentiable à gradient L-Lipshist et $\underline{\nu} I_n \preceq A_k \preceq \bar{\nu} I_n$, alors l'inégalité (2.2.10) devient :

$$\begin{aligned}
|G(y_k) - \xi|^\theta &\leqslant \kappa^{-1} L\|y_k - x_k\| + \tau\|y_k - x_k\|_{A_k} \\
&\leqslant \kappa^{-1}(L\|y_k - x_k\| + \tau(y_k - x_k)^T\sqrt{A_k}(y_k - x_k)) \\
&\leqslant \kappa^{-1}(L\|y_k - x_k\| + \tau\sqrt{\bar{\nu}}(y_k - x_k)^T I_n(y_k - x_k)) \\
&\leqslant \kappa^{-1}(L + \tau\sqrt{\bar{\nu}})\|y_k - x_k\|
\end{aligned}$$

on trouve :

$$|G(y_k) - \xi|^\theta \leqslant \beta\|x_k - y_k\| \quad \text{où} \quad \beta = \kappa^{-1}(L + \tau\sqrt{\bar{\nu}}) > 0. \tag{2.2.11}$$

A partir de l'inégalité (2.2.4), puisque $\underline{\alpha} \in (0, 1]$, on obtient :

$$G(y_k) - \xi \leqslant (1 - \underline{\alpha})(G(x_k) - \xi) + |G(y_k) - \xi|. \tag{2.2.12}$$

En outre, on peut remarquer que :

$$\begin{aligned}
\sum_{k=k^*}^{+\infty} \Delta_k^{'} &= \sum_{k=k^*}^{+\infty} ((G(x_k) - \xi)^{1-\theta} - (G(x_{k+1})^{1-\theta}) \\
&= (G(x_k^*) - \xi)^{1-\theta}
\end{aligned}$$

qui montre que $\Delta_k^{'}$ est sommable. On pose :

$$\begin{cases}
u_k = \|y_k - x_k\| \\
g_k = G(x_k) - \xi \geqslant 0 \\
g_k^{'} = |G(x_k) - \xi| \geqslant 0 \\
\beta = k^{-1}(L + \tau\sqrt{\bar{\nu}}) \\
\Delta_k = 2\underline{\lambda}^{-2}\mu^{-1}(1 - \theta)^{-1}\Delta_k^{'}
\end{cases}$$

Alors l'inégalité (2.2.9) assure l'insertion (i) du lemme d'une **suite sommable**,
$\Delta_k^{'}$ est sommable à fortiori Δ_k est sommable qui assure l'insertion (ii) du lemme d'une **suite sommable**, l'inégalité (2.2.12) assure l'insertion (iii) du lemme d'une **suite sommable**,

l'inégalité (2.2.11) assure l'insertion (iv) du lemme d'une **suite sommable**.
Alors quand $\boxed{\theta \neq 0}$ c'est-à-dire quant $\theta \in (0,1)$, d'après le lemme d'une **suite sommable**, la suite $(\|y_k - x_k\|)_{k\in\mathbb{N}}$ est sommable.
Mais ce θ est défini dans $[0,1)$ alors il peut prendre la valeur 0, donc qu'est-ce qui se passe pour $\theta = 0$. On a vu ci-dessus que pour $\theta \in [0,1)$ la suite $(x_k - y_k)_{k\in\mathbb{N}}$ converge vers 0, alors il existe $k^{**} \geqslant k^*$ tel que :

$$(\forall k \geqslant k^{**}), \quad \kappa^{-1}(L + \tau\sqrt{v})\|x_k - y_k\| < 1,$$

d'après l'inégalité (2.2.11), on obtient :

