



Laboratoire EA 2440  
**memiad**

management, économie, modélisation,  
informatique et aide à la décision

# THÈSE

pour obtenir le grade de docteur délivré par  
l'Université des Antilles



EA 2440

ECOLE DOCTORALE N°589 :

Milieu insulaire tropical à risques : protection, valorisation, santé et développement

Discipline : Mathématiques Appliquées

---

## Approche numérique de problèmes d'optimisation et applications

---

PRÉSENTÉE PAR **FLORIAN LABARRE** le 20 décembre 2021

**Directeur de thèse :** Paul-Émile MAINGÉ, MCF-HDR, Université des Antilles

**Co-encadrant de thèse :** Remy-Robert JOSEPH, MCF, Université des Antilles

COMPOSITION DU JURY:

---

<b>Isabelle</b>	<b>CHALENDAR</b>	Professeur	Université Gustave Eiffel	Examinatrice
<b>Jalal</b>	<b>FADILI</b>	Professeur	ENSI de Caen	Rapporteur
<b>Hoi An</b>	<b>LE THI</b>	Professeur	Université de Lorraine.	Examinatrice
<b>Paul-Émile</b>	<b>MAINGÉ</b>	MCF-HDR	Université des Antilles	Directeur
<b>Abdellatif</b>	<b>MOUDAFI</b>	Professeur	Université Aix-Marseille	Rapporteur
<b>Paul Silvère</b>	<b>NUIRO</b>	MCF-HDR	Université des Antilles	Examineur
<b>Hassan</b>	<b>RIAH</b>	Professeur	Université Cadi Ayyad	Rapporteur

---



# Remerciements

Je tiens naturellement à remercier mon directeur de thèse, Paul-Émile MAINGÉ pour la confiance qu'il m'a témoignée en acceptant la direction de mes travaux. Je lui suis reconnaissant de m'avoir fait bénéficier, tout au long de ce travail de sa grande compétence, de sa rigueur intellectuelle, de son dynamisme et de son efficacité certaine que je n'oublierai jamais.

Je remercie grandement Monsieur Remy-Robert JOSEPH d'avoir accepté d'être mon co-encadrant de thèse.

C'est un grand honneur pour moi de remercier les Professeurs Jalal FADILI, Abdellatif MOUDAFI et Hassan RIAHI d'avoir accepté d'être rapporteurs de cette thèse et pour m'avoir fourni une multitude de remarques pertinentes. Je remercie aussi les Professeurs Hoai An LE THI, Isabelle CHALENDAR ainsi que Monsieur Paul Silvère NUIRO d'avoir accepté de laisser leurs activités pour participer à mon jury de thèse.

Je remercie la Collectivité Territoriale de Martinique d'avoir permis le financement de cette thèse, et particulièrement Mme Alix ALMONT pour sa gentillesse.

J'adresse mes chaleureux remerciements à mon frère Julien LABARRE et son conjoint Jeremy Ratier pour leur soutien moral, à ma compagne Célia MARIE-JOSEPH d'avoir toujours cru en moi et de m'avoir motivé tout au long de cette thèse, à mon père Philippe LABARRE et à ma mère Christiane LABARRE de m'avoir supporté et pour leurs encouragements. Je n'oublie pas ma belle famille Roselyne, Henri, Alexa et Karl-Anthony MARIE-JOSEPH d'avoir toujours été présent pour me motiver à terminer ce travail.

Merci à tous ceux qui ont participé de près ou de loin à l'aboutissement de ce travail, je pense en particulier à Kephren BALTHAZARD pour nos nombreuses conversations et à Arnaud NALLAMOUTOU et sa mère.

Je remercie également Mesdames Mariette LAURENT-PEDURAND et Marie-Line REMUS pour avoir toujours répondu à mes questions pour et leur gentillesse.

## Résumé

Cette thèse est consacrée à l'étude de systèmes dynamiques dissipatifs et à la conception d'algorithmes pour la résolution de problèmes d'optimisation (tels que la minimisation de fonctions convexes et, plus généralement, le calcul de zéros d'opérateurs monotones) dans un cadre Hilbertien réel.

Dans un premier temps, nous proposons et étudions, dans un cadre continu, des dynamiques rapides de type Newton pour la résolution de ces problèmes. Le système dynamique considéré est d'abord inspiré par la reformulation d'une équation du second ordre de type Nesterov stabilisée (par l'ajout d'un amortissement géométrique Hessien) pour la minimisation de fonctions convexes. Cette stratégie est ensuite étendue au cas d'opérateurs monotones généraux. Cela nous permet de traiter simultanément les deux problèmes d'optimisation considérés via un système dissipatif du premier ordre ne faisant intervenir que le gradient de la fonction objectif et l'approximation de Yosida d'un opérateur général. Cette dernière formulation constitue un cadre particulièrement adapté à la construction de schémas numériques issus de variantes discrètes de notre modèle continu. Elle permet, en effet, de s'affranchir de la discrétisation temporelle de termes contraignants qui pourraient donner lieu à des schémas coûteux en calculs et/ou limités en application.

Puis, nous nous intéressons à la minimisation structurée de fonctions convexes non lisses, à l'aide d'un algorithme issu d'une discrétisation temporelle de notre modèle du premier ordre. L'algorithme AFB (Accelerated Forward-Backward), récemment proposé comme variante convergente de FISTA (Fast Iterative Shrinkage-Thresholding Algorithm), est obtenu comme cas particulier de notre schéma numérique. Outre les propriétés de convergence bien connues de AFB, nous démontrons la convergence rapide vers zéro des sous-gradients de la fonction objectif. Ceci n'était démontré jusqu'alors que pour des fonctions suffisamment régulières.

Finalement, concernant de la résolution d'inclusions monotones structurées plus générales, nous proposons et nous étudions un algorithme de type Forward-Backward incorporant un terme inertiel, un terme de relaxation ainsi qu'un terme correcteur. Nous démontrons, en plus de la convergence de la méthode, des propriétés proches de celles obtenues dans le cadre de la minimisation convexe structurée en termes de vitesse discrète et de résidus de points fixes.

**Mots-clés :** Systèmes dissipatifs · opérateurs maximaux monotones · minimisation structurée · inclusion monotone · méthode proximal ·

## Abstract

This thesis is devoted to the study of dissipative dynamical systems and to the design of algorithms related to optimization problems (such as convex minimization, and more generally, the computation of zeros of monotone operators) in a real Hilbert framework.

In a first step, we propose and study, in a continuous framework, fast Newton-like dynamics related to these problems. The considered dynamical system is first inspired by the reformulation of a (stabilized) second order Nesterov-type equation (by adding a geometric Hessian damping) for the minimization of convex functions. This strategy is then extended to the case of general monotone operators. This allows us to treat simultaneously the two considered optimization problems, via first-order dissipative system that only involves the gradient of the objective function and the Yosida approximation of a general operator. This last formulation constitutes a particularly well-adapted framework for the design of numerical schemes based upon discrete variants of our continuous model. Indeed, it allows us to get rid of the temporal discretization of constraining terms which could lead to computationally expensive and/or processes of limited applications.

Next, we focus on convex and non-smooth structured minimization problems, by using new class of algorithms derived from a temporal discretization of our first-order model. The process AFB (Accelerated Forward-Backward), which was recently proposed as a convergent variant of FISTA (Fast Iterative Shrinkage-Thresholding Algorithm), is obtained as a special case of our numerical scheme. In addition to the well-known convergence properties for AFB, we prove the fast convergence to zero of the subgradients of the objective function. This last feature was established so far for sufficiently smooth functions only.

Finally, regarding the numerical approach to more general structured monotone inclusions, we propose and study a Forward-Backward algorithm incorporating an inertial term, a relaxation factor and a correction term. In addition to the convergence of the method, we prove similar properties to those obtained for convex structured minimization in term of discrete velocity and fixed-point residual.

**Keywords** : Dissipative system · maximally monotone operators · structured minimization · monotone inclusion · proximal method ·



# Sommaire

Notations	ix
Introduction	1
<b>1 Rappels : travaux séminaux</b>	<b>5</b>
1.1 Contexte . . . . .	5
1.2 Minimisation de fonctions convexes . . . . .	6
1.3 Minimisation stochastique : cas du machine learning . . . . .	13
1.4 Cas général des inclusions monotones . . . . .	19
<b>2 Un modèle général du premier ordre pour des dynamiques continues de type Newton gouvernées par des opérateurs maximaux monotones</b>	<b>23</b>
2.1 Résumé et principaux résultats . . . . .	23
2.2 Article en anglais . . . . .	28
<b>3 Minimisation structurée de fonctions convexes non lisses avec décroissance rapide des sous-gradients</b>	<b>63</b>
3.1 Introduction . . . . .	63
3.2 Préliminaires . . . . .	70
3.3 Résultats généraux pour une analyse de Lyapunov . . . . .	73
3.4 Analyse de la convergence de la méthode . . . . .	80
3.5 Simulations numériques . . . . .	102
3.6 Annexe . . . . .	118

<b>4</b>	<b>Algorithmes forward-backward accélérés pour la résolution d'inclusions monotones structurées</b>	<b>125</b>
4.1	Introduction . . . . .	125
4.2	Propriétés de type Lyapunov . . . . .	132
4.3	Analyse de la convergence de la méthode. . . . .	141
4.4	Une variante plus générale de l'algorithme proposé . . . . .	152
4.5	Simulations numériques . . . . .	156
4.6	Annexe . . . . .	160
<b>5</b>	<b>Conclusions et perspectives</b>	<b>165</b>
	<b>Bibliographie</b>	<b>167</b>

# Notations

## Ensembles

$\emptyset$	: désigne l'ensemble vide.
$\mathbb{N}$	$:= \{0, 1, 2, 3 \dots\}$ désigne l'ensemble des entiers naturels.
$\mathbb{N}^*$	$:= \{1, 2, 3 \dots\}$ désigne l'ensemble des entiers naturels non nuls.
$\mathbb{Z}$	$:= \{\dots - 2, -1, 0, 1, 2 \dots\}$ désigne l'ensemble des entiers relatifs.
$\mathbb{R}$	: l'ensemble des nombres réels.
$\mathbb{R}^+$	: l'ensemble des nombres réels positif.
$L^1([a, b], \mathcal{H})$	: ensemble des fonctions Lebesgue-intégrables sur $[a, b]$ à valeurs dans $\mathcal{H}$ .
$L^\infty[0, \infty)$	: ensemble des fonctions essentiellement bornées sur $[0, +\infty[$ .
$2^{\mathcal{X}}$	: ensemble des parties de l'ensemble $\mathcal{X}$ .

## Espace de Hilbert

$\mathcal{H}$	: espace de Hilbert réel.
$\langle \cdot, \cdot \rangle$	: produit scalaire sur $\mathcal{H}$ .
$\  \cdot \ $	: norme induite par le produit scalaire sur $\mathcal{H}$ .
$\rightarrow$	: convergence forte dans $\mathcal{H}$ .
$\rightharpoonup$	: convergence faible dans $\mathcal{H}$ .
$I$	: désigne l'opérateur identité sur $\mathcal{H}$ .

## Opérateurs multivoques

$A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$	: opérateur multivoque sur $\mathcal{H}$ .
$\text{gr } A := \{(x, u) \in \mathcal{H} \times \mathcal{H}, u \in Ax\}$	: graphe de $A$ .
$\text{dom } A := \{x \in \mathcal{H}, Ax \neq \emptyset\}$	: domaine de $A$ .
$\text{ran } A := A(\mathcal{H})$	: l'image de $A$ .
$\text{zer } A := \{x \in \mathcal{H} \mid 0 \in Ax\}$	: ensemble des zéros de $A$ .
$A^{-1} := \mathcal{H} \rightarrow 2^{\mathcal{H}} : u \mapsto \{x \in \mathcal{H}, u \in Ax\}$	: inverse de $A$ .
$J_\lambda^A := (I + \lambda A)^{-1}$	: résolvante de $A$ d'indice $\lambda > 0$ .
$A_\lambda := \frac{1}{\lambda}(I - J_\lambda A)$	: approximation de Yosida de $A$ d'indice $\lambda > 0$ .

## Fonctions

$\text{gr } f := \{(x, y) \in \mathcal{H} \times \mathbb{R} \mid f(x) = y\}$	: graphe de $f : \mathcal{H} \rightarrow [-\infty, +\infty]$ .
$\text{epi } f := \{(x, y) \in \mathcal{H} \times \mathbb{R} \mid y \geq f(x)\}$	: épigraphe de $f : \mathcal{H} \rightarrow [-\infty, +\infty]$ .
$\text{dom } f := \{x \in \mathcal{H} \mid f(x) < \infty\}$	: domaine effectif de $f : \mathcal{H} \rightarrow [-\infty, +\infty]$ .
$\nabla f$	: gradient de $f : \mathcal{H} \rightarrow \mathbb{R}$ .
$\text{argmin } f$	: ensemble des minimiseurs de $f$ .
$\text{prox}_{\mu f}$	: opérateur proximal de $f$ d'indice $\mu > 0$ .
$\partial f(u)$	: sous-différentiel de Fenchel de $f$ en $u$ .

## Définitions relatives à un opérateur multivoque

Un opérateur multivoque  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  est dit

- **monotone** si

$$\forall (x, y) \in \text{gr } A, \forall (y', x') \in \text{gr } A, \langle y - y', x - x' \rangle \geq 0$$

- **$\alpha$ -fortement monotone** (avec  $\alpha > 0$ ) si  $A - \alpha I$  est monotone, i.e, si

$$\forall (x, y) \in \text{gr } A, \forall (y', x') \in \text{gr } A, \langle y - y', x - x' \rangle \geq \alpha \|x - x'\|^2,$$

- **maximal monotone** si  $A$  est monotone et s'il n'existe pas d'opérateur  $B : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  tel que  $\text{gr } A \subset \text{gr } B$  et  $A \neq B$ .
- **$\beta$ -cocoercif** (avec  $\beta > 0$ ) si  $A$  est univoque et si

$$\forall (x, x') \in \mathcal{H}^2, \langle Ax - Ax', x - x' \rangle \geq \beta \|Ax - Ax'\|^2,$$

# Introduction

Cette thèse de mathématiques porte sur la conception de systèmes dynamiques dissipatifs pour la résolution d'inclusions monotones structurés dans les espaces de Hilbert réels. Le problème général étudié est le suivant

$$(A) \quad \text{trouver } x \in \mathcal{H} \text{ tel que } 0 \in Ax + Bx,$$

où  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  et  $B : \mathcal{H} \rightarrow \mathcal{H}$  sont des opérateurs maximaux monotones et où  $\mathcal{H}$  désigne un espace de Hilbert réel. Le problème (A) intervient dans de nombreux domaines notamment dans les équations aux dérivées partielles [25, 33, 50, 95] en statistiques, en machine-learning [59, 107], en théorie des jeux [33], dans le traitement du signal [49, 33] et de l'image [41, 39, 99].

Un exemple bien connu du problème (A) est la minimisation d'une fonction convexe que l'on présente sous la forme structurée avec  $A = \partial g$  et  $B = \nabla f$  où  $\partial g$  désigne le sous-différentiel d'une fonction  $g : \mathcal{H} \rightarrow ]-\infty, +\infty]$  convexe et semi-continue inférieurement et  $\nabla f$  est le gradient d'une fonction  $f : \mathcal{H} \rightarrow ]-\infty, +\infty[$  convexe et différentiable.

De nombreuses techniques aussi bien discrètes que continues ont déjà été proposées pour traiter le problème (A) (méthode de la plus grande pente, du point proximal, de Douglas-Rachford, Forward-Backward, FISTA, Accelerated Forward-Backward etc.). Parmi ces méthodes, on peut citer AFB qui correspond à une version discrète d'un système dissipatif du second ordre permettant de résoudre le problème (A) dans le cas où  $f$  est lisse et  $g$  est non lisse. Une particularité de cet algorithme est qu'il produit une convergence en valeur en  $o(n^{-2})$  au lieu de  $\mathcal{O}(n^{-1})$  proposé par la méthode Forward-Backward qui est elle basée sur une discrétisation en temps de la méthode de la plus grande pente.

On se propose ici d'améliorer les résultats existants via l'introduction de nouveaux systèmes dissipatifs et de schémas numériques associés.

Le chapitre 1 présente des rappels concernant les principales méthodes de résolutions d'inclusions monotones. Dans ce chapitre, nous rappelons dans un premier temps les algorithmes classiques de minimisation de fonctions convexes existants ainsi les techniques couramment utilisées dans le cadre de minimisation stochastique et enfin nous rappelons quelques méthodes classiques de résolution d'inclusions monotones générales en précisant leurs avantages et inconvénients.

Le chapitre 2 est consacré à l'étude d'un modèle continu de type Newton qui s'avère approprié, du point de vue numérique, à la résolution de problèmes de minimisation convexe et plus généralement d'inclusions monotones. Des algorithmes permettant de résoudre ces problèmes ont été récemment inspirés par des discrétisations implicites en temps de la version continue (stabilisée) de la méthode du gradient accéléré de Nesterov faisant intervenir un terme d'amortissement Hessien supplémentaire (afin d'atténuer les effets d'oscillations observées au niveau des trajectoires). Cependant, en raison de la présence du terme Hessien, ces variantes discrètes nécessitent plusieurs évaluations (par itération) du gradient ou de l'opérateur proximal. Nous proposons, en ce sens, un modèle du premier ordre ne faisant plus intervenir le terme Hessien et qui peut être étendu au cas d'un opérateur maximal monotone quelconque. Notre modèle du premier ordre provient de la reformulation d'une variante liée à une équation de type Nesterov. Nous étudions simultanément les trajectoires de notre modèle en ce qui concerne la minimisation convexe et les inclusions monotones en considérant notre système lorsqu'il est gouverné par la somme du gradient d'une fonction convexe différentiable et (à une constante multiplicative près) de l'approximation de Yosida d'un opérateur maximale monotone. Il s'avère que notre modèle du premier ordre offre un nouveau cadre pour les variantes discrètes, tout en conservant les principales caractéristiques asymptotiques de l'équation de type Nesterov (stabilisée). Nous démontrons l'existence et l'unicité d'une solution pour notre système dynamique, puis, en nous appuyant sur une analyse de Lyapunov, nous étudions le comportement asymptotique des trajectoires. Finalement, nous présentons deux algorithmes (un pour la minimisation convexe structurée et un autre pour la résolution d'inclusion monotone générale) de type Forward-Backward relaxé obtenus via une discrétisation en temps de notre système dynamique.

Le chapitre 3 (sur lequel est basé l'article [72]) est destiné à l'étude d'une variante discrète du système dynamique étudié au chapitre 2 concernant la minimisation convexe structurée non lisse. La méthode AFB (« Accelerated Forward-Backward »)

---

qui a été récemment discutée comme une variante améliorée de FISTA (« Fast Iterative Thresholding Algorithm ») est retrouvée comme cas particulier de l'algorithme proposé pour un certain jeu de paramètres. Sur la base d'une suite de type énergie, nous montrons que l'algorithme étudié dans ce chapitre jouit des propriétés bien connues de AFB, à savoir que celui-ci génère des suites convergentes  $(x_n)$  qui minimisent les valeurs de la fonction objectif avec un taux de  $o(n^{-2})$ . Notre méthode offre, cependant, quelques degrés de liberté supplémentaires comparée à AFB grâce à un réglage approprié des paramètres qui peuvent être utilisés pour influencer (accélérer) la convergence des itérés. Une autre spécificité importante du schéma proposé est qu'il peut être considéré comme un modèle discret issu d'une formulation du premier ordre du système dynamique de type Newton étudié au chapitre 2. Ceci nous permet d'étendre au cadre non lisse, une propriété de convergence rapide vers zéro des sous-gradients, établie jusqu'à présent pour les dynamiques discrètes de type Newton avec des potentiels lisses uniquement. En particulier, comme nouveau résultat, nous montrons que cette dernière propriété s'applique également à AFB. Des simulations numériques sont en outre réalisées afin d'illustrer les propriétés des algorithmes considérés par rapport à d'autres algorithmes existants.

Le chapitre 4 est consacré à l'étude d'une généralisation, au cas des inclusions monotones structurées, de l'algorithme étudié au chapitre 3 dans le cadre de la minimisation convexe structurée. Plus précisément, nous développons un algorithme de type Forward-Backward pour le calcul des zéros de la somme d'un nombre fini d'opérateurs maximaux monotones. On considère, en premier lieu, une modification de la méthode classique Forward-Backward pour deux opérateurs monotones généraux, en incorporant un terme inertiel (proche des techniques d'accélération introduites par Nesterov), un terme de relaxation constant ainsi qu'un terme correcteur. Dans un cadre Hilbertien réel, nous démontrons la convergence faible des itérés générés par notre méthode ainsi que des taux  $o(n^{-1})$  en termes de vitesse discrète et de résidu du point fixe au lieu des taux classiques de  $\mathcal{O}(n^{-1})$ . Dans un second temps, l'algorithme proposé est ensuite adapté à des inclusions monotones plus générales.

Le chapitre 5 présente, quant à lui, les conclusions de cette thèse ainsi que quelques perspectives de recherches en lien avec ce travail.



# Chapitre 1

## Rappels : travaux séminaux

---

### 1.1 Contexte

Dans ce chapitre,  $\mathcal{H}$  désigne un espace de Hilbert réel pour lequel on note  $\langle \cdot, \cdot \rangle$  le produit scalaire et  $\| \cdot \| := \sqrt{\langle \cdot, \cdot \rangle}$ , la norme associée. Considérons le problème d'inclusion monotone général suivant

$$(1.1.1) \quad \text{trouver } x^* \in \mathcal{H} \text{ tel que } 0 \in Ax^* + Bx^*,$$

où  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  et  $B : \mathcal{H} \rightarrow \mathcal{H}$  sont deux opérateurs maximaux monotones. Nous présentons quelques algorithmes classiques pour la résolution du problème (1.1.1). Nous commencerons, par nous restreindre au cas particulier où  $A = \partial f$  et  $B = \partial g$  où  $f, g : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  sont des fonctions convexes, propres et semi-continues inférieurement. Avec ces considérations, le problème (1.1.1) devient le problème de minimisation suivant

$$(1.1.2) \quad \text{trouver } x^* \in \mathcal{H} \text{ tel que } \Theta(x^*) = \min_{x \in \mathcal{H}} \{\Theta(x) := f(x) + g(x)\}.$$

Nous supposons que  $S := \operatorname{argmin}_{\mathcal{H}} \Theta(x) \neq \emptyset$  et nous passons en revue les algorithmes de minimisation classiques ainsi que leurs avantages et inconvénients suivant leur domaine d'applications et les hypothèses requises sur les fonctions  $f$  et  $g$ .

## 1.2 Minimisation de fonctions convexes

Dans cette section, nous rappelons les méthodes classiques de type de descente de gradient (proximale ou non) et nous étudions leurs différentes évolutions. Notons que ces méthodes de résolution sont grandement appréciées compte tenu de leur simplicité et des faibles hypothèses (en général) requises pour leurs convergences.

### 1.2.1 La méthode de la plus grande pente

Considérons le problème (1.1.2) avec  $g \equiv 0$ , c'est-à-dire

$$(1.2.3) \quad \min_{x \in \mathcal{H}} f(x),$$

où  $f : \mathcal{H} \rightarrow \mathbb{R}$  est une fonction convexe, propre et continûment différentiable. Les systèmes dynamiques sont couramment utilisés pour résoudre le problème (1.2.3). L'un des plus connus est donné par

$$(1.2.4) \quad \dot{x}(t) + \nabla f(x(t)) = 0.$$

Notons que les points d'équilibres des trajectoires  $x(\cdot)$  générées par (1.2.4) sont exactement les éléments qui minimisent de  $f$ . D'un point de vue pratique, une discrétisation de (1.2.4) donne pour  $\lambda_k > 0$

$$(1.2.5) \quad \frac{\mathbf{x}_{k+1} - \mathbf{x}_k}{\lambda_k} + \nabla f(\mathbf{x}_k) = 0 \iff \mathbf{x}_{k+1} = \mathbf{x}_k - \lambda_k \nabla f(\mathbf{x}_k).$$

Cette méthode générale est connue sous le nom de « *méthode de la plus grande pente* » ou « *méthode de descente de gradient* ». Il existe plusieurs variantes basées sur la descente de gradient (à pas fixe, à pas optimal, etc.) suivant les valeurs prises par la suite  $(\lambda_k)$ . Lorsque  $f$  est à gradient lipschitzien, la méthode de descente du gradient associée au problème (1.2.3) possède un taux de convergence en valeurs suivant

$$(1.2.6) \quad f(\mathbf{x}_k) - f(\mathbf{x}^*) = \mathcal{O}(k^{-1}),$$

(voir [83, Corollary 2.1.2]) et si de plus  $f$  est fortement convexe, alors

$$(1.2.7) \quad f(\mathbf{x}_k) - f(\mathbf{x}^*) = \mathcal{O}(\rho^k),$$

pour un certain  $0 < \rho < 1$  (voir [83, Theorem 2.1.15]), ce qui représente une convergence linéaire.

## 1.2.2 La méthode des sous-gradients

Bien que la méthode de descente de gradient génère un algorithme simple à implémenter et nécessite des hypothèses relativement faibles, celle-ci requiert néanmoins que la fonction à minimiser soit différentiable. Ainsi, la méthode de la descente de gradient ne s'applique pas à une fonction  $F$  non différentiable, ce qui est notamment le cas lorsque l'on considère le problème (1.1.2) avec une fonction  $g$  non différentiable. Afin de contourner ce problème, une autre approche consiste à remplacer le gradient par la notion de sous-gradient.

**Définition 1.2.1.** Soit  $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  une fonction convexe propre. On dit que  $u \in \mathcal{H}$  est un sous-gradient de  $f$  en  $x \in \text{dom } f$  si

$$(1.2.8) \quad \forall y \in \mathcal{H}, \quad f(y) \geq f(x) + \langle u, y - x \rangle.$$

L'ensemble des sous-gradients de  $f$  en  $x$  est appelé **sous-différentiel de Fenchel** de  $f$  en  $x$  et est noté  $\partial f(x)$ . Autrement dit, on a

$$(1.2.9) \quad \partial f(x) := \{u \in \mathcal{H} \mid \forall y \in \mathcal{H}, f(y) \geq f(x) + \langle u, y - x \rangle\}.$$

Dans le cas où  $x \in \mathcal{H} \setminus \text{dom } f$ , on pose  $\partial f(x) = \emptyset$ .

**Remarque 1.2.1.** Notons que  $\partial f$  peut être vu comme un opérateur multivoque de  $\mathcal{H}$  dans  $2^{\mathcal{H}}$  et que lorsque  $f$  est différentiable dans la Définition (1.2.1), nous avons pour tout  $x \in \mathcal{H}$ ,  $\partial f(x) = \{\nabla f(x)\}$ . De plus, d'après la Définition (1.2.1), on observe qu'un élément  $x^* \in \mathcal{H}$  minimise  $f$  sur  $\mathcal{H}$  si et seulement si  $0 \in \partial f(x^*)$ .

Lorsqu'on ne suppose plus la régularité de la fonction  $f$ , l'équation (1.2.4) devient l'équation du gradient généralisé.

$$(1.2.10) \quad \dot{x}(t) + \partial f(x(t)) \ni 0.$$

Cette dernière a été étudiée par plusieurs auteurs (voir [21, 35]). Ceux-ci ont notamment démontré la convergence faible des trajectoires vers un élément de  $\text{argmin}_{\mathcal{H}} f$ ,

dès lors que  $\operatorname{argmin}_{\mathcal{H}} f \neq \emptyset$ . Ce résultat invite alors à considérer une discrétisation de (1.2.10) comme alternative à la méthode (1.2.5), via l'algorithme donné pour  $(\lambda_k) \subset ]0, +\infty[$  par :

$$(1.2.11) \quad \mathbf{x}_{k+1} = \mathbf{x}_k - \lambda_k g_k \quad \text{où} \quad g_k \in \partial f(x_k).$$

La méthode des sous-gradients conserve la simplicité de celle de la descente de gradient. Cependant, contrairement à celle-ci, il peut arriver que  $-g_k$  ne soit pas une direction de descente de  $f$  en  $\mathbf{x}_k$  (voir [29]). Ainsi, à chaque itération, il est nécessaire de garder à jour une liste conservant la trace de la plus petite valeur de  $f(\mathbf{x}_k)$  jusqu'à l'itération courante, pour cela, la quantité  $f_k^*$  suivante est considérée

$$(1.2.12) \quad f_k^* := \min_{i=0,1,\dots,k} f(\mathbf{x}_i),$$

la suite  $(f_k^*)_{k \geq 0}$  ainsi générée est clairement décroissante. Sous certaines conditions sur la suite de pas  $(\lambda_k)_{k \in \mathbb{N}}$  et en supposant que la fonction  $f$  soit lipschitzienne, il a été démontré (voir [23, Theorem 8.13] et [83, Theorem 3.2.2]) que

$$(1.2.13) \quad f_k^* - f(\mathbf{x}^*) = \mathcal{O}(1/\sqrt{k}),$$

où  $f(\mathbf{x}^*)$  est une valeur optimale de  $f$ . Et si de plus, on suppose  $f$  fortement convexe, alors

$$(1.2.14) \quad f_k^* - f(\mathbf{x}^*) = \mathcal{O}(1/k).$$

On constate alors que la méthode des sous-gradients se révèle être moins performante que la descente de gradient. Ainsi l'utilisation des sous-gradients pour résoudre le problème (1.1.2) ne permet pas d'avoir une convergence comparable à la descente de gradient donné par l'algorithme 1. Une autre approche s'avère donc nécessaire.