$$(\forall k \geqslant k^{**}), \quad |(G(y_k) - \xi)|^{\theta} < 1.$$

En particulier pour $\theta = 0$, deux cas peuvent se présenter :
• si $G(y_k) - \xi \neq 0$, alors on obtient $1 < 1$ impossible,
• si $G(y_k) - \xi = 0$, alors on obtient (par convention 0^0=0) $0 < 1$,
alors forcement $G(y_k) = \xi$ pour tout $k \geqslant k^{**}$.
Revenons sur l'inégalité (2.2.9), on a :

$$\|y_k - x_k\|^2 \leqslant 2\underline{\lambda}^{-2}\mu^{-1}(1-\theta)^{-1}(G(x_k) - \xi)^{\theta}\Delta_k^{'}$$

puisque la suite $(x_k - y_k)_{k\in\mathbb{N}}$ converge vers 0, alors cette inégalité n'est vraie que si le terme majorant est nul et ce cas ne se présente que quand la suite $(G(x_k))_{k\in\mathbb{N}}$ tend vers ξ.
Puisque $G(y_k) = \xi$ $(\forall k \in \mathbb{N})$, alors pour que $(G(x_k))_{k\in\mathbb{N}}$ tend vers ξ, on a forcement $x_k = y_k$ $(\forall k \in \mathbb{N})$.
En conséquence, quant $\theta = 0$, on obtient $(\forall k \in \mathbb{N})$ $x_k = y_k$, donc la suite $(\|y_k - x_k\|)_{k\in\mathbb{N}}$ est identiquement nulle, donc sommable.

De plus, selon l'égalité (2.1.1) :

$$(\forall k \in \mathbb{N}), \quad x_{k+1} = (1-\lambda_k)x_k + \lambda_k y_k \quad \Longrightarrow \quad \|x_{k+1} - x_k\| = \lambda_k\|y_k - x_k\|$$

d'après l'hypothèse sur les **paramètres**, $(\forall k \in \mathbb{N}), \quad 0 < \underline{\lambda} \leqslant \lambda_k \leqslant 1$, donc on obtient :

$$(\forall k \in \mathbb{N}), \quad \|x_{k+1} - x_k\| \leqslant \|y_k - x_k\|,$$

puisque la suite $(\|y_k - x_k\|)_{k\in\mathbb{N}}$ est sommable, alors la suite $(\|x_{k+1} - x_k\|)_{k\in\mathbb{N}}$ l'est aussi :

$$\boxed{\sum_{k=0}^{+\infty} \|x_{k+1} - x_k\| < +\infty}$$

cet dernier résultat implique que la suite $(\|x_{k+1} - x_k\|)_{k\in\mathbb{N}}$ a une longueur finie et donc la suite $(x_k)_{k\in\mathbb{N}}$ est de cauchy, alors elle converge vers un point $\hat{x}$:

$$\boxed{x_k \xrightarrow[k\to+\infty]{} \hat{x}}.$$

Comme la suite $(x_k - y_k)_{k\in\mathbb{N}}$ converge vers 0, donc on obtient :

$$\boxed{y_k \xrightarrow[k\to+\infty]{} \hat{x}}.$$

(2) Il nous reste à montrer que $\hat{x}$ est un point critique de G, c'est à dire montrer que $0 \in \partial G(\hat{x})$:
Soit $r(y_k)$ donné par l'algorithme pour tout $k \in \mathbb{N}$, on pose :

$$(\forall k \in \mathbb{N}), \quad t(y_k) = \nabla F(y_k) + r(y_k)$$

alors $(t(y_k))_{k\in\mathbb{N}}$ est une suite de la sous-différéntielle de $G(y_k)$, c'est à dire $(y_k, t(y_k))_{k\in\mathbb{N}}$ est une suite de Graph∂G.
De plus, comme on l'a déjà montré sur la preuve de l'inégalité (2.2.10) :

$$\begin{aligned}
\kappa|G(y_k) - \xi|^{\theta} &\leqslant \|\nabla F(y_k) + r(y_k)\| \\
&\leqslant \|\nabla F(y_k) - \nabla F(x_k)\| + \|\nabla F(x_k) + r(y_k)\| \\
&\leqslant \|\nabla F(y_k) - \nabla F(x_k)\| + \tau\|y_k - x_k\|_{A_k} \\
&\leqslant L\|y_k - x_k\| + \tau\sqrt{\nu}\|y_k - x_k\|
\end{aligned}$$

ainsi on a :

$$\|\nabla F(y_k) + r(y_k)\| \leqslant L\|y_k - x_k\| + \tau\sqrt{\overline{\nu}}\|y_k - x_k\|$$

donc on obtient :

$$(\forall k \in \mathbb{N}), \quad \|t(y_k)\| \leqslant (L + \tau\sqrt{\overline{\nu}})\|x_k - y_k\|. \tag{2.2.13}$$

Puisque les suites $(x_k)_{k\in\mathbb{N}}$ et $(y_k)_{k\in\mathbb{N}}$ convergent toutes les deux vers $\hat{x}$, alors :

$$\|x_k - y_k\| \xrightarrow[k\to+\infty]{} \hat{x} - \hat{x} = 0.$$

A partir de l'inégalité (2.2.13), on en déduit que la suite $(t(y_k))_{k\in\mathbb{N}}$ converge vers 0, ainsi on obtient que $(y_k, t(y_k))_{k\in\mathbb{N}}$ converge vers $(\hat{x}, 0)$. Sachant que $(y_k, t(y_k))_{k\in\mathbb{N}}$ est une suite du Graph∂G qui converge vers $(\hat{x}, 0)$, donc par la propriété de la fermeture du Graph∂G, on obtient :

$$(\hat{x}, 0) \in \text{Graph}\partial G \quad \text{qui implique que} \quad \boxed{0 \in \partial G(\hat{x})},$$

donc $\hat{x}$ est un point critique.
De plus, selon la remarque sur l'hypothèse de **régularité**, la restriction de G dans son domaine est continue. On a vu aussi que $(\forall k \in \mathbb{N}) \quad y_k \in \text{dom}R = \text{dom}G$, comme G est continue en tout point de son domaine alors on a :

$$y_k \xrightarrow[k\to+\infty]{} \hat{x} \quad \Longrightarrow \quad \boxed{G(y_k) \xrightarrow[k\to+\infty]{} G(\hat{x})},$$

il est de même pour la suite $(x_k)_{k\in\mathbb{N}}$:

$$x_k \xrightarrow[k\to+\infty]{} \hat{x} \quad \Longrightarrow \quad \boxed{G(x_k) \xrightarrow[k\to+\infty]{} G(\hat{x})}.$$

Corollaire 2.2.3. *On suppose que les hypothèses 1 à 5 sont vraies et soient $(x_k)_{k\in\mathbb{N}}$ et $(y_k)_{k\in\mathbb{N}}$ les suites données par l'algorithme **inexact Forward-Backward de métrique variable**. Alors il existe $v > 0$ tel que si*

$$G(x_0) \leqslant \underset{x\in\mathbb{R}^n}{inf} \quad G(x) + v,$$

alors les suites $(x_k)_{k\in\mathbb{N}}$ et $(y_k)_{k\in\mathbb{N}}$ convergent toutes les deux vers une solution du problème initial

$$\underset{\mathbb{R}^n}{Argmin} \quad G$$

Preuve :
Selon la remarque (1.6.1) sur l'hypothèse de **régularité**, l'ensemble $\underset{x\in\mathbb{R}^n}{\text{Argmin}} G(x)$ est compact donc borné. Soit ξ le minimum de G, alors :

$$\boxed{\xi = \min_{x\in\mathbb{R}^n} \quad G(x) < +\infty}$$

Soit $E = \{x \in \mathbb{R}^n : G(x) \leqslant \xi + \delta\}$ où $\delta > 0$, toujours selon la remarque (1.6.1) sur l'hypothèse de **régularité**, l'ensemble $E \subset \text{dom}R$ est compact donc borné, l'inégalité de **Kurdyka-Lojasiewicz** implique qu'il existe des constantes $k > 0$, $\zeta > 0$ et $\theta \in [0, 1)$ telles que pour tout $x \in E$, $|G(x) - \xi| \leqslant \zeta$, on a :

$$\forall r(x) \in \partial R(x), \quad k|G(x) - \xi|^\theta \leqslant \|\nabla F(x) + r(x)\|.$$