### 1.2.3 La méthode de descente proximale

Considérons toujours le problème (1.1.2) avec  $g \equiv 0$ , c'est-à-dire

$$(1.2.15) \quad \min_{x \in \mathcal{H}} f(x),$$

avec cette fois une fonction  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  convexe, propre et semi-continue inférieurement.

**Définition 1.2.2** (Opérateur proximal). *Soit  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  une fonction convexe propre et semi-continue inférieurement. On appelle **opérateur proximal** de  $f$  de paramètre  $\lambda > 0$  l'opérateur  $\text{prox}_{\lambda f}$  défini pour tout  $x \in \mathcal{H}$  par*

$$(1.2.16) \quad \text{prox}_{\lambda f}(x) := \underset{v \in \mathcal{H}}{\operatorname{argmin}} \left( f(v) + \frac{1}{2\lambda} \|v - x\|_2^2 \right).$$

**Remarque 1.2.2.** *Notons que pour  $x \in \mathcal{H}$ , la fonction  $f(\cdot) + 1/(2\lambda)\|\cdot - x\|_2^2$  intervenant dans (1.2.16) est fortement convexe et n'est pas partout égale à  $+\infty$ , celle-ci admet donc un unique élément  $v^*$  minimisant  $f(v) + 1/(2\lambda)\|v - x\|_2^2$  pour tout  $v \in \mathcal{H}$ . Ainsi l'opérateur  $\text{prox}_{\lambda f}$  donné par (1.2.16) est bien défini.*

Une méthode développée vers 1970 par Martinet [75, 76] (voir aussi [69, 89]) consiste à choisir  $\mathbf{x}_0 \in \mathcal{H}$  ainsi qu'une suite  $(\lambda_k) \subset ]0, +\infty[$  et effectuer pour  $k \geq 0$

$$(1.2.17) \quad \mathbf{x}_{k+1} = \text{prox}_{\lambda_k f}(\mathbf{x}_k).$$

**Remarque 1.2.3.** *Rappelons que l'opérateur proximal peut être vu comme une généralisation de l'opérateur de projection [22, 45]. En effet, lorsque  $f := \iota_C$  est la fonction indicatrice d'un convexe fermé  $C \subset \mathcal{H}$  alors pour tout  $\lambda > 0$  on a*

$$\text{prox}_{\lambda f}(x) = \underset{v \in \mathcal{H}}{\operatorname{argmin}} \left\{ \iota_C(v) + \frac{1}{2\lambda} \|v - x\|_2^2 \right\} = \underset{v \in C}{\operatorname{argmin}} \left\{ \frac{1}{2\lambda} \|v - x\|_2^2 \right\} = P_C(x),$$

où  $P_C$  désigne l'opérateur de projection sur  $C$ .

Notons par ailleurs que pour  $f \equiv 0$ , on obtient pour tout  $\lambda > 0$

$$(1.2.18) \quad \text{prox}_{\lambda f}(x) = \underset{v \in \mathcal{H}}{\operatorname{argmin}} \left\{ \frac{1}{2\lambda} \|v - x\|_2^2 \right\} = x.$$

**Remarque 1.2.4.** *L'opérateur proximal peut aussi être vu comme un moyen de créer un schéma de descente de gradient implicite. En effet, en supposant que la fonction  $f$  introduite dans la Définition 1.2.2 soit continûment différentiable, une discrétisation implicite en temps de l'équation différentielle (1.2.4) donne pour  $\lambda_k > 0$*

$$(1.2.19) \quad \frac{\mathbf{x}_{k+1} - \mathbf{x}_k}{\lambda_k} + \nabla f(\mathbf{x}_{k+1}) = 0.$$

Ainsi, en résolvant l'équation  $\mathbf{x}_{k+1} + \lambda_k \nabla f(\mathbf{x}_{k+1}) = \mathbf{x}_k$  pour  $\mathbf{x}_{k+1}$  et en utilisant la définition de l'opérateur résolvante, on obtient (1.2.17).

Le principal défaut de la méthode proximale (1.2.17) est qu'elle est difficile à mettre en pratique, car, d'après sa définition (1.2.2), à chaque itération il faut minimiser la quantité  $f(\cdot) + \frac{1}{2\lambda} \|\cdot - x\|_2^2$  pour  $x \in \mathcal{H}$ , ce qui s'avère être aussi compliqué que de minimiser  $f(\cdot)$  sauf si l'opérateur proximal est simple à calculer. Dans ce dernier cas, la méthode proximale se révèle être efficace pour résoudre le problème (1.1.2).

Lorsque la fonction  $g$  est prise en compte dans le problème (1.1.2), l'algorithme (1.2.17) devient l'algorithme de descente de gradient proximal (ou algorithme *Forward-Backward*) (voir [34, 55, 89]) donné par

$$(1.2.20) \quad \mathbf{x}_{k+1} = \text{prox}_{\lambda_k g}(\mathbf{x}_k - \lambda_k \nabla f(\mathbf{x}_k)),$$

où  $(\lambda_k)$  est une suite réelle strictement positive et où l'application  $\text{prox}$  est définie par (1.2.17). À l'aide de la Remarque 1.2.4, on observe que le schéma (1.2.20) combine une étape de descente explicite (forward) concernant la fonction régulière  $f$  et une étape implicite (backward) concernant la fonction  $g$ .

Notons que le schéma (1.2.20) peut s'écrire sous la même forme que (1.2.5),

$$(1.2.21) \quad \mathbf{x}_{k+1} = \mathbf{x}_k - \lambda_k M_{\lambda_k}(\mathbf{x}_k),$$

où  $M_\lambda(x) := \frac{1}{\lambda}(x - \text{prox}_{\lambda g}(x - \lambda \nabla f(x)))$ . L'opérateur  $M_\lambda$  est souvent appelé opérateur gradient généralisé, cette appellation se justifie par le fait qu'en prenant  $g \equiv 0$  dans l'expression de  $M_\lambda$ , on obtient  $M_\lambda = \nabla f$  d'après (1.2.18).

Faisons à présent les hypothèses suivantes :

- (A1)  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  est une fonction convexe différentiable dont le gradient est  $L$ -lipschitzien,
- (A2)  $g : \mathcal{H} \rightarrow ]-\infty, +\infty]$  est une fonction convexe propre et semi-continue inférieurement,
- (A3)  $S := \underset{\mathcal{H}}{\text{argmin}}(f + g) \neq \emptyset$ .

Il a été démontré, sous les hypothèses (A1), (A3) et (A3) ainsi que pour  $\lambda_k := \lambda \in ]0, 1/L[$ , que les itérés  $\mathbf{x}_k$  générés par l'algorithme de la descente de gradient

proximale (1.2.20) convergent faiblement vers un élément de  $S$  et vérifient les taux  $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| = \mathcal{O}(k^{-1})$  et  $(f + g)(\mathbf{x}) - \min_{\mathcal{H}}(f + g) = \mathcal{O}(k^{-1})$  (voir [22, 23]). Comme énoncé dans [44] (voir aussi [88]), cette méthode de convergence encore pour  $\lambda \in ]0, 2/L[$  mais ne permet pas d'avoir une estimation précise de la vitesse de convergence. Cependant lorsque  $f$  est fortement convexe, la suite  $(\mathbf{x}_k)$  générée par la méthode de gradient proximale (1.2.20) avec un pas constant  $\lambda_k := \lambda \in ]0, 2/L[$  converge linéairement vers l'unique élément de  $S$ .

Dans la pratique, bien que la méthode de descente proximale soit efficace et qu'elle ne demande pas trop de contraintes sur les fonctions  $f$  et  $g$ , elle s'avère être relativement lente notamment lorsque  $f$  n'est pas fortement convexe. Elle est en fait comparable à la méthode (1.2.5) que l'on retrouve en prenant  $g \equiv 0$  dans (1.1.2) (voir Remarque (1.2.3)).

#### 1.2.4 Méthodes de gradients (proximales) accélérées : boule pesante avec frottement, FISTA et AFB

En 1964 Polyak [91] développa une méthode connue sous le nom de « boule pesante avec frottement » (Heavy-ball with friction), qui s'écrit pour  $(\mathbf{x}_0, \mathbf{x}_1) \in \mathcal{H}^2$  et pour  $k \geq 0$

$$(1.2.22) \quad \mathbf{HBF} : \begin{cases} \mathbf{y}_k &= \mathbf{x}_k + \beta_k(\mathbf{x}_k - \mathbf{x}_{k-1}), \\ \mathbf{x}_{k+1} &= \mathbf{y}_k - \lambda_k \nabla f(\mathbf{x}_k), \end{cases}$$

où  $\{\beta_k, \lambda_k\} \subset ]0, +\infty[$ . La différence par rapport à la méthode de gradient classique est qu'à chaque itération, le point extrapolé  $y_k$  est utilisé au lieu de  $\mathbf{x}_k$ . L'accélération obtenue s'explique par le fait que la nouvelle itération est donnée en prenant un pas qui est une combinaison de la direction  $\mathbf{x}_k - \mathbf{x}_{k-1}$  et de la direction anti-gradient  $-\nabla f(\mathbf{x}_k)$ .

**Remarque 1.2.5.** De nombreux auteurs [2, 6, 10, 61, 56] ont analysé le modèle continu général provenant de (1.2.22) donné, pour une fonction réelle positive  $\lambda$ , par

$$(1.2.23) \quad \ddot{x}(t) + \lambda(t)\dot{x}(t) + \nabla f(x(t)) = 0.$$

Cette dernière peut être vue comme une amélioration du système (1.2.4) par l'ajout d'un terme d'amortissement  $\lambda(t)\dot{x}(t)$ . Dans le cas où  $\lambda(t) = \lambda > 0$ , Alvarez [2]

(voir aussi Attouch-Goudou-Redont [10]) a montré que les trajectoires de (1.2.23) convergent faiblement vers un élément de  $\operatorname{argmin}_{\mathcal{H}} f$  dès lors que  $f : \mathcal{H} \rightarrow \mathbb{R}$  est convexe et que  $\operatorname{argmin}_{\mathcal{H}} f \neq \emptyset$ . D'autres auteurs (voir [36, 37]) se sont intéressés à (1.2.23) pour une fonction  $\lambda(\cdot)$  plus générale et vérifiant certaines propriétés.

Le schéma (1.2.22) est à la base de nombreux algorithmes de type inertiels. Dans le cas des méthodes proximales, l'algorithme FISTA (*Fast Iterative SoftThreshold Algorithm*), proposé par Beck-Teboulle [24], figure parmi les plus connus. Notons que bien que le nom de cette méthode fasse référence à l'opérateur « soft-thresholding », celle-ci a été étendue à un opérateur plus général. L'algorithme FISTA est basé sur un schéma proposé en 1983 par Nesterov [81] et repris par Güler [57] qui se présente sous la forme

$$(1.2.24) \quad \text{FISTA} : \begin{cases} a_k &= \frac{1 + \sqrt{4a_{k-1}^2 + 1}}{2}, \\ \mathbf{y}_k &= \mathbf{x}_k + \frac{a_{k-1} - 1}{a_k} (\mathbf{x}_k - \mathbf{x}_{k-1}), \\ \mathbf{x}_{k+1} &= \operatorname{prox}_{\lambda_k g}(\mathbf{y}_k - \lambda_k \nabla f(\mathbf{y}_k)). \end{cases}$$

Bien que les schémas (1.2.24) et (1.2.22) fassent tous les deux intervenir un terme inertiel, la différence majeure se situe dans l'évaluation du gradient. En effet, celle-ci se fait en  $\mathbf{y}_k$  dans (1.2.24) et en  $\mathbf{x}_k$  pour (1.2.22). Cette différence peut être vue comme une correction de la direction à prendre pour le calcul de l'itéré  $\mathbf{x}_{k+1}$  suivant. Il a été démontré [24, Theorem 4.4], sous les hypothèses **(A1)**, **(A2)** et **(A3)**, que la suite  $(\mathbf{x}_k)_{k \geq 0}$  produite par l'algorithme (1.2.24) avec un pas constant  $\lambda_k := \lambda \in ]0, 1/L]$  vérifie  $(f + g)(\mathbf{x}) - \min_{\mathcal{H}}(f + g) = \mathcal{O}(k^{-2})$  (lorsque  $k \rightarrow +\infty$ ).

**Remarque 1.2.6.** *Le taux de convergence  $\mathcal{O}(k^{-2})$  obtenu par (1.2.24) concernant la convergence en valeur est une amélioration comparée à la méthode de descente proximale (1.2.20) et ce malgré le fait que les calculs soient essentiellement les mêmes (i.e, une descente de gradient et un calcul de prox). Notons qu'il n'a pas été prouvé que la suite des itérés  $(\mathbf{x}_k)$  générée par la méthode (1.2.24) converge. On doit cependant à Chambolle-Dossal [40] d'avoir démontré la convergence faible des itérés  $(\mathbf{x}_k)$  en utilisant une variante de FISTA, connue sous le nom d'algorithme AFB (« Accelerated Forward-Backward »). Cette dernière s'écrit pour tout  $k \geq 1$  par*

$$(1.2.25) \quad \text{AFB} : \begin{cases} \mathbf{y}_k &= \mathbf{x}_k + \frac{k-1}{k+\alpha} (\mathbf{x}_k - \mathbf{x}_{k-1}), \\ \mathbf{x}_{k+1} &= \operatorname{prox}_{\lambda g}(\mathbf{y}_k - \lambda \nabla f(\mathbf{y}_k)). \end{cases}$$

Attouch-Peypouquet [14] ont démontré, pour  $\alpha > 3$ , que la suite  $(\mathbf{x}_k)_k$  générée par (1.2.25) satisfait  $\Theta(\mathbf{x}_k) - \Theta(\mathbf{x}^*) = o(k^{-2})$  (où  $\mathbf{x}^* \in S$ ) ainsi que  $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| = o(k^{-1})$  où  $k$  désigne le nombre d'itérations.

Notons que plusieurs auteurs [16, 19, 20, 105] ont étudié le modèle continu associé à (1.2.25) (dans le cas non proximal) donné par

$$(1.2.26) \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)).$$

Cette équation est à rapprocher de (1.2.23) dans le cas particulier où  $\lambda(t) = \alpha t^{-1}$  avec  $\alpha > 0$ , ce dernier terme agissant comme un amortissement visqueux (autrement dit qui tend lentement vers 0). Il a notamment été prouvé (voir [16]) que, pour  $\alpha \geq 3$ , les trajectoires générées par (1.2.26) vérifient  $f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}(t^{-2})$  lorsque  $t \rightarrow +\infty$ .

## 1.3 Minimisation stochastique : cas du machine learning

### 1.3.1 Minimisation convexe régularisée

Considérons le problème de minimisation structuré (1.1.2)

$$(1.3.27) \quad \text{trouver } x^* \in \mathcal{H} \text{ tel que } \Theta(x^*) = \min_{x \in \mathcal{H}} \{\Theta(x) := f(x) + g(x)\},$$

où les fonctions  $f$  et  $g$  sont telles que

**(B1)**  $f := \frac{1}{n} \sum_{i=1}^n f_i$  où chaque  $f_i : \mathcal{H} \rightarrow \mathbb{R}$  sont des fonctions convexes et différentiables,

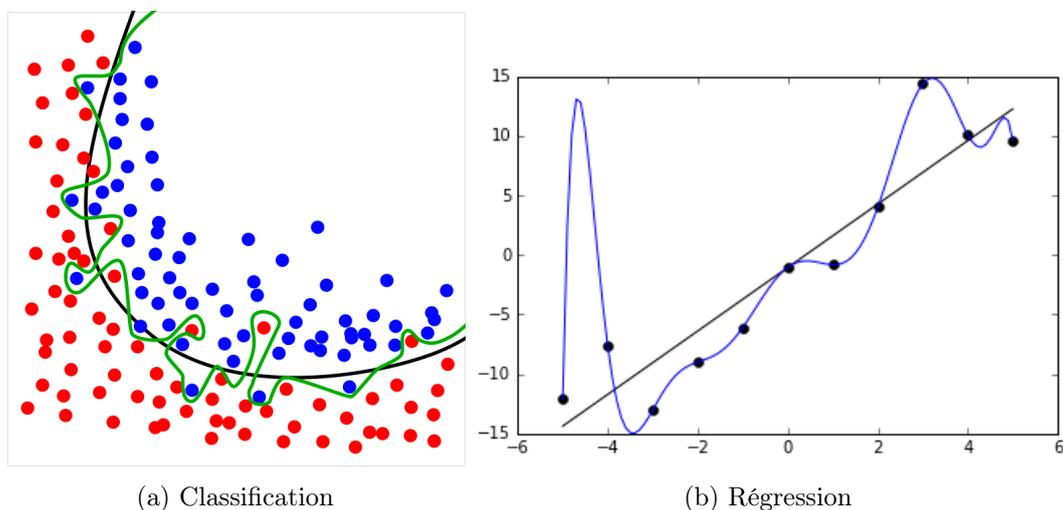
**(B2)**  $g : \mathcal{H} \rightarrow \mathbb{R} \cup \{-\infty, +\infty\}$  est une fonction convexe propre et inférieurement.

On retrouve couramment ce type de problème en statistique et en machine learning pour effectuer des régressions ou des classifications régularisées. Pour des données d'entraînement  $(a_i, b_i)$ , les fonctions  $f_i$  les plus couramment utilisées sont

- **Lasso** : avec  $f_i(w) := \frac{1}{2}(b_i - \langle x, a_i \rangle)^2$  et  $g(x) := \mu \|x\|_1$ ,
- **Ridge régularisée** :  $f_i(x) := \frac{1}{2}(b_i - \langle x, a_i \rangle)^2$  et  $g(w) := \mu \|w\|_2^2$ ,
- **Logistique régularisée** :  $f_i(x) := \ln(1 + \exp(-b_i \langle x, a_i \rangle))$ , et  $g(x) = \mu \|x\|_1$  (ou alors  $g(w) = \mu \|w\|_2^2$ ),
- **Elastic net** :  $f_i(x) := \ln(1 + \exp(-b_i \langle x, a_i \rangle))$ , et  $g(x) = \mu_1 \|x\|_1 + \mu_2 \|x\|_2^2$ .

L'ajout d'un terme régularisateur  $g$  par rapport au problème (1.3.28) se justifie par le besoin d'éviter le *surapprentissage* (« overfitting » en anglais), un phénomène qui intervient lorsque les valeurs obtenues par l'algorithme s'adaptent trop parfaitement aux données d'apprentissage et ne permettent pas de généraliser le modèle prédictif. On se retrouve alors avec un modèle qui épouse parfaitement les données d'entraînement, mais qui prédit très mal de nouvelles données. Le but de la régularisation est de pénaliser les valeurs extrêmes des paramètres lors de l'apprentissage (entraînement), car en effet, ces valeurs sont généralement dues à un surapprentissage.

La Figure 1.3.1 illustre deux cas de surapprentissage. L'un lors d'une classification (a) et l'autre lors d'une régression (b). Concernant, l'image (a), la courbe en vert représente un cas de surapprentissage, car bien que cette dernière sépare effectivement les points bleus des points rouges, la courbe ainsi obtenue ne permet pas une généralisation lors de l'ajout d'un nouveau point (bleu ou rouge). De même, concernant l'image (b) de la Figure 1.3.1, la courbe en bleu illustre un cas de surapprentissage lors d'une régression. Dans les deux images (a) et (b), la courbe en noir, bien que n'étant pas parfaite (car ne séparant ou ne passant pas par tous les points) est malgré tout une courbe idéale car elle est simple et permet d'obtenir facilement des prédictions pour de nouvelles données.

FIGURE 1.3.1 – Phénomène de surapprentissage<sup>1 2</sup>.

1. <https://upload.wikimedia.org/wikipedia/commons/1/19/Overfitting.svg>  
2. [https://upload.wikimedia.org/wikipedia/commons/6/68/Overfitted\\_Data.png](https://upload.wikimedia.org/wikipedia/commons/6/68/Overfitted_Data.png)

### 1.3.2 Méthodes de descente de gradient stochastiques

Dans cette section, nous rappelons la méthode de descentes de gradient stochastique ainsi que ces différentes évolutions. Dans un premier, nous nous restreignons au cas de (1.3.27) avec  $g \equiv 0$ , c'est à dire au problème suivant

$$(1.3.28) \quad \min_{x \in \mathcal{H}} f(x) := \frac{1}{n} \sum_{i=1}^n f_i(x).$$

Si l'on effectue la méthode de descente de gradient classique pour résoudre le problème (1.3.28), nous devons exécuter l'itération suivante

$$(1.3.29) \quad \mathbf{GD} : \mathbf{x}_{k+1} \leftarrow \mathbf{x}_k - \lambda_k \nabla f(\mathbf{x}_k),$$

avec  $(\lambda_k) \subset \mathbb{R}^+$  et où  $\nabla f$  désigne le gradient de  $f$ . D'après (1.2.6) et (1.2.7), la suite  $(\mathbf{x}_k)_{k \in \mathbb{N}}$  générée par le schéma (1.3.29) donne, dans le cas où  $\lambda$  est constant, un taux de convergence en valeurs sous-linéaire, et lorsque  $f$  est de plus fortement convexe, un taux de convergence linéaire. Bien que ce dernier taux de convergence soit intéressant, on remarque cependant que pour déterminer l'itéré  $\mathbf{x}_{k+1}$  dans (1.3.29), il est nécessaire de calculer le gradient entier de  $f$  donné par  $\nabla f = (1/n) \sum_{i=1}^n \nabla f_i$ , constitué des  $n$  gradients  $\nabla f_i$ . Cela conduit alors à effectuer beaucoup d'opération et de calculs à chaque itération, et de plus cette complexité s'accroît linéairement à mesure que  $n$  augmente. Ce schéma se révèle alors être mal adapté au problème (1.3.30). Une méthode devenue classique permettant de surmonter cette difficulté, tout en gardant la forme générale et la simplicité du schéma (1.3.29), consiste à remplacer le gradient  $\nabla f$  par un estimateur sans biais  $\nabla_k$  de celui-ci, de telle sorte que  $\mathbb{E}(\nabla_k) = \nabla f(\mathbf{x}_k)$ . Il s'agit là de l'idée générale des méthodes de gradient dites « *stochastiques* ».

Les méthodes de type gradient stochastiques ne sont pas récentes et remontent aux années 1950 [64, 96], mais elles se sont beaucoup développées ces dernières années compte tenu du développement du machine-learning et de la simplicité de leur implémentation [27, 28, 102].

Dans sa version la plus simple, l'algorithme de gradient stochastique consiste à effectuer à chaque itération

$$(1.3.30) \quad \mathbf{SG} : \mathbf{x}_{k+1} \leftarrow \mathbf{x}_k - \lambda_k \nabla_k, \quad \text{avec} \quad \nabla_k := \nabla f_{i_k},$$

où  $(\lambda_k)_{k \in \mathbb{N}} \subset \mathbb{R}^+$  désigne la longueur du pas (appelé aussi taux d'apprentissage) et où  $i_k$  est un indice choisit au hasard de façon uniforme dans  $\{1, \dots, n\}$  à chaque itération  $k$ . Le principal intérêt du schéma (1.3.30) réside dans le fait que l'estimateur  $\nabla_k$ , jouant le rôle du gradient, ne dépend plus de la taille  $n$  de l'échantillon considéré. Ce schéma est donc adapté aux problèmes impliquant un très grand nombre d'observations  $f_i$ . Néanmoins cette méthode comporte certains défauts comme la lenteur de la convergence, car l'estimateur  $\nabla_k := \nabla f_{i_k}$  possède une variance qui ne disparaît pas au fil des itérations. En effet, même en partant d'une solution optimale  $\mathbf{x}^*$ , il n'y a aucune raison que  $\nabla_k(\mathbf{x}^*) := \nabla f_{i_k}(\mathbf{x}^*) = 0$ . Afin de réduire cette variance, on choisit habituellement la suite  $(\lambda_k)_{k \in \mathbb{N}} \subset \mathbb{R}^+$  de (1.3.30) comme étant décroissante (en général de la forme  $\lambda_k = Ck^{-\alpha}$  avec  $C > 0$  et  $\alpha \in ]0, 1[$ ), mais cela entraîne un plus grand nombre d'itérations pour atteindre une solution optimale (voir [92, 93, 100]).

En effet, le taux de convergence de la méthode de descente stochastique (1.3.30) pour la résolution du problème (1.3.28), que sous certaines conditions (voir [80]) est

$$(1.3.31) \quad \mathbb{E}(f(\mathbf{x}_k) - f(\mathbf{x}^*)) = \mathcal{O}(1/\sqrt{k}),$$

dans le cas où la fonction  $f$  de (1.3.28) est convexe et est

$$(1.3.32) \quad \mathbb{E}(f(\mathbf{x}_k) - f(\mathbf{x}^*)) = \mathcal{O}(1/k),$$

lorsque  $f$  fortement convexe (voir [80]) où  $k$  désigne le nombre d'itération.

Afin d'améliorer la convergence du schéma général (1.3.30), diverses techniques ont été développées. Nous passons en revue certaines d'entre elles qui sont devenues par la suite des piliers majeurs pour le développement de nouvelles méthodes de type stochastiques.

### 1.3.3 Quelques techniques de réduction de variance

#### 1.3.3.1 Mini-lots (minibatch)

Comme mentionné précédemment, la lenteur de la convergence du schéma général de la méthode de descente de gradient stochastique (1.3.30) provient en premier lieu du terme  $\nabla_k$ . Ce dernier n'étant qu'un estimateur du gradient de  $f$ , il possède une certaine variance qui ne disparaît pas à mesure que le nombre d'itérations grandit. Une des idées les plus simple pour construire un estimateur  $\nabla_k$  plus précis consiste à utiliser non pas un seul échantillon  $f_i$  pour l'estimateur du gradient dans (1.3.30) mais une sélection de  $f_i$  où  $i \in I_k = [i_1, i_2, \dots, i_c]$  avec  $\text{card}(I_k) = c \in \mathbb{N}^*$ , autrement de prendre  $\nabla_k$  de la façon suivante :

$$(1.3.33) \quad \nabla_k := \frac{1}{c} \sum_{j=1}^c f_{i_j}.$$

Cette méthode permet d'obtenir un meilleur estimateur du gradient, cependant celle-ci s'avère est plus coûteuse en terme de calculs comparée à (1.3.30).

#### 1.3.3.2 Un estimateur plus précis

Plutôt que d'utiliser une suite  $(\lambda_k)_{k \in \mathbb{N}}$  décroissante afin de diminuer la variance, il s'avère plus judicieux d'avoir un meilleur estimateur du gradient  $\nabla f$ , i.e, remplacer  $\nabla_k := \nabla f_{i_k}$  dans (1.3.30) par un nouvel estimateur de telle sorte que la variance de ce nouvel estimateur diminue (en moyenne) au fil des itérations. D'abord Johnson et Zhang [63] pour le cas d'une seule fonction  $f$  puis Xiao et Zhang [109] pour la version proximale (correspondant au problème (1.3.27)) proposèrent de choisir l'estimateur

$$(1.3.34) \quad \nabla_k := \nabla f_{i_k}(\mathbf{x}_k) - \nabla f_{i_k}(\mathbf{x}^s) + \nabla f(\mathbf{x}^s),$$

où  $i_k$  est choisi au hasard et de manière uniforme dans  $\{1, \dots, n\}$  et où  $\mathbf{x}^s$  est un itéré que l'on garde en mémoire et que l'on met à jour toutes les  $m$  (avec  $m \geq n$ ) itérations.

On constate alors que les tentatives de réduction de la variance pour les algorithmes de type stochastiques induisent une quantité de travail et un coût supplémentaire.

## 1.3.4 Quelques techniques d'accélération de convergence

### 1.3.4.1 Utilisation de pas constants

Si l'on fait le choix d'utiliser des pas  $(\lambda_k)$  constant au lieu de pas décroissant, alors la convergence de la méthode stochastique (1.3.30) peut se décomposer en deux cas (voir [67] et [79, Proposition 2.4]). Le premier cas va dépendre de  $k$  et converger linéairement vers 0 tandis que le deuxième cas sera indépendant de  $k$  mais ne convergera pas vers 0. Ainsi, avec un pas  $\lambda_k = \lambda$  constant, l'itération (1.3.30) aura une convergence linéaire jusqu'à un certain point pour lequel les valeurs oscilleront et en général les itérations n'iront pas plus loin, car, la convergence de la méthode (1.3.30) avec un pas constant n'a été démontrée que sous des hypothèses très fortes sur les fonctions  $f_i$  (voir [104]).