C'est à dire que l'inégalité de **Kurdyka-Lojasiewicz** est vraie pour les x qui vérifient :

$$\begin{cases} G(x) - \xi \leqslant \delta \\ G(x) - \xi \leqslant \zeta, \end{cases}$$

Soit $v = \min\{\delta, \zeta\}$, alors l'inégalité de **Kurdyka-Lojasiewicz** est vraie pour les x qui vérifient :

$$G(x) \leqslant \xi + v$$

Supposons que cette inégalité est vraie pour le premier terme x_0 de la suite $(x_k)_{k\in\mathbb{N}}$:

$$\boxed{G(x_0) \leqslant \xi + v}$$

D'après le théorème (2.2.2), la suite $(G(x_k))_{k\in\mathbb{N}}$ est décroissante, alors on trouve :

$$\forall k \in \mathbb{N}, \quad G(x_k) \leqslant G(x_0) \leqslant \xi + v.$$

La suite $(x_k)_{k\in\mathbb{N}}$ converge vers $\hat{x}$ et G est sémi-continue inférieurement, alors on a :

$$G(\hat{x}) \leqslant \liminf_{k\to+\infty} G(x_k) \leqslant \xi + v$$

Alors on trouve que l'inégalité de **Kurdyka-Lojasiewicz** est vraie au point limite $\hat{x}$ de la suite $(x_k)_{k\in\mathbb{N}}$ et $(y_k)_{k\in\mathbb{N}}$. De cette façon, on a :

$$\forall t(\hat{x}) \in \partial G(\hat{x}), \quad k|G(\hat{x}) - \xi|^\theta \leqslant \|t(\hat{x})\| \tag{2.2.14}$$

Puisque $\hat{x}$ est un point critique de G alors on peut prendre :

$$0 \in \partial G(\hat{x}), \quad \text{soit} \quad t(\hat{x}) = 0.$$

Alors l'inégalité (2.2.14) devient :

$$k|G(\hat{x}) - \xi|^\theta \leqslant 0, \quad \text{implique} \quad |G(\hat{x}) - \xi|^\theta \leqslant 0$$

donc on obtient :

$$\boxed{G(\hat{x}) = \xi = \min_{x\in\mathbb{R}^n} \quad G(x)}$$

Ainsi, les suites $(x_k)_{k\in\mathbb{N}}$ et $(y_k)_{k\in\mathbb{N}}$ convergent vers $\hat{x}$ solution du problème initial min G.

Chapitre 3

Application

En raison d'une manque de temps, on n'a pas pu aborder le sujet dans un problème à plus grande échelle comme le cas de la restauration d'images Voir ([[1], p.25, sec.Application]) pour une application sur la reconstruction d'image, c'est ainsi qu'on espère aborder ce dernier dans un avenir proche. Cependant, dans cette partie nous abordons les trois algorithmes sur des simples exemples et voir ainsi leurs vitesses de convergence.

3.1 Algorithme Forward-Backward et algorithme Forward-Backward avec métrique variable

L'algorithme **Forward-Backward** et l'algorithme **Forward-Backward avec métrique variable** nécessitent que la fonction R soit différentiable, sinon la résolution de l'opérateur de proximité pour déterminer le point y_k n'aura pas de sens puisqu'on sera obligé de déterminer un élément du sous-différentiel de R en y_k sans qu'on sache le point y_k. Cependant, pour R non-différentiable c'est l'algorithme **inexact Forward-Backward avec métrique variable** qui aura plus d'intérêt.