### 1.3.4.2 Ajout d'un terme inertiel

En plus des divers choix de l'estimateur  $\nabla_k$  du gradient pour réduire la variance et de l'utilisation de pas constants, il est possible de recourir à des techniques d'accélération de type déterministe afin d'accélérer la convergence des itérés. Citons par exemple les algorithmes développés par Nitanda [85] et Allen-Zhu [1] qui combinent un terme inertiel de type Nesterov (voir [23, 24, 81]) avec les méthodes de réduction de variance énoncé dans les section précédentes.

## 1.4 Cas général des inclusions monotones

Considérons le problème d'inclusion générale (1.1.1)

$$(1.4.35) \quad \text{trouver } x^* \in \mathcal{H} \text{ tel que } 0 \in Ax^* + Bx^*,$$

où  $A$  et  $B$  sont des opérateurs  $A$  et  $B$  vérifiant les conditions suivantes

(C1)  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  est un opérateur maximal monotone,

(C2)  $B : \mathcal{H} \rightarrow \mathcal{H}$  est un opérateur maximal monotone et  $1/\beta$ -cocoercif,

(C3)  $\text{zer}(A + B) \neq \emptyset$ .

### 1.4.1 La méthode du point proximal

Dans le cas particulier où  $B \equiv 0$ , le problème (1.4.35) devient simplement

$$(1.4.36) \quad \text{trouver } x^* \in \mathcal{H} \text{ tel que } 0 \in Ax^*.$$

Une des méthodes parmi les plus couramment utilisées pour résoudre (1.4.36) est la méthode du point proximal qui consiste à effectuer pour  $x_0 \in \mathcal{H}$  et  $(\lambda_k) \subset ]0, +\infty[$  l'algorithme suivant

$$(1.4.37) \quad \mathbf{PPA} : \quad \mathbf{x}_{k+1} = J_{\lambda_k}^A(\mathbf{x}_k),$$

où  $J_{\lambda_k}^A$  est l'opérateur résolvante de  $A$  d'indice  $\lambda_k$ . Sous l'hypothèse  $\text{zer}(A) \neq \emptyset$  ainsi que sous certaines conditions sur la suite  $(\lambda_k)$ , Rockafellar [97] et Brézis-Lions [32] ont prouvé que la suite  $(\mathbf{x}_k)$  générée par (1.4.37) converge faiblement vers un élément  $\mathbf{x}^* \in \text{zer}(A)$ . Güler ([58, Corollary 5.1]), quant à lui donna un contre-exemple pour lequel la suite  $(\mathbf{x}_k)$  converge faiblement, mais pas fortement. Notons que lorsque  $A$  est supposé fortement monotone, alors la suite  $(\mathbf{x}_k)$  générée par la méthode **PPA** converge fortement (et linéairement) vers l'unique élément  $\text{zer}(A)$  (voir [22, Exemple 23.40]).

Observons que le schéma (1.4.37) est une généralisation du schéma (1.2.17) dans le cas où  $A$  est un opérateur maximal monotone quelconque et que la méthode (1.4.37) provient d'une discrétisation implicite en temps du système

$$(1.4.38) \quad \dot{x}(t) + Ax(t) \ni 0,$$

qui correspond au cas général de (1.2.10). La méthode (1.4.37) a, par la suite, fait l'objet de nombreuses variantes relaxées et inertielles (voir [3, 8, 12, 13, 46]) dans le but d'obtenir des taux de convergence similaires à ceux atteints dans le cadre de la minimisation convexe.

## 1.4.2 Méthodes d'éclatement

Lorsque  $A + B$  est maximal monotone dans (1.4.35), l'approche directe consistant à effectuer la méthode du point proximal à l'opérateur  $A + B$  n'est envisageable que dans le cas où la résolvante de l'opérateur  $A + B$  est simple à calculer, ce qui est rarement le cas en pratique. Ainsi, une astuce devenue classique par la suite est d'« éclater » le problème (1.4.35) en utilisant les résolvantes de  $A$  et de  $B$ .

### Algorithme de Douglas-Rachford

Un des algorithmes d'éclatement les plus connus et ne nécessitant pas d'hypothèses fortes sur les opérateurs  $A$  et  $B$  est la méthode Douglas-Rachford (D-R en abrégé) (voir, par exemple, [69]) dont la forme générale est donnée pour  $\mathbf{x}_0 \in \mathcal{H}$  par

$$(1.4.39) \quad \mathbf{D-R} : \quad \mathbf{x}_{k+1} = (J_\lambda^A(2J_\lambda^B - I) + (I - J_\lambda^B))(\mathbf{x}_k).$$

Sous la condition  $\text{zer}(A + B) \neq \emptyset$ , l'algorithme de D-R génère des suites  $(\mathbf{x}_k)$  qui convergent faiblement vers un élément  $x^* \in \mathcal{H}$  vérifiant  $J_\lambda^B(x^*) \in \text{zer}(A + B)$  (voir [22, Theorem 26.11]). Cependant, bien que la méthode (1.4.39) ne nécessite pas d'hypothèses particulières sur les opérateurs  $A$  et  $B$ , celle-ci requiert néanmoins le calcul des résolvantes de  $A$  et de  $B$ , ce qui peut s'avérer coûteux en pratique, voire restrictif lorsque la résolvante d'un des opérateurs est difficile à calculer.

### Algorithme Forward-Backward

Lorsque l'opérateur  $B$  est univoque, une autre approche pour résoudre (1.1.1) consiste à effectuer l'algorithme « Forward-Backward » [69, 89]

$$(1.4.40) \quad \mathbf{FB} : \quad \mathbf{x}_{k+1} = J_{\lambda_k}^A(\mathbf{x}_k - \lambda_k B\mathbf{x}_k).$$

La méthode FB est une généralisation de la méthode de gradient proximal (1.2.20) dans le cas où les sous-différentiels des fonctions sont remplacés par des opérateurs

maximaux monotones plus généraux. Notons que lorsque  $B \equiv 0$ , la méthode FB se réduit la méthode PPA. Sous les hypothèses **(C1)**-**(C3)** ainsi que pour  $\lambda \in ]0, 2\beta[$ , il a été démontré (voir par exemple [22, Theorem 25.8]) que la méthode FB génère des suites  $(\mathbf{x}_k)$  faiblement convergentes. En particulier, lorsque  $A$  est  $\alpha$ -fortement monotone ou lorsque  $B$  est à la fois fortement monotone et cocoercif, les itérés générés par le schéma (1.4.40) convergent linéairement vers un élément de  $\text{zer}(A + B)$  (voir [22, Proposition 25.9]).

### Algorithme Forward-Backward-Forward

Afin d'assouplir l'hypothèse de cocoercivité requise par l'algorithme FB, Tseng [106] proposa la méthode suivante connue sous le nom de « Forward-Backward-Forward »

$$(1.4.41) \quad \mathbf{FBF} : \begin{cases} \mathbf{y}_k &= J_\lambda^A(\mathbf{x}_k - \lambda B\mathbf{x}_k), \\ \mathbf{x}_{k+1} &= \mathbf{y}_k + \lambda B\mathbf{x}_k - \lambda B\mathbf{y}_k. \end{cases}$$

Sous la condition  $\text{zer}(A + B) \neq \emptyset$  et moyennant quelques évaluations supplémentaires de l'opérateur  $B$ , ce schéma produit des suites  $(\mathbf{x}_k)$  et  $(\mathbf{y}_k)$  qui convergent faiblement vers un élément de  $\text{zer}(A + B)$  et requiert seulement que  $B$  soit lipschitzien (voir [22, Theorem 26.17]).



# Chapitre 2

## Un modèle général du premier ordre pour des dynamiques continues de type Newton gouvernées par des opérateurs maximaux monotones

---

### 2.1 Résumé et principaux résultats

Soit  $\mathcal{H}$  un espace de Hilbert réel pour lequel nous notons  $\langle \cdot, \cdot \rangle$  le produit scalaire et  $\| \cdot \|$  la norme associée. L'objectif de ce chapitre est de proposer un système dynamique dissipatif de type Newton dans le but d'offrir un cadre adapté à la conception de nouvelles méthodes d'approximation, à la fois performante et peu coûteuse en calcul, concernant deux problèmes d'optimisation particuliers : la minimisation d'une fonction convexe  $\Phi : \mathcal{H} \rightarrow \mathbb{R}$  et le calcul d'un zéro d'un opérateur maximal monotone  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ .

La plupart des modèles concernant les méthodes accélérées se basent sur une dynamique de type Newton du second ordre faisant intervenir la Hessienne d'une fonction objectif ou la dérivée d'un opérateur dépendant du temps. Nous proposons un système dynamique ne faisant pas intervenir ces derniers termes, ceci dans le but de concevoir des algorithmes (via une discrétisation), sans termes encombrants ni hypothèses restrictives sur les opérateurs.

Dans un premier temps, nous supposons  $\Phi$  de classe  $C^2$  et nous considérons l'équation du second ordre suivante

$$(2.1.1) \quad \ddot{x}(t) + \tau(t)\dot{x}(t) + \beta \frac{d}{dt}(\nabla\Phi(x(t)) + \varepsilon A_{\lambda(t)}x(t)) + \eta(t)(\nabla\Phi(x(t)) + \varepsilon A_{\lambda(t)}x(t)) = 0,$$

où  $\nabla\Phi$  désigne le gradient de  $\Phi$ ,  $A_\lambda$  est l'approximation de Yosida de  $A$  d'indice  $\lambda > 0$ , et où les paramètres sont donnés par :

$$(2.1.2) \quad \varepsilon \geq 0, \quad \beta = 1, \quad \tau(\cdot) = \kappa - \theta - \frac{\dot{\theta}}{\theta} \quad \text{et} \quad \eta(\cdot) = \kappa - \frac{\dot{\theta}}{\theta},$$

où  $\kappa$  et  $\theta(\cdot)$  désignent respectivement une constante et une fonction réelle strictement positives. Plus précisément,  $\theta$  sera choisi de la forme

$$(2.1.3) \quad \theta(t) = \frac{\kappa\nu(t) - \dot{\nu}(t)}{\nu(t) + e},$$

où  $e$  est une constante strictement positive et  $\nu$  une fonction réelle strictement positive ( $\dot{\nu}$  désigne la dérivée de  $\nu$ ).

Nous montrons, via la proposition suivante, que (2.1.1)-(2.1.2)-(2.1.3) est équivalent à un système dynamique du premier ordre.

**Proposition 2.1.1.** *Supposons que  $\Phi : \mathcal{H} \rightarrow \mathbb{R}$  soit deux fois continûment différentiable et que  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  soit un opérateur maximal monotone. Soient  $\kappa > 0$  ainsi que  $\theta$  et  $\lambda$  deux fonctions strictement positives de classe  $C^1$  sur  $[0, +\infty[$ . Supposons qu'il existe  $\lambda_m > 0$  tel que  $\lambda(t) \geq \lambda_m > 0$  pour  $t \geq 0$ . Alors, pour tout  $(x_0, \dot{x}_0) \in \mathcal{H}^2$ , les assertions (i1) et (i2) suivantes sont équivalentes :*

(i1)  $x : [0, +\infty[ \rightarrow \mathcal{H}$  est une solution classique de (2.1.1) avec  $\dot{x}(0) = \dot{x}_0, x(0) = x_0$ .

(i2)  $(x, y) : [0, +\infty[ \rightarrow \mathcal{H}^2$ , pour un certain  $y \in C^1([0, +\infty[: \mathcal{H})$ , est une solution de classe  $C^2 \times C^1$  du système

$$(2.1.4) \quad \begin{cases} \dot{x} + \nabla\Phi(x) + \varepsilon A_{\lambda(\cdot)}x + \theta(\cdot)(y - x) = 0, \\ \dot{y} + \kappa(y - x) = 0, \end{cases}$$

avec  $x(0) = x_0$  et  $y(0) = -\frac{1}{\theta(0)}(\dot{x}_0 + \nabla\Phi(x_0) + \varepsilon A_{\lambda(0)}x_0) + x_0$ .

La proposition précédente permet de nous intéresser exclusivement au système dynamique (2.1.4), qui non seulement possède toujours du sens lorsque  $\Phi$  n'est que de classe  $C^1$  mais s'avère aussi plus simple à manipuler.

Par la suite, nous faisons les hypothèses suivantes :

- (H1)  $\Phi : \mathcal{H} \rightarrow \mathbb{R}$  est une fonction convexe différentiable et son gradient  $\nabla\Phi$  est lipschitzien sur les sous-ensembles bornés de  $\mathcal{H}$ ,
- (H2)  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  est un opérateur maximal monotone,
- (H3)  $S := (\epsilon A)^{-1}(0) \cap (\nabla\Phi)^{-1}(0) \neq \emptyset$ .

Puis, nous démontrons l'existence et l'unicité des solutions de (2.1.4) via la proposition suivante.

**Proposition 2.1.2.** *Soit  $\Phi : \mathcal{H} \rightarrow \mathbb{R}$  une fonction et  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  un opérateur vérifiant les conditions (H1)-(H3). Soit  $\kappa > 0$  et considérons des applications réelles  $\lambda(\cdot)$  et  $\theta$  vérifiant :*

$$(2.1.5) \quad \lambda \in C^0([0, +\infty[) \text{ et, pour } t \geq 0, \lambda(t) \in [\lambda_m, +\infty[ ,$$

$$(2.1.6) \quad \theta \in C^0([0, +\infty[) \text{ et, pour } t \geq 0, \theta(t) \in [0, \theta_M],$$

pour certaines constantes strictement positives  $\lambda_m$  et  $\theta_M$ . Alors pour toutes données initiales  $(x_0, y_0) \in \mathcal{H}^2$ , il existe une unique solution classique globale  $(x, y)$  du système (2.1.4) vérifiant  $(x(0), y(0)) = (x_0, y_0)$ .

Dans le but d'obtenir des résultats asymptotiques concernant les trajectoires générées par le système (2.1.4), les conditions suivantes seront requises :

$$(2.1.7a) \quad \nu(\cdot) \text{ est positive, de classe } C^2 \text{ et croissante sur } [0, +\infty[,$$

$$(2.1.7b) \quad \frac{\dot{\nu}}{\nu} \text{ est décroissante et } \frac{\dot{\nu}}{\nu} < \kappa \text{ sur } [0, +\infty[, \quad (\text{i.e } \rho > 0 \text{ et } \rho \nearrow),$$

$$(2.1.7c) \quad \text{il existe } M > 0 \text{ tel que } \sup_{t \geq 0} |\dot{\nu}(t)| \leq M.$$

Des conditions supplémentaires sur  $\nu$  seront aussi nécessaires afin d'obtenir la convergence faible des trajectoires vers un équilibre :

$$(2.1.8) \quad \nu(t) \rightarrow +\infty, \quad \dot{\nu}(t) \rightarrow l \in [0, +\infty[, \quad \ddot{\nu}(t) \rightarrow 0 \text{ (lorsque } t \rightarrow +\infty),$$

$$(2.1.9) \quad \nu \in C^3([0, +\infty[).$$

De plus, les paramètres  $\lambda(\cdot)$  et  $e$  sont choisis de telle sorte que

$$(2.1.10a) \quad 0 < \frac{2M}{\rho(0)} < e, \text{ où } \rho(0) = \kappa - \frac{\dot{\nu}(0)}{\nu(0)}.$$

$$(2.1.10b) \quad \lambda = \lambda_c \frac{(\nu + e)^2}{\theta}, \text{ pour une constante } \lambda_c > \left(\frac{\epsilon}{4}\right) \inf_{s \in \left(\frac{2M}{\rho(0)}, e\right)} \frac{1}{s(e-s)}.$$

S'agissant du cas  $\epsilon > 0$  (autrement dit lorsque  $A_\lambda$  est pris en compte dans (2.1.4)) nous considérons les hypothèses supplémentaires suivantes :

$$(2.1.11) \quad \epsilon \nu^{(3)} \in L^\infty([0, +\infty[) \quad \text{et} \quad \int_0^{+\infty} \frac{\epsilon}{\nu^3(t)} dt < +\infty.$$

En tenant compte des hypothèses précédentes, nous démontrons le théorème suivant qui concerne le comportement asymptotique des solutions générées par (2.1.4).

**Théorème 2.1.1.** *Soient  $\Phi : \mathcal{H} \rightarrow \mathbb{R}$  et  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  vérifiant **(H1)**-**(H3)**. Supposons que  $(x, y) : [0, +\infty[ \rightarrow \mathcal{H}^2$  soit une solution classique globale de (2.1.4), où  $\{\kappa, e\} \subset ]0, +\infty[$  et où  $\nu(\cdot)$  et  $\lambda(\cdot)$  satisfont respectivement (2.1.7) et (2.1.10). Alors nous avons les propriétés suivantes :*

$$(2.1.12) \quad \int_0^{+\infty} \nu(t) (\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) dt < +\infty,$$

$$(2.1.13) \quad \int_0^{+\infty} \nu(t) \|\dot{x}(t)\|^2 dt < +\infty, \quad \int_0^{+\infty} \nu(t) \|y(t) - x(t)\|^2 dt < +\infty,$$

$$(2.1.14) \quad \int_0^{+\infty} \epsilon \lambda(t) \nu(t) \|A_{\lambda(t)} x(t)\|^2 dt < +\infty, \quad \int_0^{+\infty} \nu^2(t) \|\nabla \Phi(x(t))\|^2 dt < +\infty,$$

$$(2.1.15) \quad \Phi(x(t)) - \inf_{\mathcal{H}} \Phi = o(\nu^{-2}(t)),$$

$$(2.1.16) \quad \|y(t) - x(t)\| = o(\nu^{-1}(t)), \quad \|\dot{x}(t) + \nabla \Phi(x(t)) + \epsilon A_{\lambda(t)} x(t)\| = o(\nu^{-1}(t));$$

*Si, de plus,  $\nu(\cdot)$  satisfait (2.1.8)-(2.1.11), alors :*

$$(2.1.17) \quad \|\dot{x}(t)\| = o(\nu^{-1}(t)),$$

$$(2.1.18) \quad \lim_{t \rightarrow +\infty} \epsilon \lambda(t) \|A_{\lambda(t)} x(t)\| = 0,$$

$$(2.1.19) \quad \exists \bar{x} \in S \quad t.q \quad x(\cdot) \rightharpoonup \bar{x} \text{ faiblement dans } \mathcal{H}.$$

Dans le cas particulier où  $\nu(t) = Mt + \nu_0$  (pour certaines valeurs strictement positives  $M$  et  $\nu_0$ ), nous avons le résultat suivant.

**Théorème 2.1.2.** Soient  $\Phi$  et  $A$  vérifiant **(H1)**-**(H3)**, soient  $\{\kappa, e, \nu_0, M\} \subset ]0, +\infty[$  tels que  $\nu_0 > M\kappa^{-1}$  et  $e > \frac{2M\nu_0}{\kappa\nu_0 - M}$ . Supposons que  $(x, y) : [0, +\infty[ \rightarrow \mathcal{H}^2$  est une solution classique globale de (2.1.4), avec  $\theta(t) = \frac{M\kappa t + (\kappa\nu_0 - M)}{Mt + \nu_0 + e}$  et  $\lambda(t) = \lambda_c \frac{(Mt + \nu_0 + e)^3}{(M\kappa t + (\kappa\nu_0 - M))}$  pour une certaine constante  $\lambda_c > \inf_{0 < s < \frac{2M\nu_0}{\kappa\nu_0 - M}} \frac{1}{s(e-s)}$ . Alors il existe  $\bar{x} \in S$  tel que  $x(\cdot) \rightharpoonup \bar{x}$  faiblement dans  $\mathcal{H}$ , et nous avons

$$(2.1.20) \quad \int_0^{+\infty} t(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) dt < +\infty, \quad \Phi(x(t)) - \inf_{\mathcal{H}} \Phi = o(t^{-2}),$$

$$(2.1.21) \quad \int_0^{+\infty} t\|\dot{x}(t)\|^2 dt < +\infty, \quad \|\dot{x}(t)\| = o(t^{-1}),$$

$$(2.1.22) \quad \int_0^{+\infty} t\|y(t) - x(t)\|^2 dt < +\infty, \quad \|y(t) - x(t)\| = o(t^{-1}),$$

$$(2.1.23) \quad \int_0^{+\infty} \frac{\epsilon}{t} (\lambda(t) \|A_{\lambda(t)}x(t)\|)^2 dt < +\infty,$$

$$(2.1.24) \quad \int_0^{+\infty} t^2 \|\nabla\Phi(x(t))\|^2 dt < +\infty,$$

$$(2.1.25) \quad \|\nabla\Phi(x(t)) + \epsilon A_{\lambda(t)}x(t)\| = o(t^{-1}),$$

$$(2.1.26) \quad \lim_{t \rightarrow +\infty} \epsilon \lambda(t) \|A_{\lambda(t)}x(t)\| = 0.$$

Enfin, en effectuant une discrétisation en temps pour  $h > 0$ , nous suggérons deux algorithmes selon les cas suivants :

– Dans le cas où  $A = 0$  et  $\Phi := \Theta_{sm} + \Theta_{ns}$  avec  $\Theta_{sm} : \mathcal{H} \rightarrow \mathbb{R}$  une fonction convexe différentiable et  $\Theta_{ns} : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  une fonction convexe et semi-continue inférieurement, on obtient (pour  $h > 0$ )

$$(2.1.27a) \quad z_k = x_k - h\theta_k(y_k - x_k),$$

$$(2.1.27b) \quad x_{k+1} = J_h^{\partial(\Theta_{ns})}(z_k - h\nabla(\Theta_{sm})(x_k)),$$

$$(2.1.27c) \quad y_{k+1} = y_k - h\kappa(y_k - x_k).$$

– Dans le cas où  $\Phi = 0$  et  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  est un opérateur maximal monotone, on obtient le schéma (pour  $h > 0$ )

$$(2.1.28a) \quad z_{k-1} = y_{k-1} + \frac{1+h\theta_{k-1}}{\kappa h}(y_k - y_{k-1}),$$

$$(2.1.28a) \quad x_k = \frac{\lambda_{k-1}}{\lambda_{k-1}+h}z_{k-1} + \frac{h}{\lambda_{k-1}+h}J_{\lambda_{k-1}+h}^A(z_{k-1}),$$

$$(2.1.28a) \quad y_{k+1} = (1 - h\kappa)y_k + (h\kappa)x_k.$$

Notons qu'une seule évaluation du gradient est nécessaire pour effectuer l'algorithme (2.1.27) et une seule évaluation proximale de  $A$  est requise pour (2.1.28).

## 2.2 Article en anglais

### FIRST-ORDER FRAMEWORKS FOR CONTINUOUS NEWTON-LIKE DYNAMICS GOVERNED BY MAXIMALLY MONOTONE OPERATORS<sup>1</sup>

#### Abstract

In a Hilbert framework, we discuss a continuous Newton-like model that is well-adapted in view to numerical purposes for solving convex minimization and more general monotone inclusion problems. Algorithmic solutions to these problems were recently inspired by implicit temporal discretizations of the (stabilized) continuous version of Nesterov's accelerated gradient method with an additional Hessian damping term (so as to attenuate the oscillation effects). Unfortunately, due to the presence of the Hessian term, these discrete variants require several gradients or proximal evaluations (per iteration). An alternative methodology can be realized by means of a first-order model that no more involves the Hessian term and that can be extended to the case of an arbitrary maximally monotone operator  $A$ . Our first-order model originates from the reformulation of a closely related variant to the Nesterov-like equation. Its dynamics are studied (simultaneously) with regard to convex minimization and monotone inclusions by considering it when governed by the sum of the gradient of a convex differentiable function and (up to a multiplicative constant) the Yosida approximation of a maximally monotone operator, with an appropriate adjustment of the regularization parameter. It turns out that our model, offers a new framework for discrete variants, while keeping the main asymptotic features of the (stabilized) Nesterov-like equation. Two new algorithms are then suggested relative to the considered optimization problems.

**Key words :** Asymptotic behavior · first-order differential equation · dissipative systems · maximal monotone operators · potential and non-potential operators · cocoercive operators · Yosida approximation · coupled systems · Nesterov acceleration · Newton-like convergence

---

1. Labarre, F., Maingé, PE. First-Order Frameworks for Continuous Newton-like Dynamics Governed by Maximally Monotone Operators. Set-Valued Var. Anal (2021). <https://doi.org/10.1007/s11228-021-00593-1>

---

**Mathematics Subject Classification (2010)** MSC 37N40 · MSC 46N10 · MSC 49M30 · 26 MSC 65K05 · MSC 65K10 · MSC 90B50 · MSC 90C25

### 2.2.1 Introduction

Let  $\mathcal{H}$  be a real Hilbert space with inner product and induced norm denoted by  $\langle \cdot, \cdot \rangle$  and  $\| \cdot \|$ , respectively. The purpose of this paper is to propose fast continuous Newton-like dynamics that are intended to offer a well-adapted framework for the design of new performing (reliable and computationally cheap) approximation methods for two particular problems : the minimization of a convex function  $\Phi : \mathcal{H} \rightarrow \mathbb{R}$  and the computation of a zero of an arbitrary maximally monotone operator  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ . These are issues of great importance that encompass many real-world problems arising in equilibrium theory and game theory (in the economic field) but also in partial differential equations to name but a few [25, 33, 50, 95] (also see [26, 48, 98, 105] for related topics).

The above methodology is commonly used in optimization, by considering appropriate discrete counterparts of evolution equations whose trajectories are anchored to the solution set of some considered problem. Nonetheless, such a continuous model is required to have a nice form so as to avoid the presence of some constraining terms that can lead to numerical methods that are computationally expensive or of limited effectiveness in applications. This applies even more now that most of the pertinent models (dedicated to the development of speed up optimization processes) rely upon second-order Newton-like dynamics that involves the Hessian of an objective function (see, e.g., Attouch-Chbani-Fadili-Riahi [18], Shi-Du-Jordan-Su [103], for convex minimization) or the derivative of a time-dependent operator (see, e.g., Attouch-László [12, 11], for monotone inclusions).

Specifically, we propose simultaneously (from a mathematical viewpoint), for both the considered problems, a continuous framework with a view to designing accelerated variants of (existing) related algorithms, without any embarrassing terms or restrictive assumptions on the involved operators. Two simple algorithms are then suggested from this setting. In particular, only one proximal evaluation of  $A$ , for approximating its zeroes, will be required per iteration compared with the discrete model proposed in [12] with the same acceleration purposes. A similar advantage is also obtained relative to convex differentiable minimization problems. Only one evaluation of the gradient of  $\Phi$  will be required per iteration in comparison with the

algorithms developed in [18, 103] which involve the difference of gradients of  $\Phi$  at two successive iterates (as an approximation of the Hessian of  $\Phi$ ). Below we describe our strategy.

### 2.2.1.1 A second-order system

Our strategy consists first of considering, when  $\Phi$  is  $C^2$  and for some  $\beta > 0$  and some real-valued functions  $\tau(\cdot)$  and  $\eta(\cdot)$ , the following second-order equation :

$$(2.2.1) \quad \ddot{x}(t) + \tau(t)\dot{x}(t) + \beta \frac{d}{dt} \Psi(t, x(t)) + \eta(t)\Psi(t, x(t)) = 0,$$

which is driven by the time-dependent operator  $\Psi(t, \cdot) = \nabla\Phi(\cdot) + \epsilon A_{\lambda(t)}$ , where  $\epsilon \geq 0$ ,  $\lambda(\cdot)$  is a positive mapping, and  $A_\lambda := \frac{1}{\lambda}(I - (I + \lambda A)^{-1})$  is the Yosida approximation of  $A$  (with index  $\lambda$ ). This model allows us, among others, to cover simultaneously the continuous approaches to minimization problems and monotone inclusion problems. Note that important instances of (2.2.1, in which  $\tau(t) = \alpha t^{-1}$  (for some  $\alpha > 0$ ) and  $\eta(t) = 1$ , were recently considered in the literature. Note indeed that, with these parameters  $\tau$  and  $\eta$ , when taking  $\epsilon A = 0$  (namely  $\epsilon A_\lambda = 0$ ), equation (2.2.1) reduces to

$$(2.2.2) \quad \ddot{x}(t) + \alpha t^{-1}\dot{x}(t) + \beta \nabla^2\Phi(x(t))\dot{x}(t) + \nabla\Phi(x(t)) = 0,$$

while for  $\Phi = 0$  and  $\epsilon = 1$  it becomes

$$(2.2.3) \quad \ddot{x}(t) + \alpha t^{-1}\dot{x}(t) + \beta \frac{d}{dt} (A_{\lambda(t)}x(t)) + A_{\lambda(t)}x(t) = 0.$$

Some reminders are given relative to (2.2.2) and (2.2.3) :

- The model (2.2.2), which was investigated by Attouch-Peypouquet-Redont [15], can be regarded as an over-damped dynamical system that involves two types of damping : a geometrical damping and an isotropic linear damping with a vanishing viscous parameter. Its trajectories were proved to be weakly convergent, along with the convergence rates  $\|\dot{x}(t)\| = o(t^{-1})$  (for the velocity),  $\Phi(x(t)) - \inf_{\mathcal{H}} \Phi = o(t^{-2})$  (for the function values), and a property of fast decay of the gradient illustrated by  $\int_{t_0}^{+\infty} t^2 \|\nabla\Phi(x(t))\|^2 dt < \infty$  (for  $t_0 > 0$ ). Note that the algorithms discussed in [18] were mainly inspired from temporal implicit discretizations of (2.2.2) and other second-order related models.

- The model (2.2.3) has been recently studied by Attouch-László [11]. Its trajectories, under the conditions  $\alpha > 1$ ,  $\beta \geq 0$  and  $\lambda(t) = \bar{\lambda}t^2$  with  $\bar{\lambda} > (\alpha - 1)^{-2}$ , were shown to be weakly convergent, along with the rates  $\|\dot{x}(t)\| = o(t^{-1})$ ,  $\|A_{\lambda(t)}x(t)\| = o(t^{-2})$ , and the integral estimates

$$\int_{t_0}^{+\infty} t \|\dot{x}(t)\|^2 dt < \infty, \quad \int_{t_0}^{+\infty} t^{-1} (\lambda(t) \|A_{\lambda(t)}x(t)\|)^2 dt < \infty \quad (\text{for } t_0 > 0).$$

Let us mention that the numerical process considered by Attouch-László [12] comes from a temporal implicit discretization of (2.2.3).

It would be interesting to obtain similar results that in [11, 15] by using different parameters  $\tau$  and  $\eta$  that primarily lead to a better suited framework for discrete models.

### 2.2.1.2 An equivalent first-order system

The sequel of our strategy consists of using the appropriate set of parameters :

$$(2.2.4) \quad \beta = 1, \quad \tau(\cdot) = \kappa - \theta - \frac{\dot{\theta}}{\theta} \quad \text{and} \quad \eta(\cdot) = \kappa - \frac{\dot{\theta}}{\theta},$$

for some  $\kappa > 0$  and some positive mapping  $\theta(\cdot)$ . Let us discuss the motivations for these parameters :

- On the one hand, we show (see Proposition 2.2.1) that (2.2.1) can be rewritten as the first-order differential system

$$(2.2.5a) \quad \dot{x}(t) + [\nabla\Phi(x(t)) + \epsilon A_{\lambda(t)}x(t)] + \theta(t)(y(t) - x(t)) = 0,$$

$$(2.2.5b) \quad \dot{y}(t) + \kappa(y(t) - x(t)) = 0,$$

which still makes sense when  $\Phi$  is only  $C^1$  and which can be easily generalized to the non-smooth case of  $\Phi$ .

- On the other hand, via (2.2.5) we can obtain closely related models to (2.2.2) and (2.2.3) when choosing the specific parameter

$$(2.2.6) \quad \theta(t) = \frac{\kappa\nu(t) - \dot{\nu}(t)}{\nu(t) + e},$$

along with some constant  $e > 0$  and some positive and differentiable mapping  $\nu$ . Indeed, taking (for instance)  $\nu(t) = Mt + \nu_0$  with  $M > 0$  and  $\nu_0 > 0$ , we obtain (see

Remark 2.2.3) that the second-order system (2.2.1) holds with parameters  $\tau$  and  $\eta$  such that  $\tau(t) \sim \frac{(1+\kappa\epsilon)M^{-1}}{t}$  and  $\eta(t) \sim \kappa$  (as  $t \rightarrow \infty$ ). This observation suggests that (2.2.1) with parameters given by (2.2.4) and (2.2.6) (together with an appropriate choice of  $\nu$ ) asymptotically behaves like (2.2.2) and (2.2.3), while its equivalent first-order formulation (2.2.5) is highly suitable for numerical purposes.

### 2.2.1.3 An overview of the main results

The evolution system (2.2.5)-(2.2.6) is well-posed (see Proposition 2.2.2) and it will be considered under the condition  $S := (\nabla\Phi)^{-1}(0) \cap (\epsilon A)^{-1}(0) \neq \emptyset$ . This way, based on the first-order formulation (2.2.5), we will develop a technical analysis so as to show that its trajectories converge weakly to elements of  $S$ . In specific, by the study of (2.2.5)-(2.2.6) in the potential case (which corresponds to  $\epsilon = 0$ , or  $A = 0$ ), we cover the convergence analysis of the system (relative to convex minimization)

$$(2.2.7) \quad \dot{x}(t) + \nabla\Phi(x(t)) + \theta(t)(y(t) - x(t)) = 0, \quad \dot{y}(t) + \kappa(y(t) - x(t)) = 0,$$

while (2.2.5)-(2.2.6) in the non-potential case (which corresponds to  $\epsilon > 0$  and  $\Phi = 0$ ) allows us to cover the convergence analysis of the following system (relative to monotone inclusions)

$$(2.2.8) \quad \dot{x}(t) + A_{\lambda(t)}x(t) + \theta(t)(y(t) - x(t)) = 0, \quad \dot{y}(t) + \kappa(y(t) - x(t)) = 0.$$

At first sight, the introduction of the coefficient  $\epsilon$  seems to be optional but it will be crucial in computations, for a better legibility, especially when using the co-coerciveness of  $A_\lambda$  in our estimations. Let us underline that the particular asymptotic properties of the Nesterov-like differential equations (2.2.2) and (2.2.3) are retrieved for systems (2.2.7) and (2.2.8), respectively. Under convenient assumptions on the parameters  $\theta$  and  $\lambda$  we establish (among others) the weak convergence of the trajectories of (2.2.5)-(2.2.6) towards equilibria together with the following properties (see Theorems 2.2.1 and 2.2.2) :

$$(2.2.9a) \quad \|\dot{x}(t)\| = o(t^{-1}), \quad \int_0^{+\infty} t \|\dot{x}(t)\|^2 dt < \infty,$$

$$(2.2.9b) \quad \Phi(x(t)) - \inf_{\mathcal{H}} \Phi = o(t^{-2}),$$

$$(2.2.9c) \quad \int_0^{+\infty} t^2 \|\nabla\Phi(x(t))\|^2 dt < \infty, \quad \int_0^{+\infty} \frac{\epsilon}{t} (\lambda(t) \|A_{\lambda(t)}x(t)\|)^2 dt < \infty,$$

$$(2.2.9d) \quad \epsilon \|A_{\lambda(t)}x(t)\| = o(t^{-2}),$$

where  $\lambda(t) \sim \bar{\lambda}t^2$  as  $t \rightarrow \infty$  (for some constant  $\bar{\lambda}$ ). The proof of our convergence results relies on Lyapunov properties for the functional  $\mathcal{E}_{s,q}$  related to (2.2.6)-(2.2.5) and defined for  $(s, q) \in [0, \infty) \times S$ , by

$$(2.2.10) \quad \mathcal{E}_{s,q} = \frac{1}{2} \|s(q-x) + \nu(y-x)\|^2 + \frac{s(e-s)}{2} \|x-q\|^2 + \frac{\nu(\nu+e)}{\theta} \tilde{\Phi}(x),$$

where  $\tilde{\Phi} = \Phi - \inf_{\mathcal{H}} \Phi$ .

#### 2.2.1.4 Assumptions

In order to establish our main convergence results for (2.2.5)-(2.2.6), we make the following assumptions :

- (H1)  $\Phi : \mathcal{H} \rightarrow \mathbb{R}$  is convex differentiable and its gradient  $\nabla\Phi$  is Lipchitz continuous on the bounded subsets of  $\mathcal{H}$ .
- (H2)  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  is a maximally monotone operator.
- (H3)  $S := (\epsilon A)^{-1}(0) \cap (\nabla\Phi)^{-1}(0) \neq \emptyset$ .

Given some positive parameters  $\kappa$  and  $\nu(\cdot)$  we set  $\rho := \kappa - \frac{\dot{\nu}}{\nu}$  (if  $\nu$  is differentiable) and we assume that the following conditions are satisfied :

- (2.2.11a)  $\nu(\cdot)$  is positive, of class  $C^2$  and non-decreasing on  $[0, \infty)$ ,
- (2.2.11b)  $\frac{\dot{\nu}}{\nu}$  is non-increasing and  $\frac{\dot{\nu}}{\nu} < \kappa$  on  $[0, \infty)$ , (namely  $\rho > 0$  and  $\rho \nearrow$ ),
- (2.2.11c) there exists  $M > 0$  such that  $\sup_{t \geq 0} |\dot{\nu}(t)| \leq M$ .

The parameter  $\lambda(\cdot)$  and the constant  $e$  involved in (2.2.6) are assumed to be such that :

$$(2.2.12a) \quad 0 < \frac{2M}{\rho(0)} < e, \text{ where } \rho(0) = \kappa - \frac{\dot{\nu}(0)}{\nu(0)}.$$

$$(2.2.12b) \quad \lambda = \lambda_c \frac{(\nu+e)^2}{\theta}, \quad \text{with some constant } \lambda_c > \left(\frac{\epsilon}{4}\right) \inf_{s \in \left(\frac{2M}{\rho(0)}, e\right)} \frac{1}{s(e-s)}.$$

The following additional conditions on  $\nu$  will also be needed so as to reach the (weak) convergence (to equilibria) of the trajectories :

$$(2.2.13) \quad \nu(t) \rightarrow +\infty, \quad \dot{\nu}(t) \rightarrow l \in [0, \infty), \quad \ddot{\nu}(t) \rightarrow 0 \text{ (as } t \rightarrow \infty),$$

$$(2.2.14) \quad \nu \in C^3([0, \infty)).$$

Regarding the case  $\epsilon > 0$  (which means that  $A_\lambda$  is taken into account in our problem) we need the supplementary requirements on  $\nu$  :

$$(2.2.15) \quad \epsilon\nu^{(3)} \in L^\infty[0, \infty), \quad \int_0^{+\infty} \frac{\epsilon}{\nu^3(t)} dt < \infty$$

### 2.2.1.5 Organization of the paper

The present paper is organized as follows. Section 2 exposes some specific results regarding our model (equivalence between the first and second order systems, existence and uniqueness of trajectories and basic properties regarding the involved parameters). Section 3 is devoted to the construction of a Lyapunov functional associated with our first order system. Section 4 presents our main convergence results. In Section 5 we suggest algorithms issued from discrete variants of our continuous model for monotone inclusion problems and convex minimization.

## 2.2.2 Specific results on the considered system

### 2.2.2.1 From a second to a first-order system

Let us expose the equivalence between the two systems (2.2.1) and (2.2.5).

**Remark 2.2.1.** *Let  $z \in (\epsilon A)^{-1}(0)$  (from **(H3)**), let  $x : [0, \infty) \rightarrow \mathcal{H}$ , and assume that  $\lambda(t) \geq \lambda_m > 0$  (for  $t \geq 0$ ). Then, by Lemma A.1, for any  $\{t_1, t_2\} \subset [0, \infty)$  we readily obtain*

$$\|\epsilon A_{\lambda(t_1)}x(t) - \epsilon A_{\lambda(t_2)}x(t)\| \leq 3\epsilon \frac{|\lambda(t_1) - \lambda(t_2)|}{\lambda_m^2} \|x(t) - z\|.$$

*Consequently, it is readily checked that the mapping  $\mathcal{A} : t \rightarrow \epsilon A_{\lambda(t)}x(t)$  is Lipschitz continuous on bounded intervals of  $[0, \infty)$ , provided that  $x$  and  $\lambda$  are so. This guarantees that  $\mathcal{A}$  is absolutely continuous on  $[0, \infty)$ , hence, that it is also differentiable almost everywhere on  $[0, \infty)$ .*

The following proposition is inspired by a similar result from Alvarez-Attouch-Bolte-Redont [4].

**Proposition 2.2.1.** *Suppose that  $\Phi : \mathcal{H} \rightarrow \mathbb{R}$  is twice continuously differentiable and that  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  is maximally monotone. Let  $\kappa$  be some positive constant, let  $\theta$  and  $\lambda$  be positive functions of class  $C^1$  on  $[0, \infty)$ , such that  $\lambda$  is bounded away from zero, and set  $\tau(\cdot) = \kappa - \theta - \frac{\dot{\theta}}{\theta}$ ,  $\eta(\cdot) = \kappa - \frac{\dot{\theta}}{\theta}$  and  $\Psi(t, x) = \nabla\Phi(x) + \epsilon A_{\lambda(t)}x$ , for  $(t, x) \in [0, \infty) \times \mathcal{H}$  and  $\epsilon \geq 0$ . Then, for any  $(x_0, \dot{x}_0) \in \mathcal{H}^2$ , the following statements **(i1)** and **(i2)** are equivalent :*

**(i1)**  $x : [0, \infty) \rightarrow \mathcal{H}$  is a  $C^2$  solution to the second-order differential equation

$$(2.2.16) \quad \ddot{x} + \tau\dot{x} + \frac{d}{dt}\Psi(\cdot, x) + \eta\Psi(\cdot, x) = 0,$$

with  $\dot{x}(0) = \dot{x}_0$ ,  $x(0) = x_0$ .

**(i2)**  $(x, y) : [0, \infty) \rightarrow \mathcal{H}^2$ , for some  $y \in C^1([0, \infty); \mathcal{H})$ , is a  $C^2 \times C^1$  solution to the first-order system

$$(2.2.17) \quad \begin{cases} \dot{x} + \Psi(\cdot, x) + \theta(y - x) = 0, \\ \dot{y} + \kappa(y - x) = 0, \end{cases}$$

with  $x(0) = x_0$  and  $y(0) = -\frac{1}{\theta(0)}(\dot{x}_0 + \Psi(0, x_0)) + x_0$ .

**Proof.** Let us prove that **(i1)**  $\Rightarrow$  **(i2)**. For  $t \geq 0$  we set

$$z(t) = \int_0^t [\tau(u)\dot{x}(u) + \eta(u)\Psi(u, x(u))] du - (\dot{x}_0 + \Psi(0, x_0)).$$

Then, integrating (2.2.16) on  $[0, t]$  with  $t \geq 0$  and differentiating  $z$ , gives us the following two equalities

$$(2.2.18a) \quad \dot{x}(t) + \Psi(t, x(t)) + z(t) = 0,$$

$$(2.2.18b) \quad \dot{z}(t) - \tau(t)\dot{x}(t) - \eta(t)\Psi(t, x(t)) = 0,$$

with initial conditions  $(x(0), z(0)) = (x_0, -(\dot{x}_0 + \Psi(0, x_0)))$ . Multiplying (2.2.18a) by  $\eta$ , and adding the resulting equality to (2.2.18b) allows us to get rid of the term  $\eta(t)\Psi(t, x(t))$  in (2.2.18b), so as to obtain

$$(2.2.19a) \quad \dot{x}(t) + \Psi(t, x(t)) + z(t) = 0,$$

$$(2.2.19b) \quad \eta(t)\dot{x}(t) + \eta(t)z(t) + \dot{z}(t) - \tau(t)\dot{x}(t) = 0.$$

Let us now define  $y = \frac{1}{\theta}z + x$ . It follows that  $z = \theta(y - x)$  and  $\dot{z} = \theta(\dot{y} - \dot{x}) + \dot{\theta}(y - x)$ .

Using also  $\tau = \kappa - \theta - \frac{\dot{\theta}}{\theta}$  and  $\eta = \kappa - \frac{\dot{\theta}}{\theta}$ , we can successively rewrite (2.2.19b) as follows :

$$\begin{aligned} 0 &= \eta \dot{x} + \eta z + \dot{z} - \tau \dot{x} \\ &= \left( \kappa - \frac{\dot{\theta}}{\theta} \right) \dot{x} + \left( \kappa - \frac{\dot{\theta}}{\theta} \right) \theta (y - x) + \theta (\dot{y} - \dot{x}) + \dot{\theta} (y - x) - \frac{1}{\theta} (\theta (\kappa - \theta) - \dot{\theta}) \dot{x} \\ &= \theta \dot{y} + \theta \kappa (y - x). \end{aligned}$$

Hence, as  $\theta$  is positive, we deduce that  $(x, y)$  verifies the following system

$$(2.2.20) \quad \dot{x}(t) + \Psi(t, x(t)) + \theta(t)(y(t) - x(t)) = 0,$$

$$(2.2.21) \quad \dot{y}(t) + \kappa(y(t) - x(t)) = 0,$$

with initial conditions  $(x(0), y(0)) = (x_0, -\frac{1}{\theta(0)}(\dot{x}_0 + \Psi(0, x_0)) + x_0)$ .

Let us prove that **(i2)**  $\Rightarrow$  **(i1)**. Differentiating the first equation of (2.2.17) yields

$$\ddot{x} + \frac{d}{dt} \Psi(., x) + \dot{\theta}(y - x) + \theta(\dot{y} - \dot{x}) = 0,$$

while the second equation of the system gives us  $\dot{y} = -\kappa(y - x)$ . So, by substituting the previous result into the former equation, we obtain

$$\ddot{x} + \frac{d}{dt} \Psi(., x) + (\dot{\theta} - \kappa\theta)((y - x) - \theta \dot{x}) = 0.$$

Furthermore, the first equation of (2.2.17) can be rewritten as  $y - x = -\frac{\dot{x} + \Psi(., x)}{\theta}$ . As a consequence, the prior second-order equation becomes

$$\ddot{x} + \frac{d}{dt} \Psi(., x) + \left( \kappa - \frac{\dot{\theta}}{\theta} \right) \Psi(., x) + \left( \kappa - \frac{\dot{\theta}}{\theta} - \theta \right) \dot{x} = 0.$$

This ends the proof. ■

### 2.2.2.2 Basic properties on the involved parameters

Some useful observation regarding the considered parameters are made through the following remarks.

**Remark 2.2.2.** Note that  $\theta(.)$  introduced in (2.2.6) can be rewritten as

$$(2.2.22) \quad \theta = \frac{\nu \rho}{\nu + e},$$

where  $\rho := \kappa - \frac{\dot{\nu}}{\nu}$ . This formulation of  $\theta$  will be very helpful in our upcoming computations, besides the following arguments :

**(i1)** From (2.2.22) we have  $\frac{1}{\theta} = \frac{1}{\rho} \left( \frac{e}{\nu} + 1 \right)$ . Then, as  $\nu$  and  $\rho$  are assumed to be positive (from (2.2.11a) and (2.2.11b)), we deduce that  $\theta$  is positive. Moreover,

observing that  $\frac{1}{\rho}$  and  $\frac{e}{\nu} + 1$  are non-increasing (from conditions (2.2.11a)-(2.2.11b)) entails that  $\frac{1}{\theta}$  is so, hence  $\theta$  is non-decreasing. In addition, from the definition  $\rho := \kappa - \frac{\dot{\nu}}{\nu}$ , we obviously infer that  $\rho(t) \in (0, \kappa]$  for  $t \geq 0$ , and that  $\rho(t) \rightarrow \kappa$  as  $t \rightarrow \infty$  (under condition (2.2.13)). Consequently, by (2.2.22), we obtain  $\theta(t) \in (0, \kappa)$  for  $t \geq 0$  and  $\theta(t) \rightarrow \kappa$  as  $t \rightarrow \infty$ .

(i2) It can be noticed that (2.2.11c) yields  $0 < \nu(t) \leq Mt + \nu(0)$ , which simply leads to  $\int_0^{+\infty} \frac{1}{\nu(t)} dt = +\infty$ .

**Remark 2.2.3.** The following items (j1)-(j3) are also of great importance with regard to our methodology :

(j1) Let us stress that conditions (2.2.11) and (2.2.13) guarantee that the mappings  $\tau(\cdot) = \kappa - \theta - \frac{\dot{\theta}}{\theta}$  and  $\eta(\cdot) = \kappa - \frac{\dot{\theta}}{\theta}$  (where  $\theta$  is given by (2.2.6)) are continuously differentiable on  $[0, \infty)$  and satisfy

$$(2.2.23) \quad \tau(t) \sim \frac{1}{\nu(t)} (l + \kappa e) \quad \text{and} \quad \eta(t) \sim \kappa \quad (\text{as } t \rightarrow +\infty).$$

Indeed, by Remark 2.2.2, we know that  $\theta$  is well-defined and positive on  $[0, \infty)$  under condition (2.2.11). Moreover, from  $\nu \in C^2$ ,  $\nu > 0$  and  $\theta = \frac{\kappa\nu - \dot{\nu}}{\nu + e}$  (hence  $\theta = \kappa - \frac{\dot{\nu} + \kappa e}{\nu + e}$ ), we can see that  $\theta \in C^1([0, \infty))$  and we readily get

$$(2.2.24) \quad \theta(\kappa - \theta) = \frac{(\kappa\nu - \dot{\nu})(\dot{\nu} + \kappa e)}{(\nu + e)^2} \quad \text{and} \quad \dot{\theta} = \frac{\dot{\nu}(\dot{\nu} + \kappa e) - (\nu + e)\ddot{\nu}}{(\nu + e)^2}.$$

Consequently, by these observations, we deduce that the functions  $\tau(\cdot)$  and  $\eta(\cdot)$  (in light of their definitions) belong to  $C^1([0, \infty))$  and we obtain

$$\tau = \frac{1}{\kappa\nu - \dot{\nu}} \left( (\dot{\nu} + \kappa e) \frac{(\kappa\nu - 2\dot{\nu})}{\nu + e} + \ddot{\nu} \right) \quad \text{and} \quad \eta = \kappa - \frac{1}{(\kappa\nu - \dot{\nu})} \left( \frac{\dot{\nu}(\dot{\nu} + \kappa e)}{\nu + e} - \ddot{\nu} \right).$$

So, under the the additional condition (2.2.13), we obviously obtain (2.2.23).

(j2) Let us recall that  $\lambda(\cdot)$  (defined in (2.2.12b)) is of the form  $\lambda = \frac{\lambda_c}{\theta} (\nu + e)^2$  (for some  $\lambda_c > 0$ ). So it is easily seen under condition (2.2.11) that  $\lambda(\cdot)$  is differentiable on  $[0, \infty)$ . Moreover it can be checked that condition (2.2.13) ensures that  $\lambda$  satisfies the following properties :

$$(2.2.25a) \quad \lambda(t) \sim \left( \frac{\lambda_c}{\kappa} \right) \nu^2(t) \quad (\text{as } t \rightarrow \infty);$$

$$(2.2.25b) \quad |\dot{\lambda}(t)| \leq k_2 \nu(t), \quad \text{for } t \geq 0 \quad (k_2 \text{ being some positive constant});$$

$$(2.2.25c) \quad \exists t_0 > 0 \quad \text{s.t.} \quad \lambda \text{ is non-decreasing on } [t_0, \infty).$$

Indeed, (2.2.25a) is obvious since  $\theta(t) \sim \kappa$  (as  $t \rightarrow \infty$ ) according to Remark 2.2.2. Furthermore, we simply have  $\lambda = \lambda_c \frac{(\nu+e)^3}{(\kappa\nu-\dot{\nu})}$  (by using the definition of  $\theta$ ) and from a quick calculation we obtain

$$\begin{aligned}\dot{\lambda}(t) &= \lambda_c \frac{(\nu+e)^2}{(\kappa\nu-\dot{\nu})^2} (3\dot{\nu}(\kappa\nu - \dot{\nu}) - (\nu+e)(\kappa\dot{\nu} - \ddot{\nu})) \\ &= \lambda_c \frac{(\nu+e)^2}{(\kappa\nu-\dot{\nu})^2} (\nu(2\kappa\dot{\nu} + \ddot{\nu}) - 3\dot{\nu}^2 - e(\kappa\dot{\nu} - \ddot{\nu})).\end{aligned}$$

Then, by  $\dot{\nu}(t) \rightarrow l$  and  $\ddot{\nu}(t) \rightarrow 0$  (as  $t \rightarrow \infty$ ), we can see that  $\frac{\dot{\lambda}}{\nu} \rightarrow \frac{2l\lambda_c}{\kappa}$  is bounded, which ensures (2.2.25b) and (2.2.25c) (as  $\dot{\lambda}(t)$  is positive for  $t$  large enough).

**(j3)** For  $\epsilon > 0$ , under the supplementary conditions (2.2.14)-(2.2.15), the function  $\eta(\cdot)$  is continuously differentiable on  $[0, \infty)$  and we show that  $(e + \nu)\dot{\eta}$  is bounded on  $[0, \infty)$ . Indeed, by an easy computation, for  $\nu \in C^3([0, \infty))$  we obtain  $\dot{\eta} = \frac{1}{\theta^2}(\dot{\theta}^2 - \theta\ddot{\theta})$ , together with  $\dot{\theta} = \frac{\dot{\nu}^2 + \kappa e \ddot{\nu}}{(\nu+e)^2} - \frac{\ddot{\nu}}{(\nu+e)}$  (in light of (2.2.24)), which gives us

$$\ddot{\theta} = \frac{2\ddot{\nu}\dot{\nu} + \kappa e \ddot{\nu}}{(\nu+e)^2} - 2\frac{\dot{\nu}^2(\dot{\nu} + \kappa e)}{(\nu+e)^3} - \frac{\nu^{(3)}}{(\nu+e)} + \frac{\dot{\nu}\ddot{\nu}}{(\nu+e)^2},$$

where  $\nu^{(3)}$  stands for the third derivative of  $\nu$ . Clearly, by these formulations of  $\dot{\eta}$ ,  $\dot{\theta}$  and  $\ddot{\theta}$ , along with the boundedness of  $\dot{\nu}$ ,  $\ddot{\nu}$  and  $\nu^{(3)}$  (under condition (2.2.13) and (2.2.15)), while recalling that  $\theta$  is positive and non-decreasing  $[0, \infty)$  (from Remark 2.2.2), we infer that

$$(\nu + e)|\dot{\eta}| \leq \frac{1}{\theta^2(0)}(\nu + e)|\dot{\theta}^2| + \frac{1}{\theta(0)}(\nu + e)|\dot{\theta}| \leq k_1, \quad \text{for some constant } k_1 \geq 0.$$

### 2.2.2.3 Existence and uniqueness results

Actually, we establish a result regarding the model (2.2.5) (independently of the specific formulation of  $\theta$  given by (2.2.6)).

**Proposition 2.2.2.** *Let a function  $\Phi : \mathcal{H} \rightarrow \mathbb{R}$  and an operator  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  verify conditions (H1)-(H3), let  $\kappa > 0$ , and consider real-valued mappings  $\lambda(\cdot)$  and  $\theta$  satisfying :*

$$(2.2.26) \quad \lambda \in C^0([0, \infty)) \text{ and, for } t \geq 0, \lambda(t) \in [\lambda_m, \infty) \text{ (for some constant } \lambda_m > 0);$$

$$(2.2.27) \quad \theta \in C^0([0, \infty)) \text{ and, for } t \geq 0, \theta(t) \in [0, \theta_M] \text{ (for some constant } \theta_M > 0).$$

Then for any Cauchy data  $(x_0, y_0) \in \mathcal{H}^2$ , there exists a unique global classical solution  $(x, y)$  to (2.2.5) verifying the initial condition  $(x(0), y(0)) = (x_0, y_0)$ .

**Proof.** Observe that (2.2.5) can be formulated as

$$(2.2.28) \quad \dot{U}(t) + \nabla \bar{\Phi}(U(t)) + E(t, U) = 0,$$

where  $U = (x, y)$ , while  $\bar{\Phi} : \mathcal{H}^2 \rightarrow \mathcal{H}^2$  and  $E(t, \cdot) : \mathcal{H}^2 \rightarrow \mathcal{H}^2$  are defined for any  $t \geq 0$  and for any  $(u, v) \in \mathcal{H}^2$  by

$$(2.2.29) \quad \bar{\Phi}(u, v) = \begin{pmatrix} \Phi(u) \\ 0 \end{pmatrix}^T, \quad E(t, (u, v)) = \begin{pmatrix} \phi_1(t, u, v) \\ \phi_2(t, u, v) \end{pmatrix}^T,$$

along with  $\phi_1(t, u, v) = \epsilon A_{\lambda(t)} u + \theta(t)(v - u)$  and  $\phi_2(t, u, v) = \kappa(v - u)$ . Note that (2.2.28) is governed by the sum of the convex subdifferential operator  $\partial \bar{\Phi} = \nabla \bar{\Phi}$  and the operator  $E : [0, \infty) \times \mathcal{H}^2 \rightarrow \mathcal{H}^2$ . In view of applying [31, Proposition 3.12] (see also [60, Proposition 6.2.1]) we prove that  $E$  satisfies the following properties **(a)** and **(b)** :

**(a)**  $E(t, \cdot)$  is globally Lipschitz continuous on  $\mathcal{H}^2$ . Clearly, given  $t \geq 0$  and  $(u_i, v_i) \in \mathcal{H}^2$  (for  $i = 1, 2$ ), and invoking the boundedness of  $\theta$  in (2.2.27) we readily obtain

$$\|\phi_1(t, u_1, v_1) - \phi_1(t, u_2, v_2)\| \leq \epsilon \|(A_{\lambda(t)} u_1 - A_{\lambda(t)} u_2)\| + \theta_M \|(v_1 - u_1) - (v_2 - u_2)\|.$$

Moreover, recalling that  $A_\lambda$  is  $\lambda^{-1}$ -Lipschitz continuous and assuming that  $\lambda(t) \geq \lambda_m > 0$  (for some constant  $\lambda_m$ ), we also have

$$\|(A_{\lambda(t)} u_1 - A_{\lambda(t)} u_2)\| \leq \frac{1}{\lambda_m} \|u_1 - u_2\|.$$

Then, by the previous two inequalities we get

$$\|\phi_1(t, u_1, v_1) - \phi_1(t, u_2, v_2)\| \leq \left( \frac{\epsilon}{\lambda_m} + \theta_M \right) (\|v_1 - v_2\| + \|u_1 - u_2\|);$$

while an easy computation gives us

$$\|\phi_2(t, u_1, v_1) - \phi_2(t, u_2, v_2)\| \leq \kappa (\|v_1 - v_2\| + \|u_1 - u_2\|).$$

It follows that  $E(t, \cdot)$  is globally Lipschitz continuous on  $\mathcal{H}^2$ .