On considère G définie comme la somme des fonctions F et G suivantes :

$$\begin{aligned} G : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ &\longmapsto G(x) = F(x) + R(x) \end{aligned}$$

où les fonction F et R sont définie par :

$$F(x) = 4\cos(x_1) + 4\sin(x_2 + 1) \quad \text{et} \quad R(x) = x_1^2 + x_2^2$$

1. La fonction R est effectivement propre, convexe et continue donc à fortiori sémi-continue inférieurement.
2. La fonction F est effectivement différentiable et à 4-Lipshist, en effet :

$$\nabla^2 F(x) = -4 \begin{pmatrix} \cos(x_1) & 0 \\ 0 & \sin(x_2+1) \end{pmatrix} \leqslant 4 \begin{pmatrix} \cos(x_1) & 0 \\ 0 & \sin(x_2+1) \end{pmatrix}$$

   donc on prend $A = 4I_2$ (pour l'algorithme Forward-Backward avec métrique variable).
3. La fonction G est coercive, en effet :

$$\lim_{\|x\| \to +\infty} G(x) = \lim_{\|x\| \to +\infty} 4\cos(x_1) + 4\sin(x_2+1) + x_1^2 + x_2^2 = +\infty$$

3.1.1 Résolution du problème par l'algorithme Forward-Backward :

L'algorithme **Forward-Backward** est le suivant :

$$\begin{cases} x_0 \in \mathbb{R} \\ \text{pour} \quad k \geqslant 0 \\ \qquad y_k = \text{prox}_{\lambda,R}(x_k - \lambda \nabla F(x_k)) \\ \qquad x_{k+1} = x_k + \lambda(y_k - x_k) \end{cases}$$

On a :

$$y_k = \text{prox}_{\lambda R}(x_k - \lambda \nabla F(x_k))$$

implique :

$$y_k \in \underset{x \in \mathbb{R}^n}{\text{Argmin}} \quad R(x) + \frac{1}{2\lambda} \|x - x_k + \lambda \nabla F(x_k)\|^2$$

Comme R est convexe alors ce dernier est équiuvalent à :

$$\nabla(R(y_k) + \frac{1}{2\lambda}\|y_k - x_k + \lambda \nabla F(x_k)\|^2) = 0$$
$$\Longrightarrow 2y_k + \frac{1}{\lambda}(y_k - x_k + \lambda \nabla F(x_k)) = 0$$
$$\Longrightarrow y_k = \frac{1}{2\lambda + 1}(x_k - \lambda \nabla F(x_k))$$

Résolution

```
x0=np.array([[1],[2]])
lamb=1/2

# Gradient de la fonction F
def gradf(x):
    return 4*np.cos(x[1]+1)-4*np.sin(x[0])

# Resultat de l'operateur prox qui fourni le point y
def prox(x):
        return (1/(1+2*lamb))*(x-lamb*gradf(x))

# valeur initial x0
x=x0

# on cherche une precision ( qui est l'ecaryt entre 2 sol soit ici lamb(y-x) )
norme=1

# Le conteur d'iteration initilisé à 0
cpt=0

while norme > 1e-6:
    y=prox(x)
    p=lamb*(y-x)
    x=x+p
    norme=np.linalg.norm(p)
    cpt+=1
print("La solution x est :\n", x)
print("Le nombre d'itération realisé pour calculer la solution est:\n", cpt)
```

```
La solution x est :
 [[2.66095546]
 [2.66095864]]
Le nombre d'itération realisé pour calculer la solution est:
 44
```

Tracée de la fonction G :