**(b)** For any  $T \in (0, \infty)$ ,  $(u, v) \in \mathcal{H}^2$ ,  $E(\cdot, u, v) \in L^2(0, T; \mathcal{H}^2)$ . Taking  $z \in (\epsilon A)^{-1}(0)$  (under condition (H3)), by Lemma A.1 and recalling that  $\lambda(t) \geq \lambda_m > 0$ , we obtain

$$\|\epsilon A_{\lambda(t_1)} u - \epsilon A_{\lambda(t_2)} u\| \leq 3 \frac{|\lambda(t_1) - \lambda(t_2)|}{\lambda_m^2} \|u - z\|.$$

Then, the continuity of  $\lambda$  on  $[0, \infty)$  yields that the mapping  $t \mapsto \epsilon A_{\lambda(t)} u$  is also continuous on  $[0, \infty)$ . So, in light of the definition of  $\phi_1$  and  $\phi_2$ , we infer that  $E(\cdot, u, v)$  is continuous on  $[0, \infty)$  (as are  $\phi_1$  and  $\phi_2$ ). It follows immediately the desired result of item **(b)**.

Therefore, by [31, Proposition 3.12], there exists a unique pair  $(x, y) \in C([0, \infty), \mathcal{H}^2)$  such that  $x$  and  $y$  are absolutely continuous on every compact subset of  $(0, \infty)$  and satisfy (2.2.5) almost everywhere. Hence by continuity of  $\nabla\Phi(\cdot)$  on  $\mathcal{H}$ , we infer that  $(x, y) \in C^1([0, \infty), \mathcal{H}^2)$ .  $\blacksquare$

## 2.2.3 Preliminary estimations

### 2.2.3.1 A general equality for closely related models

Consider the more general model given by

$$(2.2.30a) \quad \dot{x}(t) + \psi(t) + \theta(t)(y(t) - x(t)) = 0,$$

$$(2.2.30b) \quad \dot{y}(t) + \kappa(y(t) - x(t)) = 0,$$

$$(2.2.30c) \quad \theta(t) = \frac{\kappa\nu(t) - \dot{\nu}(t)}{\nu(t) + e},$$

where  $\psi : [0, \infty) \rightarrow \mathcal{H}$  is continuous,  $e$  is a positive constant, and  $\nu$  is a positive and differentiable mapping on  $[0, \infty)$ . As a key element of our analysis, we associate with any solution  $(x, y)$  to (2.2.30) the functional  $G_{s,q}$  defined for any couple  $(s, q) \in [0, +\infty[ \times \mathcal{H}$  by

$$(2.2.31) \quad G_{s,q} := \frac{1}{2} \|s(q - x) + \nu(y - x)\|^2 + \frac{s(e - s)}{2} \|x - q\|^2.$$

**Lemma 2.2.1.** *Let  $\psi : [0, \infty) \rightarrow \mathcal{H}$  be a continuous function, and suppose that  $(x, y) : [0, \infty) \rightarrow \mathcal{H}^2$  is a global classical solution to (2.2.30), where  $\{\kappa, e\}$  are positive constants, while  $\nu(\cdot)$  is a positive and differentiable function that verifies  $\dot{\nu} < \kappa\nu$  on  $[0, \infty)$ . Then, for any  $(s, q) \in [0, \infty) \times \mathcal{H}$ , on  $[0, \infty)$  we have*

$$(2.2.32) \quad \dot{G}_{s,q} + s(e + \nu)\langle \psi, x - q \rangle + \frac{(e + \nu)^2}{\rho} \langle \psi, \dot{x} \rangle = -\nu^2 \tilde{\mathcal{T}}_s,$$

where  $\rho = \kappa - \frac{\dot{\nu}}{\nu}$ , together with

$$(2.2.33) \quad \tilde{\mathcal{T}}_s = \rho \left\| y - x + \frac{1}{\theta} \left( 1 - \frac{e-s}{2(e+\nu)} \right) \dot{x} \right\|^2 + \left( \frac{e-s}{4\rho\nu} \right) \left( \frac{s+3e}{\nu} + 4 \right) \|\dot{x}\|^2.$$

**Proof.** Observe that (2.2.31) can be written as  $G_{s,q} = seb + s\nu a + \nu^2 c$ , where  $b(\cdot) := \frac{1}{2} \|x - q\|^2$ ,  $a(\cdot) := \langle q - x, y - x \rangle$  and  $c(\cdot) := \frac{1}{2} \|y - x\|^2$ . Consequently, the

derivative of  $G_{s,q}$  on  $[0, \infty)$  is given by

$$(2.2.34) \quad \dot{G}_{s,q} = se\dot{b} + s\dot{v}a + s\nu\dot{a} + \nu^2\dot{c} + 2\dot{\nu}\nu c.$$

The rest of the proof can be divided into the following steps **(A1)**-**(A3)** :

**(A1)** In order to estimate the time derivative  $\dot{G}_{s,q}$ , we evaluate separately each of the terms occurring on the right side of (2.2.34). Let us prove the following basic results on  $a(\cdot)$  and  $c(\cdot)$  :

$$(2.2.35a) \quad a = -\theta^{-1}P + \theta^{-1}W,$$

$$(2.2.35b) \quad \dot{a} = -Q + (-1 + \kappa\theta^{-1})P - \kappa\theta^{-1}W,$$

$$(2.2.35c) \quad \dot{b} = -P, \quad \dot{c} = -\kappa\|u\|^2 - Q,$$

where  $P = \langle q - x, \dot{x} \rangle$ ,  $Q = \langle \dot{x}, u \rangle$ ,  $W = \langle \psi, x - q \rangle$  and  $u = y - x$ .

From the definitions of  $a(\cdot)$  and  $c(\cdot)$ , we immediately have

$$\dot{a} = -Q - P + R \quad \text{and} \quad \dot{c} = -\kappa\|u\|^2 - Q, \quad \text{where } R := \langle \dot{y}, q - x \rangle.$$

From (2.2.5) we get  $\dot{x} + \psi - \frac{\theta}{\kappa}\dot{y} = 0$ . Taking the scalar product of each side of this latter equality with  $q - x$  amounts to  $P - W - \frac{\theta}{\kappa}R = 0$ , which can be rewritten as  $R = \frac{\kappa}{\theta}P - \frac{\kappa}{\theta}W$ . Moreover, by  $a = \langle q - x, u \rangle$  and  $u = -\frac{1}{\kappa}\dot{y}$ , we deduce that  $a = -\frac{1}{\kappa}R$ . Hence, taking into account the latter formulation of  $R$  yields  $a = -\frac{1}{\theta}P + \frac{1}{\theta}W$ , that is (2.2.35a). Next, by combining this formulation of  $R$  with the previous last estimate of  $\dot{a}$ , we get  $\dot{a} = -Q + (-1 + \frac{\kappa}{\theta})P - \frac{\kappa}{\theta}W$ , that is (2.2.35b). The first result in (2.2.35c) is obvious from the definition of  $b$ . In addition, from  $c(t) = (1/2)\|u(t)\|^2$  with  $u = y - x$ , it is deduced that  $\dot{c} = \langle \dot{u}, u \rangle$ . Then, by  $\dot{u} = \dot{y} - \dot{x}$  and  $\dot{y} = -\kappa u$ , we obtain the second result in (2.2.35c).

**(A2)** Now, we are in position to prove that

$$(2.2.36) \quad \dot{G}_{s,q} = -(\nu^2 + s\nu) \langle \dot{x}, u \rangle - s(\nu + e)W - \nu^2 \left( \kappa - \frac{\dot{\nu}}{\nu} \right) \|u\|^2,$$

where  $u = y - x$  and  $W = \langle \psi, x - q \rangle$ . From (2.2.34), we directly get

$$\dot{G}_{s,q} = se\dot{b} + s\dot{v}a + s\nu\dot{a} + \nu^2\dot{c} + 2\dot{\nu}\nu c.$$

So, using (2.2.35a)-(2.2.35b)-(2.2.35c) and  $\dot{b} = -P$ , we have

$$\begin{aligned} \dot{G}_{s,q} &= -seP + s\dot{\nu} \left( -\frac{1}{\theta(t)}P + \frac{1}{\theta(t)}W \right) \\ &\quad + s\nu \left( -Q + \left( -1 + \frac{\kappa}{\theta} \right) P - \frac{\kappa}{\theta}W \right) + \nu^2 \left( -\kappa\|u\|^2 - Q \right) + \dot{\nu}\nu\|u\|^2. \end{aligned}$$

Hence, recalling that  $\theta = \frac{\kappa\nu - \dot{\nu}}{\nu + e}$ , we get  $e = -\frac{\dot{\nu}}{\theta} + \nu(-1 + \frac{\kappa}{\theta})$ , so that the previous equality reduces to

$$\dot{G}_{s,q} = -(\nu^2 + s\nu)Q + \frac{s}{\theta}(\dot{\nu} - \kappa\nu)W + \nu(\dot{\nu} - \kappa\nu)\|u\|^2.$$

Thus, recalling that  $\dot{\nu} - \kappa\nu = -\theta(\nu + e)$  leads to the desired result.

**(A3)** Now, we are in position to prove the main result of the lemma. Take  $q \in \mathcal{H}$ . According to Lemma 2.2.1 we have

$$\dot{G}_{s,q} + s(e + \nu)W = -(\nu^2 + s\nu)Q - \nu^2\rho\|u\|^2.$$

Moreover, in light of  $\dot{x} + \psi + \theta u = 0$  (from (2.2.5)), we readily get

$$\|\dot{x}\|^2 + \langle \psi, \dot{x} \rangle + \theta \langle u, \dot{x} \rangle = 0.$$

Thus, multiplying this last equality by  $\mu := \frac{(e+\nu)^2}{\rho}$  and adding the resulting equality to the previous estimate on  $\dot{G}_{s,q}$  amounts to

$$\dot{G}_{s,q} + s(e + \nu)W + \mu \langle \psi, \dot{x} \rangle = -(\nu^2 + s\nu + \mu\theta) \langle \dot{x}, u \rangle - \nu^2\rho\|u\|^2 - \mu\|\dot{x}\|^2,$$

or equivalently

$$\dot{G}_{s,q} + s(e + \nu)W + \mu \langle \psi, \dot{x} \rangle = -\nu^2 \tilde{\mathcal{T}}_s,$$

where  $\tilde{\mathcal{T}}_s = w \langle \dot{x}, u \rangle + \sigma \|\dot{x}\|^2 + \rho \|u\|^2$ , with  $w = 1 + \frac{s}{\nu} + \frac{\mu\theta}{\nu^2}$  and  $\sigma = \frac{\mu}{\nu^2}$ . Now we give an alternative formulation of the mapping  $\tilde{\mathcal{T}}_s$ . One can easily check that  $\tilde{\mathcal{T}}_s$  can be rewritten as

$$\tilde{\mathcal{T}}_s = \rho \left\| u + \frac{w}{2\rho} \dot{x} \right\|^2 + \left( \frac{4\rho\sigma - w^2}{4\rho} \right) \|\dot{x}\|^2.$$

Moreover, by  $\theta = \frac{\rho\nu}{e+\nu}$  and  $\mu = \frac{(\nu+e)^2}{\rho}$ , we get  $\frac{\mu\theta}{\nu^2} = \frac{e}{\nu} + 1$ , and so, by  $w = 1 + \frac{s}{\nu} + \frac{\mu\theta}{\nu^2}$  (hence  $w = \frac{s+e}{\nu} + 2$ ) and noticing that  $\rho = \theta \left( \frac{e}{\nu} + 1 \right)$ , we obtain

$$\frac{w}{2\rho} = \frac{\frac{s+e}{\nu} + 2}{2\theta \left( \frac{e}{\nu} + 1 \right)} = \frac{1}{\theta} \left( 1 - \frac{\frac{1}{2\nu}(e-s)}{\frac{e}{\nu} + 1} \right) = \frac{1}{\theta} \left( 1 - \frac{e-s}{2(e+\nu)} \right).$$

Observing in addition that  $\sigma = \frac{(\nu+e)^2}{\rho\nu^2}$ , we also obtain

$$4\sigma\rho - w^2 = 4 \left( \frac{e}{\nu} + 1 \right)^2 - \left( \frac{s+e}{\nu} + 2 \right)^2 = \left( \frac{e-s}{\nu} \right) \left( \frac{s+3e}{\nu} + 4 \right).$$

So, combining the last expression of  $\tilde{\mathcal{T}}$  and the previous two results gives us

$$\tilde{\mathcal{T}}_s = \rho \left\| u + \frac{1}{\theta} \left( 1 - \frac{e-s}{2(e+\nu)} \right) \dot{x} \right\|^2 + \frac{1}{4\rho} \left( \frac{e-s}{\nu} \right) \left( \frac{s+3e}{\nu} + 4 \right) \|\dot{x}\|^2.$$

This along with our last estimation on yields the desired result. ■

### 2.2.3.2 Exhibiting a Lyapunov functional

In this section we state our main result regarding the Lyapunov-like function  $\mathcal{E}_{s,q}$  introduced in (2.2.10).

**Lemma 2.2.2.** *Let (H1)-(H3) hold, and let  $(x, y) : [0, +\infty[ \rightarrow \mathcal{H}^2$  be a classical solution to (2.2.5)-(2.2.6), with positive constants  $\{\kappa, e\}$ , a function  $\nu(\cdot)$  verifying (2.2.11), and a positive mapping  $\lambda(\cdot)$ . Then, for  $(s, q) \in [0, e] \times S$ , on  $[0, \infty)$ , we have*

$$(2.2.37) \quad \dot{\mathcal{E}}_{s,q} + (s - e_c)v\tilde{\Phi}(x) + \epsilon s \lambda v \|A_\lambda x\|^2 + \epsilon \frac{v^2}{\rho} \langle A_\lambda x, \dot{x} \rangle \leq -\nu^2 \tilde{\mathcal{T}}_s.$$

where  $e_c = \frac{2M}{\rho(0)}$ ,  $v = e + \nu$ ,  $\rho = \kappa - \frac{\dot{\nu}}{\nu}$ , and  $\tilde{\mathcal{T}}_s$  is defined in (2.2.33).

**Proof.** Note that (2.2.5)-(2.2.6) enters the setting of (2.2.30) with  $\psi = \nabla\Phi(x) + \epsilon A_\lambda x$ . Take  $q \in S$  (hence  $\epsilon A_\lambda(q) = 0$  and  $\nabla\Phi(q) = 0$ ). In light of Lemma 2.2.1, with this particular expression of  $\psi$ , we obtain

$$\dot{G}_{s,q} + svW + \frac{v^2}{\rho}Z = -\nu^2\tilde{\mathcal{T}}_s,$$

where  $W = \langle \nabla\Phi(x) + \epsilon A_\lambda x, x - q \rangle$  and  $Z = \langle \nabla\Phi(x) + \epsilon A_\lambda x, \dot{x} \rangle$ . Let us estimate  $W$  and  $Z$ . Using the convexity of  $\Phi$  and the  $\lambda$ -co-coerciveness of  $A_\lambda$ , by  $\tilde{\Phi}(x) = \Phi(x) - \Phi(q)$  we get  $\langle \nabla\Phi(x), x - q \rangle \geq \tilde{\Phi}(x)$  and  $\langle \epsilon A_\lambda x, x - q \rangle \geq \epsilon \lambda \|A_\lambda x\|^2$ . It follows immediately that  $W \geq \tilde{\Phi}(x) + \epsilon \lambda \|A_\lambda x\|^2$ . Moreover, noticing that  $\frac{d}{dt}(\tilde{\Phi}(x)) = \langle \nabla\Phi(x), \dot{x} \rangle$ , we simply have  $Z = \frac{d}{dt}(\tilde{\Phi}(x)) + \epsilon \langle A_\lambda x, \dot{x} \rangle$ . Setting  $\mu = \frac{v^2}{\rho}$  (or equivalently  $\mu = \frac{\nu(\nu+e)}{\theta}$ , as  $\theta = \frac{\nu\rho}{\nu+e}$  (from (2.2.22))), by these two results in light of the previous estimate on  $\dot{G}_{s,q}$ , we are led to

$$(2.2.38) \quad \dot{G}_{s,q} + \epsilon C + D \leq -\nu^2 \tilde{\mathcal{T}}_s,$$

where  $C(\cdot) = s \lambda v \|A_\lambda x\|^2 + \mu \langle A_\lambda x, \dot{x} \rangle$  and  $D = sv\tilde{\Phi}(x) + \mu \frac{d}{dt}(\tilde{\Phi}(x))$ . Let us evaluate  $D$ . From its definition we clearly have  $D = \frac{d}{dt}(\mu\tilde{\Phi}) + (sv - \dot{\mu})\tilde{\Phi}$ . Concerning the derivative of  $\mu$ , by  $\mu = \frac{1}{\rho} \times (\nu + e)^2$  we also have  $\dot{\mu} = \frac{1}{\rho} \times \frac{d}{dt}v^2 + \frac{d}{dt}\left(\frac{1}{\rho}\right) \times v^2$ ; hence, noticing from condition (2.2.11b) that  $\rho$  is non-decreasing (thus  $\frac{d}{dt}\left(\frac{1}{\rho}\right) \leq 0$ ), and recalling that  $\sup_{t \geq 0} \|\dot{\nu}(t)\| \leq M$  (from (2.2.11c)), we get

$$\dot{\mu} \leq \frac{1}{\rho(0)} \times \frac{d}{dt}(e + \nu)^2 = \frac{2\dot{\nu}(e+\nu)}{\rho(0)} \leq \frac{2Mv}{\rho(0)}.$$

Consequently, by the previous estimates we deduce that

$$D \geq \frac{d}{dt}(\mu\tilde{\Phi}(x)) + \left[s - \frac{2M}{\rho(0)}\right]v\tilde{\Phi}(x).$$

So, combining this last inequality with (2.2.38) entails (2.2.37). ■

**Remark 2.2.4.** Clearly, by (2.2.12) and setting  $e_c = \frac{2M}{\rho(0)}$ , we have  $e_c \in (0, e)$  and  $\lambda = \lambda_c \frac{(\nu+e)^2}{\theta}$ , where  $\lambda_c$  is assumed to verify  $\lambda_c > (\frac{\epsilon}{4}) \inf_{s \in (e_c, e)} \frac{1}{s(e-s)}$ . Moreover, we recall that  $\theta = \frac{(\kappa\nu - \dot{\nu})}{e+\nu}$  (from (2.2.6)). This amounts to  $\lambda = \lambda_c \frac{(\nu+e)^3}{(\kappa\nu - \dot{\nu})} = \lambda_c \frac{\mu(e+\nu)}{\nu}$  (where  $\mu = \frac{\nu(e+\nu)^2}{\kappa\nu - \dot{\nu}}$ ). So, it is not difficult to see for  $\epsilon > 0$  that the assumptions (2.2.12) on the function  $\lambda(\cdot)$  can be alternatively expressed as :  $e_c \in (0, e)$  and  $\lambda = \frac{\mu(e+\nu)}{\nu s_c (e-s_c) \delta}$ , for some  $s_c \in (e_c, e)$  and some  $\delta > 0$  such that  $\epsilon\delta \in [0, 4)$ .

**Proposition 2.2.3.** *Let  $\Phi : \mathcal{H} \rightarrow \mathbb{R}$  and  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  verify (H1)-(H3) and let  $(x, y) : [0, +\infty[ \rightarrow \mathcal{H}^2$  be a global classical solution of (2.2.5)-(2.2.6), with positive constants  $\{\kappa, e\}$  and mappings  $\nu(\cdot)$  and  $\lambda(\cdot)$  satisfying (2.2.11) and (2.2.12). Then, for  $(s, q) \in (0, e] \times S$ , we have the following inequality on  $[0, \infty)$  :*

$$(2.2.39) \quad \dot{\mathcal{E}}_{s,q} + (s - e_c) v(\Phi - \Phi(q)) + s\lambda v \epsilon \|A_\lambda x + \frac{v}{2s\lambda\rho} \dot{x}\|^2 \leq -\nu^2 \mathcal{T}_s,$$

where  $e_c = \frac{2M}{\rho(0)}$ ,  $v = e + \nu$ ,  $\rho = \kappa - \frac{\dot{\nu}}{\nu}$ , while  $\mathcal{T}_s$  is defined, for some  $s_c \in (e_c, e)$  and some  $\delta > 0$  (occurring in Remark 2.2.4) verifying  $\epsilon\delta \in [0, 4)$ , by

$$(2.2.40) \quad \mathcal{T}_s = \rho \left\| y - x + \frac{1}{\theta} \left( 1 - \frac{e-s}{2(e+\nu)} \right) \dot{x} \right\|^2 + \frac{1}{4\rho\nu} \left( (e-s) \left( \frac{s+3e}{\nu} + 4 - \epsilon\delta \right) + \frac{\epsilon\delta}{s} (s-s_c)(e-s_c-s) \right) \|\dot{x}\|^2.$$

So, the functional  $\mathcal{E}_{s_c,q}$  is non-increasing and convergent as  $t \rightarrow \infty$ .

**Proof.** It can be noticed that (2.2.39) reduces to (2.2.37) when  $\epsilon = 0$ . Now, we let  $\epsilon > 0$  and we consider the quantity  $C(\cdot) := s\lambda v \|A_\lambda x\|^2 + \mu \langle A_\lambda x, \dot{x} \rangle$  that arises in the left side of (2.2.37). To estimate  $C$ , we take  $s \in (0, e)$  and we set  $k_s = \frac{\mu}{2s\lambda v}$ . So, by  $\|A_\lambda x\|^2 = \|A_\lambda x + k_s \dot{x} - k_s \dot{x}\|^2$  we simply have

$$\|A_\lambda x\|^2 = \|A_\lambda x + k_s \dot{x}\|^2 - 2k_s \langle A_\lambda x, \dot{x} \rangle - k_s^2 \|\dot{x}\|^2.$$

Hence, by  $\mu = 2s\lambda v k_s$  (from definition of  $k_s$ ) along with the definition of  $C$ , we get

$$\begin{aligned} C &= s\lambda v \|A_\lambda x + k_s \dot{x}\|^2 + (\mu - 2s\lambda v k_s) \langle A_\lambda x, \dot{x} \rangle - s\lambda v k_s^2 \|\dot{x}\|^2 \\ &= s\lambda v \|A_\lambda x + k_s \dot{x}\|^2 - s\lambda v k_s^2 \|\dot{x}\|^2, \end{aligned}$$

or equivalently  $C = s\lambda v \|A_\lambda x + k_s \dot{x}\|^2 - \nu^2 \gamma \|\dot{x}\|^2$ , where  $\gamma := \nu^{-2} (s\lambda v k_s^2)$ . This, in light of (2.2.37), amounts to

$$\dot{\mathcal{E}}_{s,q} + [s - e_c] v \tilde{\Phi}(x) + s\lambda v \epsilon \|A_\lambda x + k_s \dot{x}\|^2 \leq -\nu^2 (\tilde{\mathcal{T}}_s - \epsilon \gamma \|\dot{x}\|^2).$$

Observe from Remark 2.2.4 that  $\lambda = \frac{\mu v}{\nu s_c (e-s_c) \delta}$ , for some  $s_c \in (e_c, e)$  and some  $\delta > 0$  such that  $\epsilon\delta \in [0, 4)$ . In addition, by the definition of  $\gamma$  we get  $\gamma = \nu^{-2} (s\lambda v k_s^2) = \nu^{-2} \frac{\mu^2}{4s\lambda v}$  (since  $k_s = \frac{\mu}{2s\lambda v}$ ), hence, from  $\lambda = \frac{\mu v}{\nu s_c (e-s_c) \delta}$ , we get

$$\gamma = \nu^{-2} \frac{\mu^2}{4sv} \times \frac{\nu s_c(e-s_c)\delta}{\mu v} = \frac{s_c(e-s_c)\delta}{4s\rho\nu}.$$

Therefore, setting  $u = y - x$  and using the definition of  $\tilde{\mathcal{T}}_s$  in Lemma 2.2.1, we have

$$\begin{aligned} & \tilde{\mathcal{T}}_s - \epsilon\gamma \|\dot{x}\|^2 - \rho \left\| u + \frac{1}{\theta}\dot{x} - \frac{e-s}{2\rho\nu}\dot{x} \right\|^2 \\ &= \left( \frac{e-s}{4\rho\nu} \right) \left( \frac{s+3e}{\nu} + 4 \right) \|\dot{x}\|^2 - \epsilon \frac{s_c\delta}{4s\rho\nu} (e-s_c) \|\dot{x}\|^2 \\ &= \left( \frac{e-s}{4\rho\nu} \right) \left( \frac{s+3e}{\nu} + 4 - \epsilon\delta \right) \|\dot{x}\|^2 + \left( \frac{\delta}{4\rho\nu} \right) \left( \epsilon(e-s) - \epsilon \frac{s_c(e-s_c)}{s} \right) \|\dot{x}\|^2 \\ &= \left( \frac{e-s}{4\rho\nu} \right) \left( \frac{s+3e}{\nu} + 4 - \epsilon\delta \right) \|\dot{x}\|^2 + \left( \frac{\delta}{4\rho\nu} \right) \left( \frac{\epsilon}{s} \right) (s-s_c)(e-s_c-s) \|\dot{x}\|^2. \end{aligned}$$

It follows from our previous estimate on  $\dot{\mathcal{E}}_{s,q}$  that

$$\dot{\mathcal{E}}_{s,q} + [s - e_c] v \tilde{\Phi}(x) + s\lambda v \epsilon \|A_\lambda x + k_s \dot{x}\|^2 \leq -\nu^2 \mathcal{T}_s,$$

where  $\mathcal{T}_s = \tilde{\mathcal{T}}_s + \left( \frac{\epsilon}{s} \right) (s-s_c)(e-s_c-s) \left( \frac{\delta}{4\rho\nu} \right) \|\dot{x}\|^2$ . This leads to (2.2.39). The remainder of the proof is obvious.  $\blacksquare$

## 2.2.4 Asymptotic properties of a classical solution

### 2.2.4.1 Main convergence results

#### 2.2.4.2 The general case

In this section we expose our main asymptotic results regarding (2.2.5)-(2.2.6).

**Theorem 2.2.1.** *Let  $\Phi : \mathcal{H} \rightarrow \mathbb{R}$  and  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  verify **(H1)**-**(H3)** and assume that  $(x, y) : [0, +\infty[ \rightarrow \mathcal{H}^2$  is a global classical solution to (2.2.5)-(2.2.6), where  $\{\kappa, e\}$  are positive constants,  $\nu(\cdot)$  and  $\lambda(\cdot)$  satisfy (2.2.11) and (2.2.12). Then the following properties are reached :*

$$(2.2.41) \quad \int_0^{+\infty} \nu(t) (\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) dt < \infty,$$

$$(2.2.42) \quad \int_0^{+\infty} \nu(t) \|\dot{x}(t)\|^2 dt < \infty, \quad \int_0^{+\infty} \nu(t) \|y(t) - x(t)\|^2 dt < \infty,$$

$$(2.2.43) \quad \int_0^{+\infty} \epsilon \lambda(t) \nu(t) \|A_{\lambda(t)} x(t)\|^2 dt < \infty, \quad \int_0^{+\infty} \nu^2(t) \|\nabla \Phi(x(t))\|^2 dt < \infty,$$

$$(2.2.44) \quad \Phi(x(t)) - \inf_{\mathcal{H}} \Phi = o(\nu^{-2}(t)),$$

$$(2.2.45) \quad \|y(t) - x(t)\| = o(\nu^{-1}(t)), \quad \|\dot{x}(t) + \nabla \Phi(x(t)) + \epsilon A_{\lambda(t)} x(t)\| = o(\nu^{-1}(t));$$

If, in addition, (2.2.13)-(2.2.15) are satisfied by  $\nu(\cdot)$ , then :

$$(2.2.46) \quad \|\dot{x}(t)\| = o(\nu^{-1}(t)),$$

$$(2.2.47) \quad \lim_{t \rightarrow +\infty} \epsilon \lambda(t) \|A_{\lambda(t)} x(t)\| = 0,$$

$$(2.2.48) \quad \exists \bar{x} \in S, \quad \text{s.t. } x(\cdot) \rightharpoonup \bar{x} \text{ weakly in } \mathcal{H}.$$

Theorem 2.2.1 will be proved in Section 2.2.5.

The next result is a direct consequence of Theorem 2.2.1 relative to the special case when  $\nu(t) = Mt + \nu_0$  (for some positive constants  $M$  and  $\nu_0$ ).

**Theorem 2.2.2.** *Let  $\Phi$  and  $A$  verify (H1)-(H3), let  $\{\kappa, e, \nu_0, M\}$  be positive constants such that  $\nu_0 > M\kappa^{-1}$  and  $e > \frac{2M\nu_0}{\kappa\nu_0 - M}$  and assume that  $(x, y) : [0, +\infty[ \rightarrow \mathcal{H}^2$  is a global classical solution to (2.2.5), along with  $\theta(t) = \frac{\kappa Mt + (\kappa\nu_0 - M)}{Mt + \nu_0 + e}$  and  $\lambda(t) = \lambda_c \frac{(Mt + \nu_0 + e)^3}{(\kappa Mt + (\kappa\nu_0 - M))}$  for some constant  $\lambda_c > \inf_{0 < s < \frac{2M\nu_0}{\kappa\nu_0 - M}} \frac{1}{s(e-s)}$ . Then there exists  $\bar{x} \in S$  such that  $x(\cdot) \rightharpoonup \bar{x}$  weakly in  $\mathcal{H}$ , and the following properties are reached :*

$$(2.2.49) \quad \int_0^{+\infty} t(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) dt < \infty, \quad \Phi(x(t)) - \inf_{\mathcal{H}} \Phi = o(t^{-2}),$$

$$(2.2.50) \quad \int_0^{+\infty} t\|\dot{x}(t)\|^2 dt < \infty, \quad \|\dot{x}(t)\| = o(t^{-1}),$$

$$(2.2.51) \quad \int_0^{+\infty} t\|y(t) - x(t)\|^2 dt < \infty, \quad \|y(t) - x(t)\| = o(t^{-1}),$$

$$(2.2.52) \quad \int_0^{+\infty} \frac{\epsilon}{t} (\lambda(t) \|A_{\lambda(t)}x(t)\|)^2 dt < \infty,$$

$$(2.2.53) \quad \int_0^{+\infty} t^2 \|\nabla\Phi(x(t))\|^2 dt < \infty,$$

$$(2.2.54) \quad \|\nabla\Phi(x(t)) + \epsilon A_{\lambda(t)}x(t)\| = o(t^{-1}),$$

$$(2.2.55) \quad \lim_{t \rightarrow +\infty} \epsilon \lambda(t) \|A_{\lambda(t)}x(t)\| = 0.$$

**Proof.** It can be checked in this setting that conditions (2.2.11) holds whenever  $\nu_0 > M\kappa^{-1}$ . Conditions (2.2.13)-(2.2.15) are also obviously satisfied, while (2.2.12a) reduces to  $e > \frac{2M\nu_0}{\kappa\nu_0 - M}$ . Theorem 2.2.2 is then directly deduced from Theorem 2.2.1.

■

### 2.2.4.3 Two particular cases

There we specialize our convergence results to the two systems (2.2.7) and (2.2.8) separately.

**Theorem 2.2.3.** *Let  $\Phi : \mathcal{H} \rightarrow \mathbb{R}$  be a convex differentiable function whose gradient  $\nabla\Phi$  is Lipschitz continuous on bounded sets and verifies  $S_{\nabla\Phi} := (\nabla\Phi)^{-1}(0) \neq \emptyset$ . Let  $\{\kappa, e, \nu_0, M\}$  be positive constants such that  $\nu_0 > M\kappa^{-1}$ , and let  $(x, y) : [0, +\infty[ \rightarrow \mathcal{H}^2$  be a global classical solution to (2.2.7), along with  $\theta(t) = \frac{\kappa Mt + (\kappa\nu_0 - M)}{Mt + \nu_0 + e}$ . Then there exists  $\bar{x} \in S_{\nabla\Phi}$  such that  $x(\cdot) \rightharpoonup \bar{x}$  weakly in  $\mathcal{H}$ , and we have :*

$$(2.2.56) \quad \int_0^{+\infty} t(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) dt < \infty, \quad \Phi(x(t)) - \inf_{\mathcal{H}} \Phi = o(t^{-2}),$$

$$(2.2.57) \quad \int_0^{+\infty} t\|\dot{x}(t)\|^2 dt < \infty, \quad \|\dot{x}(t)\| = o(t^{-1}),$$

$$(2.2.58) \quad \int_0^{+\infty} t\|y(t) - x(t)\|^2 dt < \infty, \quad \|y(t) - x(t)\| = o(t^{-1}),$$

$$(2.2.59) \quad \int_0^{+\infty} t^2\|\nabla\Phi(x(t))\|^2 dt < \infty, \quad \|\nabla\Phi(x(t))\| = o(t^{-1}).$$

**Proof.** Theorem 2.2.3 is immediately deduced from Theorem 2.2.2 with  $\epsilon = 0$ . ■

**Theorem 2.2.4.** *Let  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  be a maximally monotone operator such that  $S_A := A^{-1}(0) \neq \emptyset$ , let  $\{\kappa, e, \nu_0, M\}$  be positive constants such that  $\nu_0 > M\kappa^{-1}$  and  $e > \frac{2M\nu_0}{\kappa\nu_0 - M}$ , and assume that  $(x, y) : [0, +\infty[ \rightarrow \mathcal{H}^2$  is a global classical solution to (2.2.8), along with  $\theta(t) = \frac{\kappa Mt + (\kappa\nu_0 - M)}{Mt + \nu_0 + e}$  and  $\lambda(t) = \lambda_c \frac{(Mt + \nu_0 + e)^3}{(\kappa Mt + (\kappa\nu_0 - M))}$  for some constant  $\lambda_c$  verifying  $\lambda_c > \inf_{0 < s < \frac{2M\nu_0}{\kappa\nu_0 - M}} \frac{1}{s(e-s)}$ . Then there exists  $\bar{x} \in S_A$  such that  $x(\cdot) \rightharpoonup \bar{x}$  weakly in  $\mathcal{H}$ , and we have :*

$$(2.2.60) \quad \int_0^{+\infty} t\|\dot{x}(t)\|^2 dt < \infty, \quad \|\dot{x}(t)\| = o(t^{-1}),$$

$$(2.2.61) \quad \int_0^{+\infty} t\|y(t) - x(t)\|^2 dt < \infty, \quad \|y(t) - x(t)\| = o(t^{-1}),$$

$$(2.2.62) \quad \int_0^{+\infty} \frac{1}{t} (\lambda(t) \|A_{\lambda(t)}x(t)\|)^2 dt < \infty,$$

$$(2.2.63) \quad \lim_{t \rightarrow +\infty} \lambda(t) \|A_{\lambda(t)}x(t)\| = 0.$$

**Proof.** Theorem 2.2.4 follows from Theorem 2.2.2 with  $\epsilon = 1$  and  $\Phi = 0$ . ■

Observe that the Newton-types convergence results obtained for (2.2.2) and (2.2.3) by Attouch-Peypouquet-Redont [15] and Attouch-László [11], respectively, are also valid for the proposed systems (2.2.7) and (2.2.8).

#### 2.2.4.4 Some related results

Let us underline that Theorem 2.2.1 gives us sufficient conditions on parameters that ensure the convergence of the trajectories of the first-order system (2.2.5)-(2.2.6) towards an element of  $\nabla\Phi^{-1}(0) \cap (\epsilon A)^{-1}(0)$ , that is a common solution of a convex

minimization problem and a monotone inclusion, where  $\Phi$  is smooth  $C^1$  convex function with locally Lipschitz continuous gradient and where  $A$  is a maximally monotone operator. So far, most of the research efforts in the literature have been devoted to solving these two kinds of problems separately.

As special instances of Theorem 2.2.1 we cover both the particular issues. These results (see Theorems 2.2.3 and 2.2.4) allow us to reach the same convergence rates that were obtained relative to the second-order Newton-like systems (2.2.2) and (2.2.3), by Attouch-Peypouquet-Redont [15] (for minimizing a  $C^2$  convex function) and Attouch-László [11] (for an arbitrary monotone inclusion), respectively. Recall that Attouch-László [12] also considered a discrete variant of (2.2.3) of proximal-type and showed that its iterates inherit the good features of the continuous model.

Other recent developments on minimizing a  $C^2$  convex function  $\Phi$  deal with dynamical systems combining geometric damping with temporal scaling. Let us mention, for instance, the approaches by Attouch-Balhag-Chbani-Riahi [17], Attouch-Chbani-Fadili-Riahi [18] and Shi-Du-Jordan-Su [103] which enter the second-order system

$$(2.2.64) \quad \ddot{x}(t) + \gamma(t)\dot{x}(t) + \beta(t)\nabla^2\Phi(x(t))\dot{x}(t) + b(t)\nabla\Phi(x(t)) = 0,$$

where the coefficients  $\gamma, \beta : [t_0, +\infty[ \rightarrow \mathbb{R}_+$  (for some  $t_0 > 0$ ) take account of the viscous and Hessian-driven damping, respectively, and where  $b : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is a time scale parameter. Note that (2.2.64) reduces to (2.2.2) when  $\beta(t) = \beta$  (for some constant  $\beta > 0$ ) and  $b(t) = 1$ . Contrary to Theorem 2.2.1, the studies in [17, 18, 103] do not prove the convergence of the trajectories, and no estimates are given regarding the velocity. Only convergence rates of the gradient and the function values are estimated, depending on the involved parameters. It turns out for bounded temporal scaling that the optimal estimates reached by the authors are  $\int_{t_0}^{+\infty} t^2 \|\nabla\Phi(x(t))\|^2 dt < \infty$  (fast decay of the gradient) and  $\Phi(x(t)) - \inf_{\mathcal{H}} \Phi = \mathcal{O}(t^{-2})$  (for the function values), which are obtained as a particular case of Theorem 2.2.1, with an even better rate  $\Phi(x(t)) - \inf_{\mathcal{H}} \Phi = o(t^{-2})$ . These optimal properties were found out in [103, Theorem 5 and Corollary 4.2], in Euclidean spaces, for (2.2.64) when  $\gamma(t) = 3t^{-1}$ ,  $\beta(t) = \beta$  (for some  $\beta > 0$ ) and  $b(t) = 1 + 3\beta t^{-1}$ , relative to smooth  $C^2$  convex and Lipschitz continuous functions  $\Phi$ . The same properties were stated in [18, Theorem 2.3], in Hilbert spaces, regarding (2.2.64) when  $\gamma(t) = \alpha t^{-1}$

(for some  $\alpha \geq 3$ ),  $\beta(t) = \beta$  (with  $\beta > 0$ ) and  $b(t) = 1 + \beta t^{-1}$ , relative to smooth  $C^2$  convex functions  $\Phi$  with locally Lipschitz continuous gradient. Other significant accelerations are proposed in [18, 17], based upon reparametrization in time of non-autonomous dissipative dynamical systems. This comes into play when using unbounded temporal scaling  $b(\cdot)$ . An major concern with this latter approach is to lead to discrete counterparts of proximal-types with unbounded proximal indexes. Indeed the resulting algorithms appear to be somewhat difficult to adapt to the context of structured minimization problems.

## 2.2.5 Proof of Theorem 2.2.1

Suppose that  $(x, y)$  is a global classical solution to (2.2.5)-(2.2.6) under conditions (H1)-(H3), (2.2.11) and (2.2.12), and introduce the following notations :  $v = e + \nu$ ,  $\mu = \frac{v^2}{\rho}$  and  $k_{s_c} = \frac{\mu}{2s_c \lambda v}$ , where  $\rho$  and  $s_c$  are given in Proposition 2.2.3, and set  $e_c = \frac{2M}{\rho(0)}$ .

### 2.2.5.1 Proof of items (2.2.41)-(2.2.45)

This first part of the proof can be divided into the following steps (B1)-(B3) :

(B1) Let us prove (2.2.41)-(2.2.42) and other useful estimates. As an immediate consequence of Proposition 2.2.3 we establish the following results :

$$(2.2.65a) \quad \Phi(x(t)) - \inf_{\mathcal{H}} \Phi \leq \frac{\mathcal{E}_{s_c, q}(0)\kappa}{v^2(t)},$$

$$(2.2.65b) \quad \frac{s_c(e-s_c)}{2} \|x - q\|^2 \leq \mathcal{E}_{s_c, q}(0),$$

$$(2.2.65c) \quad \int_0^{+\infty} v(t)(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) dt \leq \frac{\mathcal{E}_{s_c, q}(0)}{s_c - e_c},$$

$$(2.2.65d) \quad \int_0^{+\infty} \epsilon \lambda(t) v(t) \left\| A_{\lambda(t)} x(t) + k_{s_c}(t) \dot{x}(t) \right\|^2 dt < \frac{\mathcal{E}_{s_c, q}(0)}{s_c},$$

$$(2.2.65e) \quad \int_0^{+\infty} \nu(t) \|\dot{x}(t)\|^2 dt \leq \frac{4\kappa \mathcal{E}_{s_c, q}(0)}{(4-\epsilon\delta)(e-s_c)},$$

$$(2.2.65f) \quad \int_0^{+\infty} \nu^2(t) \left\| \theta(t) u(t) + \left(1 - \frac{e-s_c}{2v(t)}\right) \dot{x}(t) \right\|^2 dt \leq \frac{\mathcal{E}_{s_c, q}(0)\kappa^2}{\rho(0)},$$

$$(2.2.65g) \quad \int_0^{+\infty} \nu(t) \|u(t)\|^2 dt \leq \frac{2\kappa \mathcal{E}_{s_c, q}(0)}{\theta^2(0)} \left( \frac{4}{(4-\delta)(e-s_c)} + \frac{\kappa}{\rho(0)} \right).$$

Note that estimates (2.2.41)-(2.2.42) follow directly from (2.2.65c), (2.2.65e) and (2.2.65g).

Let  $(s, q) \in [0, e] \times S$ . By the definition of  $\mathcal{E}_{s, q}$  in (2.2.10), we have  $\mu \tilde{\Phi}(x) \leq \mathcal{E}_{s, q}$ , where  $\mu := \frac{v^2}{\rho}$  and  $\tilde{\Phi} := \Phi - \Phi(q)$ , as well as  $\frac{s_c(e-s)}{2} \|x - q\|^2 \leq \mathcal{E}_{s, q}$ . Moreover,

we know that  $\mathcal{E}_{s_c, q}$  is non-increasing (by Proposition 2.2.3) and, as  $\rho \leq \kappa$  (from Remark 2.2.2), we also have  $\mu \geq \frac{v^2}{\kappa}$ . Then it is deduced that  $\frac{v^2}{\kappa} \tilde{\Phi} \leq \mathcal{E}_{s_c, q}(0)$  and that  $\frac{s_c(e-s_c)}{2} \|x - q\|^2 \leq \mathcal{E}_{s_c, q}(0)$ , which leads to (2.2.65a) and (2.2.65b). Furthermore, by integrating the inequality (with  $s = s_c$ ) in Proposition 2.2.3 and observing that  $\mathcal{E}_{s_c, q}$  is non-negative, we have (for  $t \geq 0$ )

$$(2.2.66) \quad \begin{aligned} & (s_c - e_c) \int_0^t v(r) \tilde{\Phi}(x(r)) \, dr \\ & + s_c \int_0^t \epsilon \lambda(r) v(r) \|A_{\lambda(r)} x(r) + k_{s_c}(r) \dot{x}(r)\|^2 \, dr \\ & + \int_0^t \nu^2(r) \mathcal{T}_{s_c}(r) \, dr \leq \mathcal{E}_{s_c, q}(0) - \mathcal{E}_{s_c, q}(t) \leq \mathcal{E}_{s_c, q}(0), \end{aligned}$$

with  $k_{s_c} = \frac{v}{2s_c \lambda \rho}$  and  $\mathcal{T}_{s_c} = \rho \|u + \theta^{-1} g \dot{x}\|^2 + \frac{\epsilon}{4\rho} \left( \frac{e-s_c}{\nu} \right) \left( \frac{s_c+3e}{\nu} + 4 - \epsilon \delta \right) \|\dot{x}\|^2$ , where  $g(t) = 1 - \frac{e-s_c}{2v(t)}$ . So, recalling that  $s_c - e_c > 0$  and that  $\tilde{\Phi}$  and  $\mathcal{T}_{s_c}$  are non-negative mappings, by passing to the limit as  $t \rightarrow +\infty$  in the previous inequality yields  $\int_0^{+\infty} v(r) \tilde{\Phi}(x(r)) \, dr < \infty$  and  $\int_0^{+\infty} \epsilon \lambda(r) v(r) \|A_{\lambda(r)} x(r) + k_{s_c}(r) \dot{x}(r)\|^2 \, dr < \infty$ , namely (2.2.65c) and (2.2.65d). It is also immediate from (2.2.66) that

$$(2.2.67) \quad \int_0^{+\infty} \nu^2(r) \mathcal{T}_{s_c}(r) \, dr \leq \mathcal{E}_{s_c, q}(0).$$

On the one hand, by definition of  $\mathcal{T}_{s_c}$  and by  $\rho \leq \kappa$  (from Remark 2.2.2), we have

$$\nu^2 \mathcal{T}_{s_c} \geq \frac{(e-s_c)(4-\epsilon\delta)}{4\kappa} \nu \|\dot{x}\|^2,$$

which, in light of (2.2.67) and recalling that  $s_c \in (e_c, e)$  and that  $\epsilon\delta \in [0, 4)$ , entails  $\int_0^{+\infty} \nu(r) \|\dot{x}(r)\|^2 \, dr \leq \frac{4\kappa}{(e-s_c)(4-\epsilon\delta)} \mathcal{E}_{s_c, q}(0)$ , that is (2.2.65e). On the other hand, again using the above definition of  $\mathcal{T}_{s_c}$  while reminding that  $\rho(0) \leq \rho$  and  $\theta \leq \kappa$  (from Remark 2.2.2), we readily get

$$\nu^2 \mathcal{T}_{s_c} \geq \frac{\rho(0)}{\kappa^2} \nu^2 \|\theta u + g \dot{x}\|^2.$$

This, by (2.2.67), gives us  $\int_0^{+\infty} \nu^2(t) \|\theta(t)u(t) + g(t)\dot{x}(t)\|^2 \, dt \leq \frac{\kappa^2 \mathcal{E}_{s_c, q}(0)}{\rho(0)}$ , that is (2.2.65f). Let us prove (2.2.65g). It is not difficult to check that  $0 \leq g \leq 1$  as  $e - s_c > 0$ , while we know that  $\theta$  is positive and that  $\theta \geq \theta(0)$  (as  $\theta$  is non-decreasing). So, by convexity of the square norm and by  $\nu \|u\|^2 = \frac{4\nu}{\theta^2} \|\frac{1}{2}(\theta u + g \dot{x}) - \frac{1}{2}g \dot{x}\|^2$  we obtain

$$\nu \|u\|^2 \leq \frac{2\nu}{\theta^2} (\|\theta u + g \dot{x}\|^2 + g^2 \|\dot{x}\|^2) \leq \frac{2}{\theta^2(0)} (\nu \|\theta + g \dot{x}\|^2 + \nu \|\dot{x}\|^2).$$

Combining this last inequality with (2.2.65e) and (2.2.65f) entails (2.2.65g).

(B2) Let us prove (2.2.43) given by :

$$(2.2.68a) \quad \int_0^{+\infty} \epsilon \lambda(t) v(t) \|A_{\lambda(t)} x(t)\|^2 dt < \infty,$$

$$(2.2.68b) \quad \int_0^{+\infty} \nu^2(t) \|\nabla \Phi(x(t))\|^2 dt < \infty.$$

**Proof.** Using the notations of Proposition 2.2.3, we get

$$\lambda v \|A_{\lambda} x\|^2 = 4\lambda v \left\| \frac{1}{2} (A_{\lambda} x + k_{s_c} \dot{x}) - \frac{1}{2} k_{s_c} \dot{x} \right\|^2,$$

where  $v = e + \nu$  and where  $\lambda$  and  $k_{s_c}$  are positive quantities. Then, by convexity of the square norm, we deduce that

$$(2.2.69) \quad \lambda v \|A_{\lambda} x\|^2 \leq 2\lambda v \|A_{\lambda} x + k_{s_c} \dot{x}\|^2 + 2\lambda v k_{s_c}^2 \|\dot{x}\|^2.$$

From  $k_{s_c} = \frac{\mu}{2s_c \lambda v}$ , we have  $k_{s_c}^2 \lambda v = \left( \frac{\mu}{2s_c \lambda v} \right)^2 \lambda v = \frac{\mu^2}{4s_c^2 \lambda v}$ , while Remark 2.2.4 tells us that  $\lambda = \frac{\mu v}{\nu s_c C_{\delta}}$ , where  $C_{\delta} = (e - s_c)\delta$  for some constant  $\delta > 0$ . Then a simple computation leads us to  $k_{s_c}^2 \lambda v = \frac{\mu^2}{4s_c^2 v} \frac{\nu s_c C_{\delta}}{\mu v} = \frac{\mu \nu C_{\delta}}{4s_c v^2}$ , hence, as  $\mu = \frac{v^2}{\rho}$  (again from Remark 2.2.4) and as  $\rho$  is non-decreasing, we obtain  $k_{s_c}^2 \lambda v = \frac{\nu C_{\delta}}{4s_c \rho} \leq \frac{C_{\delta}}{4s_c \rho(0)} \nu$ . So by  $\int_0^{+\infty} \nu(t) \|\dot{x}(t)\|^2 dt < \infty$  (from (2.2.65e)), we get  $\int_0^{+\infty} \lambda(t) v(t) k_{s_c}^2(t) \|\dot{x}(t)\|^2 dt < \infty$ , while we know that  $\int_0^{+\infty} \epsilon \lambda(t) v(t) \|A_{\lambda(t)} x(t) + k_{s_c}(t) \dot{x}(t)\|^2 dt < \infty$  (from (2.2.65d)). Hence, by these results and (2.2.69), we get  $\int_0^{+\infty} \epsilon \lambda(t) v(t) \|A_{\lambda(t)} x(t)\|^2 dt < \infty$ , that is (2.2.68a). Now, we simply have

$$(2.2.70) \quad \nu^2 \|\nabla \Phi(x)\|^2 \leq 2\nu^2 \|\nabla \Phi(x) + \epsilon A_{\lambda} x\|^2 + 2\epsilon^2 \nu^2 \|A_{\lambda} x\|^2.$$

As  $\nabla \Phi(x) + \epsilon A_{\lambda} x = -\theta u - \dot{x}$  (from (2.2.5)), we obviously get  $\|\nabla \Phi(x) + \epsilon A_{\lambda} x\| = \|\theta u + \dot{x}\|$ . Consequently, setting  $g = 1 - \frac{e-s_c}{2v}$ , by the convexity of the square norm we successively obtain

$$\begin{aligned} \nu^2 \|\nabla \Phi(x) + \epsilon A_{\lambda} x\|^2 &= 4\nu^2 \left\| \frac{1}{2} (\theta u + g\dot{x}) + \frac{1}{2} \frac{e-s_c}{2v} \dot{x} \right\|^2 \\ &\leq 2\nu^2 \|\theta u + g\dot{x}\|^2 + \frac{(e-s_c)^2 \nu^2}{2v^2} \|\dot{x}\|^2 \\ &\leq 2\nu^2 \|\theta u + g\dot{x}\|^2 + \frac{(e-s_c)^2}{2} \|\dot{x}\|^2 \quad (\text{since } v = \nu + e). \end{aligned}$$

So, by  $\int_0^{+\infty} \nu^2(t) \|\theta(t)u(t) + g(t)\dot{x}(t)\|^2 < \infty$  (from (2.2.65f)) and  $\int_0^{+\infty} \nu(t) \|\dot{x}(t)\|^2 dt < \infty$  (from (2.2.65e)), we immediately deduce  $\int_0^{+\infty} \nu^2(t) \|\nabla \Phi(x(t)) + \epsilon A_{\lambda(t)} x(t)\|^2 dt < \infty$ . The last result (2.2.68b) is immediate from (2.2.70) in view of (2.2.68a) and the latter estimate.

**(B3)** Let us prove (2.2.44)-(2.2.45). In this regard, we prove the following results :

$$(2.2.71a) \quad \lambda(t)v(t) \sim \left(\frac{\lambda_c}{\kappa}\right) v^3(t) \text{ (for some positive constant } \lambda_c),$$

$$(2.2.71b) \quad \int_0^{+\infty} v^2(t) |\langle A_{\lambda(t)}x(t), \dot{x}(t) \rangle| dt < \infty,$$

$$(2.2.71c) \quad \lim_{t \rightarrow +\infty} \nu^2(t)(\Phi(x(t)) - \Phi(q)) = 0,$$

$$(2.2.71d) \quad \lim_{t \rightarrow +\infty} \nu(t)\|u(t)\| = 0,$$

$$(2.2.71e) \quad \lim_{t \rightarrow +\infty} \nu(t)\|\dot{x}(t) + \nabla\Phi(x(t)) + \epsilon A_{\lambda(t)}x(t)\| = 0.$$

Observe that (2.2.44)-(2.2.45) follow from (2.2.71c)-(2.2.71e).

The first item (2.2.71a) is immediate from  $\lambda(t) \sim \left(\frac{\lambda_c}{\kappa}\right)\nu^2(t)$  (according to (2.2.25a)) and from  $\nu(t) \sim v(t)$  (by  $v = \nu + e$ ), as  $t \rightarrow +\infty$ . Next, using Lemma 2.2.2 with  $s = 0$  we get (for  $t \geq 0$ )

$$(2.2.72) \quad \dot{\mathcal{E}}_{0,q}(t) \leq e_c v(t) \tilde{\Phi}(x(t)) - \frac{v^2(t)}{\rho(t)} \langle A_{\lambda(t)}x(t), \dot{x}(t) \rangle,$$

where  $\tilde{\Phi} = \Phi - \Phi(q)$ . Moreover, using the Holder inequality and recalling that  $\rho(t) \geq \rho(0)$  (for  $t \geq 0$ ), we get

$$\begin{aligned} & \int_0^{+\infty} \left| \frac{v^2(t)}{\rho(t)} \langle A_{\lambda(t)}x(t), \dot{x}(t) \rangle \right| dt \\ & \leq \frac{1}{\rho(0)} \left( \int_0^{+\infty} v^3(t) \|A_{\lambda(t)}x(t)\|^2 dt \right)^{\frac{1}{2}} \cdot \left( \int_0^{+\infty} v(t) \|\dot{x}(t)\|^2 dt \right)^{\frac{1}{2}}, \end{aligned}$$

as  $\int_0^{+\infty} v(t) \|\dot{x}(t)\|^2 dt < \infty$  (by (2.2.65e)) and  $\int_0^{+\infty} v^3(t) \|A_{\lambda(t)}x(t)\|^2 dt < \infty$  (by (2.2.68a) and (2.2.71a)). This trivially yields (2.2.71b).

We also know that  $\int_0^{+\infty} v(t) \tilde{\Phi}(x(t)) dt < \infty$  (from (2.2.65c)). Then by these last two results, in light of (2.2.72), we classically deduce that  $\lim_{t \rightarrow +\infty} \mathcal{E}_{0,q}(t)$  exists. It can also be checked that  $\mathcal{E}_{0,q}$  verifies

$$\frac{1}{\nu(t)} \mathcal{E}_{0,q}(t) = \frac{1}{2} \nu(t) \|u(t)\|^2 + \frac{v^2(t)}{\nu(t)\rho(t)} \tilde{\Phi}(x(t)),$$

Furthermore, as  $t \rightarrow \infty$ , we simply have  $\frac{v^2(t)}{\nu(t)\rho(t)} \sim \left(\frac{1}{\kappa}\right)v(t)$ , because  $\nu(t) \sim v(t)$  and  $\rho(t) \rightarrow \kappa$  (in light of Remark 2.2.2-(i2)). Therefore, by  $\int_0^{+\infty} v(t) \tilde{\Phi}(x(t)) dt < \infty$  (from (2.2.65c)) and  $\int_0^{+\infty} \nu(t) \|u(t)\|^2 dt < +\infty$  (from (2.2.65g)), we get  $\int_0^{+\infty} \frac{1}{\nu(t)} \mathcal{E}_{0,q}(t) dt < \infty$ . Whence, as  $\lim_{t \rightarrow +\infty} \mathcal{E}_{0,q}(t)$  exists and  $\int_0^{+\infty} \frac{1}{\nu(t)} dt = \infty$  (from Remark 2.2.2), we deduce that  $\lim_{t \rightarrow +\infty} \mathcal{E}_{0,q}(t) = 0$ . This in view of the definition of  $\mathcal{E}_{0,q}$  entails (2.2.71c) and (2.2.71d). It remains to prove (2.2.71e). From (2.2.5), we simply have  $\|\dot{x} +$

$\nabla\Phi(x) + \epsilon A_\lambda x \| = \|\theta u\|$ . It follows from  $\theta \in (0, \kappa)$  (see Remark 2.2.2) that

$$\nu \|u\| \geq \frac{\nu}{\kappa} \|\theta u\| = \frac{\nu}{\kappa} \|\dot{x} + \nabla\Phi(x) + \epsilon A_\lambda x\|.$$

Hence, (2.2.71e) is straightforwardly deduced from (2.2.71d).