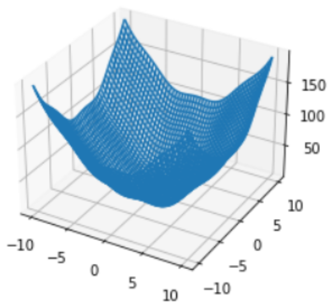
```python
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')

def f(x1,x2):
    return 4*np.cos(x1)+4*np.sin(x2+1)
def r(x1,x2):
    return x1**2+x2**2
def g(x1,x2):
    return f(x1,x2)+r(x1,x2)

x1 = np.linspace(-10,10,50)
x2 = np.linspace(-10,10,50)

# fabrication de la matrice de cordonnées x,y
x1, x2= np.meshgrid(x1, x2)
ax.plot_wireframe(x1, x2, g(x1, x2), rstride=1, cstride=1)

plt.show()
```



3.1.2 Résolution du problème par l'algorithme Forward-Backward avec métrique variable :

L'algorithme **Forward-Backward avec métrique variable** est le suivant :

$$\textbf{(MVFB)} \begin{cases} x_0 \in \mathbb{R}^n \quad \text{fixé} \\ \text{pour} \quad k \geqslant 0 : \\ \qquad y_k = \text{prox}_{\gamma_k^{-1} A_k, R}(x_k - \gamma_k A_k^{-1} \nabla F(x_k)) \\ \qquad x_{k+1} = x_k + \lambda_k (y_k - x_k) \end{cases}$$

On a :

$$y_k = \text{prox}_{\gamma_k^{-1} A_k, R}(x_k - \gamma_k A_k^{-1} \nabla F(x_k))$$

implique :

$$y_k \in \underset{x \in \mathbb{R}^n}{\text{Argmin}} \quad R(x) + \frac{1}{2} \| x - x_k + \gamma_k A_k^{-1} \nabla F(x_k) \|^2_{\gamma_k^{-1} A_k}$$

Comme R est convexe donc ce dernier est équivalent à :

$$\begin{aligned} & \nabla(R(y_k) + \frac{1}{2} \| y_k - x_k + \gamma_k A_k^{-1} \nabla F(x_k) \|^2_{\gamma_k^{-1} A_k}) = 0 \\ \text{A est symetrique} \quad & \Longrightarrow 2y_k + \frac{1}{\gamma_k} A_k (y_k - x_k + \gamma_k A_k^{-1} \nabla F(x_k)) = 0 \\ & \Longrightarrow y_k = (2\gamma_k I_2 + A_k)^{-1} (A_k x_k - \gamma_k \nabla F(x_k)) \end{aligned}$$

où la matrice A_k vérifie l'hypothèse de majoration, ici on a $A = 4I_2$. Pour l'hypothèse sur les paramètres, on peut prendre $\gamma_k = \frac{1}{3}$ et $\lambda_k = \frac{1}{2}$, donc on peut prendre :

$$\underline{\eta} = \gamma_k \lambda_k \quad \text{et} \quad \bar{\eta} = 1$$

Ainsi $\tau = (\gamma_k \lambda_k)^{-1} \sqrt{\bar{\nu}}$ où $\bar{\nu}$ la plus grande valeur propre de A qui est égale à 4.

```
x0=np.array([[1],[2]])
I=np.array([[1,0],[0,1]])
A=4*I
lamb=1/2
gamma=1/3

#Matrice du grandient de F
def matricegrad(x):
    return np.array([[-4*np.sin(x[0,0])],[4*np.cos(x[1,0]+1)]])


# La hessienne de F
def hess(x):
    return -4*np.cos(x[0,0])-4*np.sin(x[1,0]+1)


#matrice hessienne
def matricehess(x):
    return np.array([[-4*np.cos(x[0,0]),0],[0,-4*np.sin(x[1,0]+1)]])


#Matrice (2Ilamb+A)
D=2*gamma*I+A
B=np.linalg.inv(D)

# Resultat de l'operateur prox qui fourni le point y
def prox(x):
    return np.dot(B,(np.dot(A,x)-gamma*matricegrad(x)))
```

```
norme=1
# Le conteur d'iteration initilisé à 0
cpt=0
x=x0

#Resolution du probleme
while norme > 1e-6:
    y=prox(x)
    p=lamb*(y-x)
    x=x+p
    norme=np.linalg.norm(p)
    cpt+=1
print("La solution x est :\n", x)
print("Le nombre d'itération realisé pour calculer la solution est:\n", cpt)
```

```
La solution x est :
 [[1.89549425]
 [1.97097231]]
Le nombre d'itération realisé pour calculer la solution est:
 150
```