### 2.2.5.2 Proof of items (2.2.46)-(2.2.48)

In this second part of the proof we assume that  $\nu(\cdot)$  satisfies the additional conditions (2.2.13)-(2.2.15) and we divide the rest of the proof into the following steps (C1)-(C3) :

(C1) (Estimate (2.2.46)). At once we prove that

$$(2.2.73) \quad \lim_{t \rightarrow +\infty} \nu(t) \|\dot{x}(t)\| = 0.$$

For simplification purposes, given  $h > 0$ ,  $t \geq 0$  and  $f : [0, +\infty[ \rightarrow \mathcal{H}$ , we denote  $f_h(t) = f(t+h)$  and  $\Delta_h f(t) = f(t+h) - f(t)$  (hence  $\Delta_h f = f_h - f$ ). Let  $t_0 > 0$ , and let a pair  $(x, y)$  verify (2.2.5) on  $[0, \infty)$ . Then for  $t > t_0$  and  $h > 0$ , we have

$$\dot{x}(t) + (\theta u)(t) + \Psi(t, x(t)) = 0, \text{ where } u = y - x \text{ and } \Psi(t, x) = \nabla\Phi(x) + \epsilon A_{\lambda(t)}x.$$

Thus omitting the variable  $t$  we obtain

$$(2.2.74) \quad \Delta_h \dot{x} + \Delta_h(\theta u) + \Delta_h(\Psi(\cdot, x)) = 0,$$

It is also simply checked from an easy computation that

$$\Delta_h(\theta u) = u_h \Delta_h \theta + \theta \Delta_h u = (y_h - x_h) \Delta_h \theta + \theta(\Delta_h y - \Delta_h x).$$

Then (2.2.74) can be rewritten as

$$\Delta_h \dot{x} + \Delta_h \theta(y_h - x_h) + \theta(\Delta_h y - \Delta_h x) + \Delta_h \Psi(\cdot, x) = 0.$$

Furthermore, as  $\sup_{t \geq 0} \|\dot{v}(t)\| \leq M$  (from (2.2.11c)), we readily have

$$\frac{1}{2} \nu^2 \frac{d}{dt} \|\Delta_h x\|^2 \geq \frac{1}{2} \frac{d}{dt} (\nu^2 \|\Delta_h x\|^2) - M \nu \|\Delta_h x\|^2.$$

Hence, in light of this last estimate, by taking the scalar product of the previous equality with  $\nu^2 \Delta_h x$  and noticing that  $\langle \Delta_h \dot{x}, \Delta_h x \rangle = \frac{1}{2} \frac{d}{dt} \|\Delta_h x\|^2$ , we obtain

$$(2.2.75) \quad \begin{aligned} & \frac{1}{2} \frac{d}{dt} (\nu^2 \|\Delta_h x\|^2) - M \nu \|\Delta_h x\|^2 + \nu^2 \langle \Delta_h \theta(y_h - x_h), \Delta_h x \rangle \\ & + \nu^2 \theta \langle \Delta_h y - \Delta_h x, \Delta_h x \rangle + \nu^2 \langle \Delta_h \Psi(\cdot, x), \Delta_h x \rangle \leq 0. \end{aligned}$$

Moreover, by  $\Psi(\cdot, x(\cdot)) = \nabla\Phi + \epsilon A_\lambda(x)$  and thanks to the monotonicity of  $\nabla\Phi$ , we get

$$\langle \Delta_h \Psi(\cdot, x), \Delta_h x \rangle \geq \epsilon \langle A_{\lambda_h}(x_h) - A_\lambda(x), \Delta_h x \rangle,$$

so, by monotonicity of  $A_{\lambda_h}$ , we are led to

$$\langle \Delta_h \Psi(\cdot, x), \Delta_h x \rangle \geq \epsilon \langle A_{\lambda_h}(x) - A_\lambda(x), \Delta_h x \rangle.$$

Let  $z \in (\epsilon A)^{-1}(0)$ . From (2.2.96) of Lemma A.1, and remembering that  $\lambda$  is non-decreasing on  $[t_0, \infty)$  for  $t_0$  large enough (by Remark 2.2.3), we also have, on  $[t_0, \infty)$ ,

$$\|\epsilon A_{\lambda_h}(x) - \epsilon A_\lambda(x)\| \leq \frac{3\epsilon}{\lambda^2} \|x - z\| \times |\Delta_h \lambda|.$$

The previous two results ensure that

$$\langle \Delta_h \Psi(\cdot, x), \Delta_h x \rangle \geq -\frac{3\epsilon}{\lambda^2} \|x - z\| \times |\Delta_h \lambda| \times \|\Delta_h x\|.$$

Thus, by (2.2.75) we infer that

$$(2.2.76) \quad \begin{aligned} & \frac{1}{2} \frac{d}{dt} (\nu^2 \|\Delta_h x\|^2) - M \nu \|\Delta_h x\|^2 + \nu^2 \langle \Delta_h \theta(y_h - x_h), \Delta_h x \rangle \\ & + \nu^2 \theta \langle \Delta_h y - \Delta_h x, \Delta_h x \rangle - \nu^2 \frac{3\epsilon}{\lambda^2} \|x - z\| \times |\Delta_h \lambda| \times \|\Delta_h x\| \leq 0. \end{aligned}$$

So, integrating (2.2.76) on  $[s, t]$  (where  $t_0 \leq s < t$ ) and dividing by  $h^2$  yields

$$(2.2.77) \quad \begin{aligned} & \frac{1}{2} \|\nu(t) \frac{\Delta_h x(t)}{h}\|^2 - \frac{1}{2} \|\nu(s) \frac{\Delta_h x(s)}{h}\|^2 - M \int_s^t \nu(r) \|\frac{\Delta_h x(r)}{h}\|^2 dr \\ & + \int_s^t \nu^2(r) \langle \frac{\Delta_h \theta(r)}{h} (y_h(r) - x_h(r)), \frac{\Delta_h x(r)}{h} \rangle dr \\ & + \int_s^t \nu^2(r) \theta(r) \langle \frac{\Delta_h y(r)}{h} - \frac{\Delta_h x(r)}{h}, \frac{\Delta_h x(r)}{h} \rangle dr \\ & - 3\epsilon \int_s^t \frac{\nu^2(r)}{\lambda^2(r)} \|x(r) - z\| \times |\frac{\Delta_h \lambda(r)}{h}| \times \|\frac{\Delta_h x(r)}{h}\| dr \leq 0. \end{aligned}$$

Now we consider the following arguments :

- Obviously, as  $\nu$  is assumed to be positive and  $C^3$  on  $[0, \infty)$ , by definition of  $\theta(\cdot)$  we observe that  $\dot{\theta}$  is of class  $C^1$ , so that  $\dot{\theta}$  is Lipschitz continuous on  $[s, t]$ .

- The condition  $(\epsilon A)^{-1}(0) \neq \emptyset$  and the property of  $(1/\lambda)$  - Lipschitz continuity of  $A_\lambda$  ensure that, for any given  $z \in (\epsilon A)^{-1}(0)$ , we have  $\epsilon \|A_{\lambda(t)} x(t)\| \leq (\epsilon/\lambda(t)) \|x(t) - z\|$ . Then, reminding that  $\nabla \Phi$  is assumed to be Lipschitz continuous on bounded subset of  $\mathcal{H}$ , and noticing that  $(x, y)$  (as a classical solution) is bounded on  $[s, t]$ , we infer from (2.2.5) that  $\dot{x}$  and  $\dot{y}$  are bounded on this interval. It follows that  $x$  and  $y$  are Lipschitz continuous on  $[s, t]$ .

- In view of Remark 2.2.1, we can see that the mapping  $t \mapsto \epsilon A_{\lambda(t)}(x(t))$  is Lipschitz continuous on  $[s, t]$ , because  $\lambda$  is assumed to be bounded away from zero. Observe that this property of Lipschitz continuity is also satisfied by the mapping  $t \mapsto \nabla \Phi(x(t))$ , because  $x$  and  $y$  are bounded on  $[s, t]$ . Therefore we deduce from (2.2.5) and the Lipschitz continuity of  $x$  and  $y$  on  $[s, t]$  that  $\dot{x}$  and  $\dot{y}$  are also Lipschitz continuous on  $[s, t]$ .

Thus, we infer that the quantities  $\frac{\Delta_h x}{h}$ ,  $\frac{\Delta_h \theta}{h}$ ,  $\frac{\Delta_h y}{h}$  converge uniformly on  $[s, t]$  to

$\dot{x}$ ,  $\dot{\theta}$  and  $\dot{y}$ , respectively, (as  $h \rightarrow 0$ ). It is also readily checked that  $\|\frac{\Delta_h x}{h}\|$  converges uniformly on  $[s, t]$  to  $\|\dot{x}\|$  (as  $\frac{\Delta_h x}{h} \rightarrow \dot{x}$  on  $[s, t]$ ), while noticing that  $\lambda$  is  $C^2$  on  $[0, \infty)$ , which guarantees that  $\frac{\Delta_h \lambda}{h}$  and  $|\frac{\Delta_h \lambda}{h}|$  converge uniformly on  $[s, t]$  to  $\dot{\lambda}$  and  $|\dot{\lambda}|$ , respectively (as  $h \rightarrow 0$ ). Hence, passing to the limit in (2.2.77), as  $h \rightarrow 0$ , amounts to

$$(2.2.78) \quad \begin{aligned} & \frac{1}{2} \|\nu(t)\dot{x}(t)\|^2 - \frac{1}{2} \|\nu(s)\dot{x}(s)\|^2 - M \int_s^t \nu(r) \|\dot{x}(r)\|^2 dr \\ & + \int_s^t \nu^2(r) \dot{\theta}(r) \langle y(r) - x(r), \dot{x}(r) \rangle dr + \int_s^t \nu^2 \theta(r) \langle \dot{y}(r) - \dot{x}(r), \dot{x}(r) \rangle dr \\ & - 3\epsilon \int_s^t \frac{\nu^2(r)}{\lambda^2(r)} |\dot{\lambda}(r)| \times \|x(r) - z\| \times \|\dot{x}(r)\| dr \leq 0. \end{aligned}$$

In addition, from  $y - x = -\kappa^{-1} \dot{y}$  and  $\dot{y} = \frac{\kappa}{\theta} (\dot{x} + \Psi(\cdot, x))$ , we readily have

$$y - x = -\frac{1}{\theta} (\dot{x} + \Psi(\cdot, x)) \quad \text{and} \quad \dot{y} - \dot{x} = \left(\frac{\kappa}{\theta} - 1\right) \dot{x} + \frac{\kappa}{\theta} \Psi(\cdot, x),$$

which implies that

$$\dot{\theta} \langle y - x, \dot{x} \rangle = -\frac{\dot{\theta}}{\theta} \|\dot{x}\|^2 - \frac{\dot{\theta}}{\theta} \langle \Psi(\cdot, x), \dot{x} \rangle,$$

and

$$\theta \langle \dot{y} - \dot{x}, \dot{x} \rangle = (\kappa - \theta) \|\dot{x}\|^2 + \kappa \langle \Psi(\cdot, x), \dot{x} \rangle.$$

From these last two equalities we get

$$\dot{\theta} \langle y - x, \dot{x} \rangle + \theta \langle \dot{y} - \dot{x}, \dot{x} \rangle = \tau \|\dot{x}\|^2 + \eta \langle \Psi(\cdot, x), \dot{x} \rangle,$$

where  $\tau(\cdot) = \kappa - \theta - \frac{\dot{\theta}}{\theta}$  and  $\eta(\cdot) = \kappa - \frac{\dot{\theta}}{\theta}$ . This in light of (2.2.78) entails

$$(2.2.79) \quad \begin{aligned} & \frac{1}{2} \|\nu(t)\dot{x}(t)\|^2 - \frac{1}{2} \|\nu(s)\dot{x}(s)\|^2 - M \int_s^t \nu(r) \|\dot{x}(r)\|^2 dr \\ & + \int_s^t \tau(r) \nu^2(r) \|\dot{x}(r)\|^2 dr + \int_s^t \eta \nu^2 \langle \Psi(r, x(r)), \dot{x}(r) \rangle dr \\ & - 3\epsilon \int_s^t \frac{\nu^2(r)}{\lambda^2(r)} |\dot{\lambda}(r)| \times \|x(r) - z\| \times \|\dot{x}(r)\| dr \leq 0. \end{aligned}$$

Let  $\Gamma(\cdot)$  be defined (for  $t \geq t_0$ ) by

$$(2.2.80) \quad \begin{aligned} \Gamma(t) & := \frac{1}{2} \|\nu(t)\dot{x}(t)\|^2 - M \int_{t_0}^t \nu(r) \|\dot{x}(r)\|^2 dr \\ & + \int_{t_0}^t \tau(r) \nu^2(r) \|\dot{x}(r)\|^2 dr + \int_{t_0}^t \eta(r) \nu^2 \langle \Psi(r, x(r)), \dot{x}(r) \rangle dr \\ & - 3 \int_{t_0}^t \epsilon \frac{\nu^2(r)}{\lambda^2(r)} |\dot{\lambda}(r)| \times \|x(r) - z\| \times \|\dot{x}(r)\| dr. \end{aligned}$$

It is clear from (2.2.79) that  $\Gamma(t) \leq \Gamma(s)$  for  $t_0 \leq s < t$ , and so  $\Gamma(\cdot)$  is non-increasing on  $[t_0, \infty)$ . In order to conclude that  $\Gamma$  is convergent we prove the convergence (as  $t \rightarrow +\infty$ ) of each of the four integrals involved in (2.2.80) through the following items **(r1)**-**(r4)** :

**(r1)** : From (2.2.65e), we directly obtain  $\int_{t_0}^{+\infty} \nu(r) \|\dot{x}(r)\|^2 dr < \infty$ .

**(r2)** : Since  $\tau(t) \sim (l + \kappa e) \frac{1}{\nu(t)}$  as  $t \rightarrow \infty$  (from Remark 2.2.3) we immediately

have  $\tau(t)\nu^2(t) \sim (l + \kappa e)\nu(t)$  (as  $t \rightarrow +\infty$ ); hence **(r1)** yields

$$\int_{t_0}^{\infty} \tau(r)\nu^2(r)\|\dot{x}(r)\|^2 dr < \infty.$$

**(r3)** : As  $\Psi(\cdot, x) := \nabla\Phi(x) + \epsilon A_{\lambda(\cdot)}x$  and recalling that  $\langle \nabla\Phi(x), \dot{x} \rangle = \frac{d}{dt} \left( \tilde{\Phi}(x) \right)$  (where  $\tilde{\Phi} = \Phi - \Phi(q)$ ), we have

$$(2.2.81) \quad \begin{aligned} & \int_{t_0}^t (\eta\nu^2)(r) \langle \Psi(r, x(r)), \dot{x}(r) \rangle dr \\ &= \int_0^t (\eta\nu^2)(r) \frac{d}{dr} \left( \tilde{\Phi}(x) \right) (r) dr + \int_0^t \epsilon (\eta\nu^2)(r) \langle A_{\lambda(r)}x(r), \dot{x}(r) \rangle dr. \end{aligned}$$

Let us evaluate separately the two quantities arising on the right-side of the previous equality. On the one hand, integrating by parts the first integral gives us

$$(2.2.82) \quad \int_{t_0}^t (\eta\nu^2)(r) \frac{d}{dr} \left( \tilde{\Phi}(x) \right) (r) dr = \eta(t)\nu^2(t)\tilde{\Phi}(x(t)) - \int_{t_0}^t \frac{d}{dr} (\eta\nu^2)(r) \tilde{\Phi}(x(r)) dr.$$

Note that, as  $t \rightarrow +\infty$ , we have  $\eta(t) \sim \kappa$  (from Remark 2.2.3) and  $\nu^2(t)\tilde{\Phi}(x(t)) \rightarrow 0$ , and so  $\eta(t)\nu^2(t)\tilde{\Phi}(x(t)) \rightarrow 0$ . In addition, by a simple computation we have  $\frac{d}{dr} (\eta\nu^2) = \dot{\eta}\nu^2 + 2\eta\nu\dot{\nu}$ , hence, by boundedness of  $\dot{\nu}$ ,  $\eta$  and  $\nu\dot{\eta}$  (from Remark 2.2.3), we infer that there exists a positive constant  $K_2$  verifying (for  $r \geq 0$ )

$$\left| \frac{d}{dr} (\eta\nu^2)(r) \tilde{\Phi}(x(r)) \right| \leq K_2 \nu(r) \tilde{\Phi}(x(r)).$$

So, in light of  $\int_0^{+\infty} \nu(r) \tilde{\Phi}(x(r)) dr < \infty$  (from (2.2.65c)), we deduce the convergence of the term  $\int_{t_0}^t \frac{d}{dr} (\eta\nu^2)(r) \tilde{\Phi}(x(r)) dr$  (as  $t \rightarrow \infty$ ). These last two results in light of (2.2.82) entail the convergence (as  $t \rightarrow \infty$ ) of first integral on the right side of (2.2.81). On the other hand, by  $\eta(t) \sim \kappa$  (as  $t \rightarrow \infty$ ) and by  $\int_{t_0}^{\infty} \epsilon \nu^2(r) |\langle A_{\lambda}x(r), \dot{x}(r) \rangle| dr < \infty$  (according to (2.2.71b)), we obviously get the convergence (as  $t \rightarrow \infty$ ) of the second term on the right side of (2.2.81). Therefore, in light of (2.2.81), we conclude that the term  $\int_{t_0}^t (\eta\nu^2)(r) \langle \Psi(r, x(r)), \dot{x}(r) \rangle dr$  is convergent (as  $t \rightarrow \infty$ ).

**(r4)** : Remark 2.2.3 tells us that  $\dot{\lambda}$  satisfies  $|\dot{\lambda}| \leq k_2\nu$  (for some constant  $k_2$ ) and (as  $t \rightarrow \infty$ ) that  $\lambda(t) \sim \left(\frac{\lambda_c}{\kappa}\right) \nu^2(t)$  (or equivalently  $\frac{\nu^3(t)}{\lambda^2(t)} \sim \left(\frac{\kappa}{\lambda_c}\right)^2 \frac{1}{\nu(t)}$ ). So,  $\frac{\nu^3}{\lambda^2} |\dot{\lambda}|$  is bounded, while we know that  $x(\cdot)$  is bounded (from (2.2.65b)), hence there exists some constant  $C_z \geq 0$  such that

$$(2.2.83) \quad \epsilon \frac{\nu^2}{\lambda^2} |\dot{\lambda}| \times \|x - z\| \times \|\dot{x}\| \leq C_z \epsilon \nu^{-1} \|\dot{x}\|.$$

In addition, we have  $\epsilon \nu^{-1} \|\dot{x}\| = \frac{\epsilon}{\nu^{3/2}} \cdot \nu^{1/2} \|\dot{x}\|$ ; integrating this equality between  $t_0$  and  $t$  and using the Holder inequality give us

$$\int_{t_0}^t \epsilon \nu^{-1}(r) \|\dot{x}(r)\| dr \leq \left( \int_{t_0}^t \frac{1}{\nu^3(r)} dr \right)^{1/2} \cdot \left( \int_{t_0}^t \nu(r) \|\dot{x}(r)\|^2 dr \right)^{1/2}.$$

This, according to  $\int_0^{+\infty} \frac{\epsilon}{\nu^3(r)} dr < \infty$  (by hypothesis) and  $\int_{t_0}^t \nu(r) \|\dot{x}(r)\|^2 dr < \infty$  (from **(r1)**) yields  $\int_{t_0}^{+\infty} \nu^{-1}(r) \|\dot{x}(r)\| dr < \infty$ . As a consequence, by (2.2.83), we get  $\int_0^{+\infty} \epsilon \frac{\nu^2(r)}{\lambda^2(r)} |\dot{\lambda}(r)| \times \|x(r) - z\| \times \|\dot{x}(r)\| dr < \infty$ .

Finally, we have shown that all the integrals involved in (2.2.80) are convergent. So, recalling that  $\Gamma$  is non-increasing on  $[t_0, \infty)$  yields the boundedness of  $\Gamma$ . Hence, it is deduced that  $\lim_{t \rightarrow +\infty} \Gamma(t)$  exists, which obviously implies that  $\nu^2(t) \|\dot{x}(t)\|^2 \rightarrow \ell$  (for some constant  $\ell \geq 0$ ) as  $t \rightarrow +\infty$ . Let us prove by contradiction that  $\ell = 0$ . Suppose that  $\ell > 0$ . This clearly implies that  $\nu(t) \|\dot{x}(t)\|^2 \sim \frac{\ell}{\nu(t)}$  (as  $t \rightarrow +\infty$ ), which, by  $\int_0^{+\infty} \nu(t) \|\dot{x}(t)\|^2 dt < \infty$  (according to (2.2.65e)) yields  $\int_0^{+\infty} \frac{1}{\nu(t)} dt < \infty$ . This contradicts Remark 2.2.2. Therefore  $\ell = 0$  and so  $\lim_{t \rightarrow +\infty} \nu(t) \|\dot{x}(t)\| = 0$ .

**(C2)** (Estimate (2.2.47)). At once we establish the following result

$$(2.2.84) \quad \lim_{t \rightarrow +\infty} \epsilon \lambda(t) A_{\lambda(t)} x(t) = 0.$$

**Proof.** Let  $q \in (\epsilon A)^{-1}(0)$ . Given  $s \geq 0$  and  $t \geq 0$ , and using Lemma A.1 with  $\gamma = \lambda(t)$ ,  $\delta = \lambda(s)$ ,  $x = x(t)$  and  $y = x(s)$ , we get

$$\|\epsilon \lambda(t) A_{\lambda(t)} x(t) - \epsilon \lambda(s) A_{\lambda(s)} x(s)\| \leq 2\epsilon \|x(t) - x(s)\| + 2\epsilon \|x(t) - q\| \frac{|\lambda(t) - \lambda(s)|}{\lambda(t)}.$$

Dividing by  $t - s$  with  $t > s$  and letting  $s$  tends to  $t$ , we obtain

$$(2.2.85) \quad \left\| \frac{d}{dt} (\epsilon \lambda(t) A_{\lambda(t)} x(t)) \right\| \leq 2\epsilon \|\dot{x}(t)\| + 2\epsilon \|x(t) - q\| \frac{|\dot{\lambda}(t)|}{\lambda(t)},$$

for almost every  $t > 0$ . Next, Remark 2.2.3 gives us  $|\dot{\lambda}| \leq k_2 \nu$  (for some constant  $k_2$ ) and (as  $t \rightarrow \infty$ )  $\lambda(t) \sim \left(\frac{\lambda_c}{\kappa}\right) \nu^2(t)$  (hence  $\frac{\dot{\lambda}(t)}{\lambda(t)} \sim \left(\frac{\kappa}{\lambda_c}\right) \dot{\lambda}(t) \nu^{-2}(t)$ ). Hence, it is readily checked that there exists some constant  $c_1 \geq 0$  such that  $\frac{|\dot{\lambda}|}{\lambda} \leq \frac{c_1}{\nu}$ . Moreover, from (2.2.73), we can see that there exists  $c_2 \geq 0$  such that  $\|\dot{x}\| \leq \frac{c_2}{\nu}$ . So, combining (2.2.85) with these last two estimates, we obtain, for some constant  $c \geq 0$ ,

$$(2.2.86) \quad \left\| \frac{d}{dt} (\epsilon \lambda(t) A_{\lambda(t)} x(t)) \right\| \leq \frac{\epsilon c}{\nu(t)} \quad \text{for a.e. } t \geq 0.$$

Note that, setting  $\omega(t) := \|\epsilon \lambda(t) A_{\lambda(t)} x(t)\|^2$ , we classically have (for a.e  $t \geq 0$ )

$$(2.2.87) \quad \left| \frac{d}{dt} \omega(t) \right| \leq 2 \left\| \frac{d}{dt} (\epsilon \lambda(t) A_{\lambda(t)} x(t)) \right\| \cdot \|\epsilon \lambda(t) A_{\lambda(t)} x(t)\|.$$

In addition, using the  $\lambda^{-1}$  Lipschitz continuity of  $A_\lambda$  we clearly have

$$\lambda(t)\|\epsilon A_{\lambda(t)}x(t)\| \leq \epsilon\|x(t) - q\|.$$

So, as  $x(\cdot)$  is bounded, we deduce that  $\|\epsilon\lambda(t)A_{\lambda(t)}x(t)\|$  is bounded. Therefore by (2.2.87) in light of (2.2.86) we deduce that there exists a positive constant (again denoted  $c$ ) such that  $|\frac{d}{dt}\omega(t)| \leq \frac{\epsilon c}{\nu(t)}$  for a.e  $t \geq 0$ . Now, as  $t \rightarrow \infty$ , again by  $\lambda(t) \sim \left(\frac{\lambda_c}{\kappa}\right)\nu^2(t)$ , and noticing that  $v(t) \sim \nu(t)$ , we obviously have  $\frac{v(t)}{\lambda(t)} \sim \left(\frac{\kappa}{\lambda_c}\right)\frac{1}{\nu(t)}$ , while  $\int_0^{+\infty} \frac{v(t)}{\lambda(t)} \cdot \|\epsilon\lambda(t)A_{\lambda(t)}x(t)\|^2 dt < \infty$  (from (2.2.68a)). It is then immediately deduced that  $\int_0^{+\infty} \omega(t) \cdot \frac{1}{\nu(t)} dt < \infty$ . So, as  $\frac{1}{\nu} \notin L^1(0, \infty)$  (from Remark 2.2.2), invoking Lemma B.2 yields the desired result.

**(C3)** (Convergence of the iterates). Let us prove that  $x(t)$  converges weakly to some element of  $S$ , as  $t \rightarrow +\infty$ .

Let  $q \in S := (\epsilon A)^{-1}(0) \cap (\nabla\Phi)^{-1}(0)$ . From Proposition 2.2.3, we know that  $\lim_{t \rightarrow +\infty} \mathcal{E}_{s_c, q}(t)$  exists. Let us recall from (2.2.10) that

$$\mathcal{E}_{s_c, q}(t) = \frac{1}{2}\|s_c(q - x(t)) + \nu(t)u(t)\|^2 + \frac{1}{2}s_c(e - s_c)\|q - x(t)\|^2 + \mu(t)\tilde{\Phi}(x(t)),$$

where  $\mu = \frac{\nu(\nu+e)}{\theta}$  (hence  $\mu = \frac{(e+\nu)^2}{\rho}$ , as  $\theta = \frac{\nu\rho}{\nu+e}$  (from (2.2.22))) and  $\tilde{\Phi} = \Phi - \Phi(q)$ . Clearly, we have  $\lim_{t \rightarrow +\infty} \nu(t)\|u(t)\| = 0$  (from (2.2.45)). Moreover, by (2.2.44) we have  $\lim_{t \rightarrow +\infty} \nu^2(t)\tilde{\Phi}(x(t)) = 0$ , thus, recalling that  $\rho$  is bounded away from zero (from condition (2.2.11b)), we obtain  $\lim_{t \rightarrow +\infty} \mu(t)\tilde{\Phi}(x(t)) = 0$  (because  $\mu(t) \sim \frac{\nu^2(t)}{\rho(t)}$ , as  $t \rightarrow +\infty$ , under condition (2.2.13)). So, in light of the previous arguments, we deduce that

$$\begin{aligned} \lim_{t \rightarrow +\infty} \mathcal{E}_{s_c, q}(t) &= \lim_{t \rightarrow +\infty} \left( \frac{1}{2}\|s_c(q - x(t))\|^2 + \frac{1}{2}s_c(e - s_c)\|q - x(t)\|^2 \right) \\ &= \lim_{t \rightarrow +\infty} \left( \frac{1}{2}s_c^2 + \frac{1}{2}s_c(e - s_c) \right) \|q - x(t)\|^2 \\ &= \frac{s_c e}{2} \lim_{t \rightarrow +\infty} \|q - x(t)\|^2. \end{aligned}$$

Hence,  $\lim_{t \rightarrow +\infty} \|x(t) - q\|$  exists. Now, let  $\bar{x}$  be a weak sequential cluster point of  $x(t)$ . Namely, there exists a sequence  $(t_n)_{n \geq 0}$  such that  $\lim_{n \rightarrow +\infty} t_n = +\infty$  and  $(x(t_n))_{n \geq 0}$  weakly converges to  $\bar{x}$  as  $n \rightarrow +\infty$ . Whence, using the lower semi-continuity of  $\Phi$  and by  $\lim_{t \rightarrow +\infty} \tilde{\Phi}(x(t)) = 0$  (thanks to (2.2.44)), we get

$$0 = \lim_{n \rightarrow +\infty} \Phi(x(t_n)) - \inf_{\mathcal{H}} \Phi \geq \Phi(\bar{x}) - \inf_{\mathcal{H}} \Phi \geq 0.$$

Thus we deduce that  $\bar{x} \in (\nabla\Phi)^{-1}(0)$ . It remains to show that  $\bar{x} \in (\epsilon A)^{-1}(0)$ . Indeed, for  $\epsilon = 0$ , then  $(\epsilon A)^{-1}(0) = \mathcal{H}$ , so that the result is obvious. Now let  $\epsilon > 0$ . Using (b) from Proposition A.1, we obtain

$$(2.2.88) \quad A_{\lambda(t_n)}(x(t_n)) \in A(x(t_n) - \lambda(t_n)A_{\lambda(t_n)}(x(t_n))).$$

Moreover, from  $\lim_{t \rightarrow +\infty} \|\lambda(t)A_{\lambda(t)}x(t)\| = 0$  (in view of (2.2.84)) we immediately infer that  $\lambda(t_n)A_{\lambda(t_n)}(x(t_n)) \rightarrow 0$  strongly as  $n \rightarrow +\infty$ . Hence, as  $\lambda(t_n) \rightarrow +\infty$  we infer that  $A_{\lambda(t_n)}(x(t_n)) \rightarrow 0$  strongly in  $\mathcal{H}$ . Therefore, passing to the limit as  $n \rightarrow +\infty$  in (2.2.88) and using the fact that  $A$  is a sequentially demi-closed operator (as it is maximally monotone), we obtain  $0 \in A(\bar{x})$ , that is  $\bar{x} \in (\epsilon A)^{-1}(0)$ . Thus, applying the well-known Opial's lemma [87] gives us the desired result. ■

## 2.2.6 Two accelerated proximal algorithms

In this section, we suggest two inertial proximal algorithms based upon implicit temporal discretizations of the systems (2.2.7) and (2.2.8). However the study of these algorithms is out of the scope of this work and will be carried out in a future research. In what follows we take  $h > 0$ , and we set  $t_k = kh$ ,  $x_k = x(t_k)$ ,  $y_k = y(t_k)$ ,  $\theta_k = \theta(t_k)$  and  $\lambda_k = \lambda(t_k)$  for all  $k \geq 0$ .