3.2 Algorithme inexact Forward-Backward avec métrique variable

On considère G définie comme la somme des fonctions F et G suivantes :

$$\begin{aligned} G : \mathbb{R} &\longrightarrow \mathbb{R} \\ &\longmapsto G(x) = F(x) + R(x) \end{aligned}$$

où les fonction F et R sont definie par :

$$F(x) = 4\cos(x+1) \quad \text{et} \quad R(x) = |x|$$

1. La fonction R est effectivement propre, convexe et continue donc à fortiori sémi-continue inférieurement.
2. La fonction F est effectivement différentiable et à 4-Lipshist, en effet :

$$\nabla^2 F(x) = -4\cos(x) \leqslant 4\cos(x)$$

donc on prend $A = 4$ (pour l'algorithme Forward-Backward avec métrique variable).

3. La fonction G est coercive, en effet :

$$\lim_{\|x\|\to+\infty} G(x) = \lim_{\|x\|\to+\infty} 4\cos(x+1) + |x| = +\infty$$

Tracé de la fonction G :

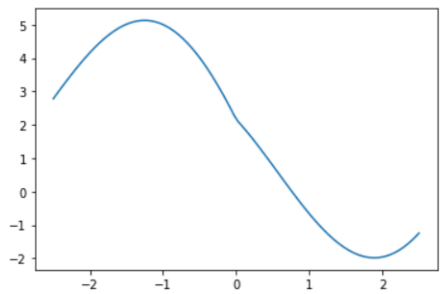
```
x = np.linspace(-2.5,2.5,100)

#fonction F
def fonctf(x):
    return 4*cos(x+1)

#fonction R
def fonctr(x):
    return abs(x)

#fonction G
def fonctg(x):
    return fonctr(x)+fonctf(x)
plt.plot(x,fonctg(x))
```

```
[<matplotlib.lines.Line2D at 0x20297360490>]
```



Comme on l'a déjà expliquer sur le premier chapitre, résoudre le problème $\inf\{G\}$, revient tout d'abord à chercher le point y solution du problème $\inf\{V\}$ par la méthode de la plus grande pente avec recherche linéaire, où V la fonction définie par :

$$\begin{aligned} V :& \mathbb{R} \longrightarrow \mathbb{R} \\ x \longmapsto& V(x) = R(x) + \frac{1}{2}\|x - (x_k - \gamma_k A_k^{-1}\nabla F(x_k))\|^2_{\gamma_k^{-1}A_k} \end{aligned}$$

qui, en retour, nous permet de calculer x solution du problème $\inf\{G\}$.
L'algorithme est le suivante :

$$\begin{cases} \gamma_k = \frac{1}{3}, \quad \lambda_k = \frac{1}{2} \\ A = 4I_2 \\ \tau = 2(\gamma_k\lambda_k)^{-1}, \quad x_0 \in \text{dom}R \ \text{ fixés} \\ x = x_0 \\ \text{while} \ \ \text{norme} > 10^{-6} : \\ \qquad\qquad \text{def} \ \ \text{solution}(x_k) : \\ \qquad\qquad\qquad z_0 = x_k \\ \qquad\qquad\qquad \text{while} \ \ \|\nabla F(z_n) - r_n\| > \tau\|z_n - x_k\| : \\ \qquad\qquad\qquad\qquad\qquad z_{n+1} = z_n - \text{armijo}(z_n)v_n \\ \qquad\qquad\qquad \text{return} \ \ z_n \\ \qquad y_k = \text{solution}(x_k) \\ \qquad p = \lambda(y_k - x_k) \\ \qquad x_{k+1} = x_k + p \\ \qquad \text{norme} = \text{norme(p)} \end{cases}$$

où $\text{armijo}(z_n)$ le **pas d'Armijo** du problème $\inf\{V\}$ et $v_n \in \partial V(z_n)$, on a :

$$v_n = r_n + \gamma_k^{-1}A(z_n - x_k) + \nabla F(x_k), \quad r_n \in \partial R(z_n)$$

Le sous-différentiel de R au point x est donné par :

$$\begin{cases} 1 \quad \text{si} \quad x > 0 \\ \alpha \in [-1, 1] \quad \text{si} \quad x = 0 \\ -1 \quad \text{si} \quad x < 0 \end{cases}$$

```python
gamma=1/3
A=4
lamb=1/2
# fonction R
def fctr(x):
    return la.norm(x)

# le sous differentiel de R
def soudifr(y):
    if y>0:
        r=1
    elif y<0:
        r=-1
    else:
        r=0.5
    return r

#fonction F
def fctf(x):
    return 4*cos(x+1)

#gradiant de F
def gradf(x):
    return -4*sin(x+1)

#fonction V
def v(y,x,A):
    return fctr(y)+0.5*np.vdot((1/gamma)*A*(y-x+gamma*(1/A)*gradf(x)),y-x+gamma*(1/A)*gradf(x))

# le sous differentiel de V
def soudiffv(y,x,A):
    return soudifr(y)+(1/gamma)*A*(y-x)+gradf(x)
```

```python
#Pas d'armijo
beta=1
c1=0.5
def armijo(y,x,A):
    j0=v(y,x,A);
    j_prime_0=-np.vdot(soudiffv(y,x,A),soudiffv(y,x,A))
    t=beta
    while v(y-t*soudiffv(y,x,A),x,A)>(j0+t*c1*j_prime_0):
        t=t/2
    return t

# condition d'arret du boucle en k ( qui est l'ecaryt entre 2 sol soit ici lamb(y-x): initialisé à 1)
norme=1
# Le conteur d'iteration initilisé à 0
cpt=0
x=2
tau=(gamma*lamb)**(-1)*sqrt(A)
while norme > 1e-6:
    z=x
    while (la.norm(gradf(x)+soudifr(z))>tau*sqrt(np.vdot(A*(z-x),z-x))):
        z=z-armijo(z,x,A)*soudiffv(z,x,A)
    p=lamb*(z-x)
    x=x+p
    norme=la.norm(p)
    cpt+=1
print("La solution x est :\n", x)
print("Le nombre d'itération realisé pour calculer la solution est:\n", cpt)
```

```
La solution x est :
 1.8889192721541526
Le nombre d'itération realisé pour calculer la solution est:
 75
```

Bibliographie

[1] Emilie Chouzenoux, Jean-Christophe Pesquet, Audrey Repetti, *Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function*, Fevrier 2013.

[2] Clara Barbanson-Graziussi, *Correction des Effets de Relief en Spectro-imagerie Aéroportée* [sec. 5.2], Novembre 2018.

[3] Assarrar Abdelghani, *Le sous-différentiel au sens de l'analyse convexe et le sous-différentiel au sens de Clarke*,

[4] Grégoire Allaire, *Introduction à la modélisation Mathématique et à la simulation numérique* [sec. 9 et 10], 11 Juillet 2015.

[5] Arman Molla, Samuel Nicolay, *Analyse et optimisation convexe sur des espaces de Banach* [sec. 4], 2017.

[6] Champion Thierry, *Cours d'analyse numérique Master 2 Math.appliquée univ. tln*, 2021-2022.