### 2.2.6.1 The convex minimization case

Consider the structured convex minimization problem

$$(2.2.89) \quad \min\{\Theta(x) := \Theta_{ns}(x) + \Theta_{sm}(x); \quad x \in \mathcal{H}\},$$

where  $\Theta_{ns}$  is a convex and proper lower semi-continuous function and  $\Theta_{sm}$  is a convex differentiable function. Given  $(x_k, y_k) \in \mathcal{H}^2$  and based upon an approximate finite difference of (2.2.7), we seek for  $(x_{k+1}, y_{k+1}) \in \mathcal{H}^2$  verifying the numerical scheme

$$\begin{cases} \frac{1}{h}(x_{k+1} - x_k) + \partial\Theta_{ns}(x_{k+1}) + \nabla\Theta_{sm}(x_k) + \theta_k(y_k - x_k) \ni 0, \\ \frac{1}{h}(y_{k+1} - y_k) + \kappa(y_k - x_k) = 0, \end{cases}$$

where  $\partial\Theta_{ns}$  denotes the Fenchel sub-differential of the (possibly non-smooth) convex function  $\Theta_{ns}$ . This can be simply rewritten as

$$\begin{cases} x_{k+1} = J_h^{\partial(\Theta_{ns})}(x_k - h\nabla(\Theta_{sm})(x_k) - h\theta_k(y_k - x_k)) \\ y_{k+1} = y_k - h\kappa(y_k - x_k), \end{cases}$$

where  $J_h^{\partial(\Theta_{ns})} = (I + h\partial(\Theta_{ns}))^{-1}$  is the resolvent operator of  $\partial(\Theta_{ns})$  with index  $h$ . Consequently, in light of Theorem 2.2.3, a good candidate for approximating (2.2.89)

is probably given by

$$(2.2.90a) \quad z_k = x_k - h\theta_k(y_k - x_k),$$

$$(2.2.90b) \quad x_{k+1} = J_h^{\partial(\Theta_{ns})}(z_k - h\nabla(\Theta_{sm})(x_k)),$$

$$(2.2.90c) \quad y_{k+1} = y_k - h\kappa(y_k - x_k).$$

Note that when  $h\kappa = 1$ , by (2.2.90c) we have  $y_{k+1} = x_k$  (for  $k \geq 0$ ), hence, for  $k \geq 1$ , (2.2.90) reduces to

$$(2.2.91a) \quad z_k = x_k + h\theta_k(x_k - x_{k-1}),$$

$$(2.2.91b) \quad x_{k+1} = J_h^{\partial(\Theta_{ns})}(z_k - h\nabla(\Theta_{sm})(x_k)).$$

This is nothing but the form of the well-known "heavy ball with friction" method (see, e.g., [86]). Other variants of (2.2.90) can be obtained by replacing  $x_k$  by  $z_k$  in the gradient of  $\Theta_{sm}$ , which leads to some Nesterov-type accelerated forward-backward algorithm (see [9, 14, 82, 83, 84]) or by swapping  $x_k$  and  $z_k$  in the operator  $J_{h\partial(\Theta_{ns})}$  (see [74]).

### 2.2.6.2 The monotone inclusion case

Consider the monotone inclusion problem

$$(2.2.92) \quad \text{find } x^* \in \mathcal{H} \text{ such that } 0 \in Ax^*,$$

where  $A$  is maximally monotone operator. Given  $(x_k, y_k) \in \mathcal{H}^2$  and based upon an approximate finite difference of (2.2.8), we compute the  $(x_{k+1}, y_{k+1}) \in \mathcal{H}^2$  by

$$\text{performing } \begin{cases} \frac{1}{h}(x_{k+1} - x_k) + A_{\lambda_k}x_{k+1} + \theta_k(y_k - x_k) = 0, \\ \frac{1}{h}(y_{k+1} - y_k) + \kappa(y_k - x_k) = 0, \end{cases}$$

$$\text{which is equivalent to } \begin{cases} x_{k+1} = J_h^{A_{\lambda_k}}(x_k - h\theta_k(y_k - x_k)), \\ y_{k+1} = y_k - h\kappa(y_k - x_k), \end{cases}$$

where  $J_h^{A_{\lambda_k}} = (I + hA_{\lambda_k})^{-1}$  is the resolvent operator of the maximally monotone operator  $A_{\lambda_k}$  with index  $h$ . Set  $z_k = (1 + h\theta_k)x_k - h\theta_k y_k$ . Then we alternatively have

$z_k = x_k - h\theta_k(y_k - x_k)$  and the previous system can be rewritten as

$$(2.2.93a) \quad z_k = (1 + h\theta_k)x_k - h\theta_k y_k,$$

$$(2.2.93b) \quad x_{k+1} = J_h^{A\lambda_k}(z_k),$$

$$(2.2.93c) \quad y_{k+1} = (1 - h\kappa)y_k + h\kappa x_k.$$

Moreover, using Proposition A.1 (f), we have  $(A_\lambda)_\mu = A_{\lambda+\mu}$ . Then by (2.2.93b) we simply deduce that

$$x_{k+1} = J_h^{A\lambda_k}(z_k) = \frac{\lambda_k}{\lambda_k+h} z_k + \frac{h}{\lambda_k+h} J_{\lambda_k+h}^A z_k.$$

Furthermore, by (2.2.93c) we readily have  $x_k = \frac{1}{\kappa h} y_{k+1} - \frac{1-\kappa h}{\kappa h} y_k$ , which by (2.2.93a) amounts to  $z_k = y_k + \frac{1+h\theta_k}{\kappa h} (y_{k+1} - y_k)$ . Finally, using the above expressions of  $x_{k+1}$  and  $z_k$ , along with that of  $y_{k+1}$  given by (2.2.93c), we are led to

$$(2.2.94a) \quad z_{k-1} = y_{k-1} + \frac{1+h\theta_{k-1}}{\kappa h} (y_k - y_{k-1}),$$

$$(2.2.94b) \quad x_k = \frac{\lambda_{k-1}}{\lambda_{k-1}+h} z_{k-1} + \frac{h}{\lambda_{k-1}+h} J_{\lambda_{k-1}+h}^A(z_{k-1}),$$

$$(2.2.94c) \quad y_{k+1} = (1 - h\kappa)y_k + (h\kappa)x_k.$$

Interesting convergence properties can be expected from (2.2.94) in light of Theorem 2.2.4. Observe that (2.2.94) requires only one evaluation of the proximity operator of  $A$  per iteration compared with (PRINAM) in [12]. However these two algorithms can be regarded as being complemented.

## 2.2.7 Appendix

### 2.2.7.1 The Yosida approximation

The Yosida approximation enjoys numerous nice properties which often facilitating and simplifying calculations. Some of them are recalled below (see [31, 30]) :

**Proposition A.1.** *Let  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  be a maximal monotone operator and set  $J_\lambda^A := (I + \lambda A)^{-1}$ . Then, we have the following properties :*

- (a)  $J_{\lambda A}$  is single-valued, everywhere defined and nonexpansive,
- (b)  $\forall v \in \mathcal{H}, \forall \lambda > 0, A_\lambda v \in A(J_\lambda^A v)$ ,
- (c)  $\forall (u, v) \in \mathcal{H}^2, \forall \lambda > 0, \langle A_\lambda u - A_\lambda v, u - v \rangle \geq \lambda \|A_\lambda u - A_\lambda v\|^2$ ,
- (d)  $\forall (u, v) \in \mathcal{H}^2, \forall \lambda > 0, \|A_\lambda u - A_\lambda v\| \leq \frac{1}{\lambda} \|u - v\|$ ,

$$(e) \quad \forall \lambda > 0, A^{-1}(\{0\}) = A_{\lambda}^{-1}(\{0\}),$$

$$(f) \quad \forall \lambda > 0, \forall \mu > 0, (A_{\lambda})_{\mu} = A_{\lambda+\mu}.$$

**Lemma A.1.** *Let  $\gamma, \delta > 0$  and  $x, y \in \mathcal{H}$ . Then for  $z \in A^{-1}(\{0\})$ , we have*

$$(2.2.95) \quad \|\gamma A_{\gamma}x - \delta A_{\delta}y\| \leq 2\|x - y\| + 2\frac{|\gamma-\delta|}{\gamma}\|x - z\|,$$

$$(2.2.96) \quad \|A_{\gamma}x - A_{\delta}y\| \leq \left(3\frac{|\delta-\gamma|}{\delta\gamma} \times \|x - z\| + \frac{2}{\delta}\|x - y\|\right).$$

**Proof.** The proof of (2.2.95) can be found in [13]. Let us prove (2.2.96). To get this we simply have

$$\begin{aligned} A_{\gamma}x - A_{\delta}y &= \frac{1}{\delta} (\delta A_{\gamma}x - \delta A_{\delta}y) \\ &= \frac{1}{\delta} ((\delta - \gamma)A_{\gamma}x + (\gamma A_{\gamma}x - \delta A_{\delta}y)), \end{aligned}$$

hence

$$\|A_{\gamma}x - A_{\delta}y\| \leq \frac{1}{\delta} (|\delta - \gamma| \times \|A_{\gamma}x\| + \|\gamma A_{\gamma}x - \delta A_{\delta}y\|).$$

Consequently, by  $\|A_{\gamma}x\| \leq \frac{1}{\gamma}\|x - z\|$  and using (2.2.95), we obtain

$$\begin{aligned} \|A_{\gamma}x - A_{\delta}y\| &\leq \frac{1}{\delta} \left( \frac{|\delta-\gamma|}{\gamma} \times \|x - z\| + 2\|x - y\| + 2\frac{|\gamma-\delta|}{\gamma}\|x - z\| \right) \\ &= \frac{1}{\delta} \left( 3\frac{|\delta-\gamma|}{\delta\gamma} \times \|x - z\| + 2\|x - y\| \right), \end{aligned}$$

that is the desired inequality. ■

### 2.2.7.2 A technical result

**Lemma B.2** ([13] Lemma A.5). *Let  $\omega, \eta : [0, +\infty[ \rightarrow [0, +\infty[$  be absolutely continuous functions such that  $\eta \notin L^1(0, +\infty)$  and which satisfy  $\int_0^{+\infty} \omega(t)\eta(t)dt < \infty$ , along with  $|\dot{\omega}(t)| \leq \eta(t)$  for almost every  $t > 0$ . Then  $\lim_{t \rightarrow +\infty} \omega(t) = 0$ .*

## Acknowledgements

The authors would like to thank the two anonymous referees for their careful readings of the manuscript and their insightful comments and observations.

# Chapitre 3

## Minimisation structurée de fonctions convexes non lisses avec décroissance rapide des sous-gradients

---

### 3.1 Introduction

Dans ce chapitre,  $\mathcal{H}$  désigne un espace de Hilbert réel muni d'un produit scalaire noté  $\langle \cdot, \cdot \rangle$  et de la norme induite notée  $\| \cdot \|$ . Nous nous intéressons à des méthodes du premier ordre pour la résolution de problèmes de minimisation structurés non lisses. Ce type d'algorithme a gagné en popularité, en raison de ses caractéristiques remarquables qui en font une alternative intéressante aux méthodes de points intérieurs, coûteuses en calcul. Le contenu de ce chapitre est basé sur l'article [72].

#### 3.1.1 Présentation du problème et de l'algorithme proposé.

L'objectif de ce chapitre est de résoudre le problème de minimisation convexe structuré suivant :

$$(3.1.1) \quad \min\{\Theta(x) := f(x) + g(x) : x \in \mathcal{H}\},$$

où les fonctions  $f$  et  $g$  vérifient les conditions suivantes :

(3.1.2a)  $g : \mathcal{H} \rightarrow ]-\infty, +\infty]$  est convexe, propre et semi-continue inférieurement,

(3.1.2b)  $f : \mathcal{H} \rightarrow ]-\infty, +\infty[$  est convexe et continûment différentiable,  
et son gradient  $\nabla f$  est  $L$ -lipschitzien,

(3.1.2c)  $S := \operatorname{argmin}_{\mathcal{H}} \Theta \neq \emptyset$ .

Les conditions ci-dessus sont couramment utilisées pour résoudre le problème (3.1.1), en supposant de plus que l'application proximale de  $g$  (voir chapitre 1, Définition 1.2.2) ne soit pas difficile à calculer. Notons que (3.1.1) inclut de nombreux problèmes concrets intervenant dans le traitement d'images, la vision par ordinateur, l'apprentissage automatique et le traitement du signal. Le problème (3.1.1) se pose par exemple en acquisition comprimée avec  $g(x) = \|x\|_1$  (voir [24, 40]) et comprend également d'autres problèmes bien étudiés tels que les inégalités variationnelles.

**Remarque 3.1.1.** L'application proximale de  $g$  d'indice  $\mu > 0$  est définie pour  $x \in \mathcal{H}$  par  $\operatorname{prox}_{\mu g}(x) := \operatorname{argmin}_{y \in \mathcal{H}} (g(y) + (2\mu)^{-1}\|x - y\|^2)$ , autrement dit, il s'agit de l'opérateur résolvante  $J_{\mu}^{\partial g} := (I + \mu \partial g)^{-1}$ , où  $\partial g : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  désigne le sous-différentiel de Fenchel de  $g$ . Il est bien connu pour être univoque et partout défini.

Nous discutons, dans ce chapitre, d'un algorithme de type inertiel relaxé (que nous appellerons RIGPA), permettant de résoudre le problème (3.1.1)-(3.1.2). Celui-ci est donné, pour des suites  $\{z_n, x_n, y_n\} \subset \mathcal{H}$  par

### RIPGA

► **Initialisation :** Soient  $\{x_0, y_0\} \subset \mathcal{H}$ ,  $(\nu_n) \subset ]0, +\infty[$  et  $e > 0$ .

► **Étape principale :** Pour  $n \geq 1$  et  $\{y_{n-1}, y_n\}$  donnés, on effectue

$$(3.1.3a) \quad z_{n-1} = y_n + \alpha_{n-1}(y_n - y_{n-1}),$$

$$(3.1.3b) \quad x_n = \operatorname{prox}_{\mu g}(z_{n-1} - \mu \nabla f(z_{n-1})),$$

$$(3.1.3c) \quad y_{n+1} = (1 - \kappa)y_n + \kappa x_n,$$

où le terme inertiel  $(\alpha_n) \subset ]0, \kappa^{-1}[$  est défini par

$$(3.1.4) \quad \alpha_n = \kappa^{-1} \left( \frac{\nu_n + (1 - \kappa)e}{\nu_{n+1} + e} \right).$$

Il sera démontré (voir Section 3.2.2) que ce schéma peut être vu comme une variante discrète d'un système dynamique du premier ordre récemment étudié.

Il convient de noter qu'un algorithme inertiel de type *Forward-Backward* relaxé semblable à (3.1.3) a été récemment étudié par Attouch-Cabot [8], en utilisant toutefois des paramètres variables  $(\kappa_n)$  et  $(\mu_n)$  au lieu des valeurs constantes  $\kappa$  et  $\mu$ . De plus, ce dernier incorpore un terme de relaxation de la forme  $(1 - \kappa_n)z_{n-1}$  à la place de  $(1 - \kappa)y_n$  dans l'algorithme proposé (3.1.3)-(3.1.4). En outre, le paramètre d'extrapolation  $\alpha_n$  de [8] est compris entre 0 et 1, tandis que pour la méthode proposée (3.1.3)-(3.1.4), le terme  $\alpha_n$  peut être supérieur à 1 pour  $\kappa \in ]0, 1[$  (car  $\alpha_n \rightarrow \kappa^{-1}$ , dès lors que  $\nu_n \rightarrow +\infty$  quand  $n \rightarrow +\infty$ ). Une approche similaire combinant relaxation et terme inertiel a aussi été étudiée par Iutzeler-Hendrickx [62] concernant les opérateurs moyennés (averaged).

### 3.1.2 Quelques rappels et liens avec l'état de l'art.

Mettons en évidence quelques liens entre la méthode proposée (3.1.3)-(3.1.4) et d'autres algorithmes de type proximal existants pour résoudre (3.1.1)-(3.1.2) :

- Pour  $\nu_n = 0$  et  $\kappa = 1$  (et donc  $\alpha_n \equiv 0$ ), l'algorithme (3.1.3)-(3.1.4) devient la méthode *Forward-Backward* (FB) (voir [Chapitre 1, Section 1.2.3])

$$(3.1.5) \quad y_{n+1} = \text{prox}_{\mu g}(y_n - \mu \nabla f(y_n)),$$

qui a été considérée sous la condition

$$(3.1.6) \quad \mu \in \left] 0, \frac{1}{L} \right[.$$

Rappelons que (3.1.5) est l'un des premiers algorithmes célèbres permettant de résoudre (3.1.1)-(3.1.2). Ce dernier génère des suites (faiblement) convergentes  $(y_n)$  qui vérifient le taux de convergence  $\Theta(y_n) - \min_{\mathcal{H}} \Theta = \mathcal{O}(n^{-1})$  pour la minimisation des valeurs de la fonction objectif  $\Theta := f + g$  (voir chapitre 1, Section 1.2.3). La méthode FB (3.1.5) a par la suite été améliorée via, notamment, l'algorithme *Fast Iterative Soft Thresholding Algorithm* (FISTA) proposé par Beck-Teboulle [24], sur la base des idées de Nesterov [81] et de Güler [57]. Il a été démontré que FISTA génère des itérés  $(y_n)$  qui garantissent le taux de convergence  $\Theta(y_n) - \min_{\mathcal{H}} \Theta = \mathcal{O}(n^{-2})$ . Cependant la convergence des itérés n'a pas été établie.

- Pour  $\kappa = 1$ ,  $e = \alpha - 1$  et  $\nu_n = n$ , le schéma (3.1.3)-(3.1.4) se réduit à la méthode AFB (*Accelerated Forward-Backward*), qui est une légère variante de FISTA introduite par Chambolle-Dossal [40] (voir aussi Attouch-Peypouquet [14]) comme étant l’algorithme inertiel :

$$(3.1.7) \quad \begin{aligned} z_{n-1} &= y_n + \frac{n-1}{n+\alpha-1}(y_n - y_{n-1}), \\ y_{n+1} &= \text{prox}_{\mu g}(z_{n-1} - \mu \nabla f(z_{n-1})), \end{aligned}$$

pour une certaine constante  $\alpha > 0$ . Rappelons (voir chapitre 1, Section 1.2.4), sous la condition (3.1.6) et pour  $\alpha > 3$ , que l’algorithme AFB génère des suites  $(y_n)$  faiblement convergentes vérifiant le taux  $\Theta(y_n) - \min_{\mathcal{H}} = o(n^{-2})$ , au lieu des taux  $\mathcal{O}(n^{-1})$  et  $\mathcal{O}(n^{-2})$  obtenu respectivement pour (3.1.5) et FISTA. Des liens intéressants entre AFB et les systèmes dynamiques continus du second ordre ont été établis (voir, par exemple, May [77], Attouch-Chbani-Peypouquet-Redont[16]). Ces liens se révèlent utiles pour comprendre le phénomène d’accélération résultant de la présence du terme inertiel dans (3.1.7). En particulier, il a été prouvé que AFB provient d’une discrétisation du modèle continu du second ordre suivant, qui fait intervenir un amortissement linéaire isotrope et un paramètre « visqueux », c’est-à-dire qui tend lentement vers 0 (voir [16])

$$(3.1.8) \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla\Theta(x(t)) = 0.$$

Les schémas de type inertiel ont été considérablement développés ces dix dernières années dans le cadre de problèmes d’optimisation structurés : par exemple, les algorithmes pour la minimisation convexe [9], ceux intégrant des procédures de recherche linéaire (Cruz-Nghia [47], Nesterov [82], Scheinberg-Goldfarb-Bai [101]), mais aussi ceux pour la résolution d’inclusions monotones (Moudafi-Oliny [78], Lorentz-Pock [70]).

### 3.1.3 Liens avec quelques systèmes dynamiques continus.

Ce travail a été inspiré par un article d’Attouch-Peypouquet-Redont [15] dans lequel les auteurs ont étudié le système dynamique du second ordre

$$(3.1.9) \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta \nabla^2\Theta(x(t))\dot{x}(t) + \nabla\Theta(x(t)) = 0,$$

où  $\beta$  et  $\alpha$  sont des constantes strictement positives et où  $\Theta$  est une fonction réelle convexe et lisse. Ce système comprend deux types d'amortissement : un amortissement géométrique et un amortissement linéaire isotrope comprenant un paramètre visqueux (voir aussi [4, 7]). Il a été démontré, pour  $\alpha > 3$ , que les trajectoires  $x(\cdot) : [t_0, +\infty[ \rightarrow \mathcal{H}$  (avec  $t_0 > 0$ ) générées par (3.1.9) convergent faiblement vers un élément de  $\operatorname{argmin}\Theta$  (lorsque  $t \rightarrow +\infty$ ) et vérifient  $\Theta(x(t)) - \min_{\mathcal{H}} \Theta = o(t^{-2})$  ainsi que  $\int_{t_0}^{+\infty} t^2 \|\nabla\Theta(x(t))\|^2 dt < +\infty$  (cette dernière propriété représente la convergence rapide vers zéro des gradients).

Le modèle (3.1.9) peut être interprété comme une variante améliorée du modèle continu de AFB (donnée par (3.1.8)) qui fait intervenir un amortissement géométrique de manière à réduire les oscillations observées dans les trajectoires de (3.1.7). Une caractéristique intéressante de (3.1.9) est que ce modèle peut être étendu (dans un certain sens) au cas non lisse de  $\Theta$  via un système dynamique différentiel du premier ordre.

Nous baserons notre étude sur un modèle continu discuté dans [66] qui s'avère être étroitement lié à (3.1.9). Ce modèle est donné par le système dynamique du second ordre

$$(3.1.10) \quad \ddot{x}(t) + a(t)\dot{x}(t) + \nabla^2\Theta(x(t))\dot{x}(t) + b(t)\nabla\Theta(x(t)) = 0,$$

avec  $a$  et  $b$  des applications définies par

$$(3.1.11) \quad a(\cdot) := \frac{1}{\theta}(\theta(\kappa - \theta) - \dot{\theta}) \quad \text{et} \quad b(\cdot) := \frac{1}{\theta}(\kappa\theta - \dot{\theta}),$$

où  $\kappa > 0$  et où  $\theta$  est défini, pour  $e > 0$  et pour une fonction  $\nu$  différentiable strictement positive, par

$$(3.1.12) \quad \theta(t) = \frac{\kappa\nu(t) - \dot{\nu}(t)}{\nu(t) + e},$$

$\dot{\nu}$  étant la dérivée de  $\nu$ .

Un premier point important établi dans [66] (voir également la Proposition 3.2.2) est que (3.1.10) peut être étendu au cas non lisse de  $\Theta$  via le système différentiel du

premier ordre

$$(3.1.13) \quad \begin{cases} \dot{x}(t) + \partial\Theta(x(t)) + \theta(t)(y(t) - x(t)) \ni 0, \\ \dot{y}(t) + \kappa(y(t) - x(t)) = 0. \end{cases}$$

Un deuxième point essentiel mentionné dans [66] est que (3.1.10) (avec un choix approprié de  $\nu$ ) se comporte asymptotiquement comme un système sur-amorti de la forme (3.1.9). Les auteurs ont montré, dans le cas particulier où  $\nu(t) = t + \nu_0$  (avec  $\nu_0 > 0$ ), que les fonctions  $a, b$  et  $\theta$  définies respectivement par (3.1.11) et (3.1.12) satisfont

$$(3.1.14) \quad \theta(t) \rightarrow \kappa, \quad a(t) \sim \frac{1 + \kappa e}{t} \quad \text{et} \quad b(t) \rightarrow \kappa \quad (\text{lorsque } t \rightarrow +\infty).$$

Ainsi, la formulation simple de (3.1.13) offre un cadre intéressant pour la conception d'une nouvelle classe d'algorithmes de minimisation, comme proposé dans ce chapitre. Nous montrerons en effet (voir Section 3.2.2) que (3.2.4)-(3.2.5) peut être considéré comme une approximation par différences finies de (3.1.13)-(3.1.12).

### 3.1.4 Esquisse des principaux résultats.

Nous proposons une nouvelle classe d'algorithmes, donnée par (3.1.3)-(3.1.4) et qui inclue AFB comme cas particulier (lorsque  $\kappa = 1$  et  $\nu_n = n$ ) pour résoudre des problèmes de minimisation structurés convexes non lisses.

Plusieurs résultats de convergence sont établis concernant la méthode (3.1.3)-(3.1.4) dans le cadre de paramètres  $(\nu_n)$  généraux ou particuliers. L'approche développée dans ce chapitre repose sur des propriétés de Lyapunov d'une suite de type énergie  $\mathcal{E}_n(s, q)$  (définie par (3.3.2)) associée au schéma (3.1.3)-(3.1.4).

Les principales propriétés asymptotiques de (3.1.3)-(3.1.4) sont établies par les Théorèmes 3.4.1 et 3.4.2. Le Théorème 3.4.1 est consacré au cas particulier où  $\nu_n$  est de la forme  $\nu_n = n + \nu_0$  (pour un certain  $\nu_0 > 0$ ) tandis que le Théorème 3.4.2 fait appel à une suite  $(\nu_n)$  définie récursivement vérifiant  $\nu_{n+1} - \nu_n \sim 1$  (lorsque  $n \rightarrow +\infty$ ). Ces deux choix de  $(\nu_n)$  assurent la décroissance de  $\mathcal{E}_n(s, q)$  pour  $n$  assez grand, mais le dernier choix garantit cette décroissance à partir de  $n = 1$ .

Plus précisément, nous prouvons, sous les conditions

$$(3.1.15) \quad \kappa \in ]0, 2[ \quad \text{et} \quad e > 2\kappa^{-1},$$

que l'algorithme (3.1.3)-(3.1.4) génère des suites (faiblement) convergentes  $(x_n)$  qui vérifient un  $o(n^{-2})$  pour la minimisation des valeurs de la fonction. Notons que ces propriétés sont bien connue pour AFB lorsque  $e > 2$  (voir chapitre 1, Section 1.2.4).

De plus, nous étendons au cadre non lisse la propriété de convergence rapide vers zéro des sous-gradients de la fonction objectif, établie jusqu'à présent pour les potentiels lisses uniquement. Plus précisément, toujours pour des paramètres vérifiant (3.1.15), nous prouvons l'existence de sous-gradients  $\xi_n^* \in \partial\Theta(x_n)$  qui satisfont

$$(3.1.16) \quad \sum_n n^2 \|\xi_n^*\|^2 < +\infty.$$

Le phénomène (3.1.16) a été récemment découvert par Shi-Du-Jordan-Su [103] pour une nouvelle classe d'algorithmes de type Nesterov (basées sur des EDO à haute résolution) pour la minimisation de fonctions convexes lisses et lipschitziennes dans des espaces euclidiens. Notons que la convergence des itérés produits par ces algorithmes n'a pas été prouvée. Des résultats similaires à celui de [103] ont également été proposés par Attouch-Chbani-Fadili-Riahi [18] dans le cadre des espaces de Hilbert. Rappelons que les meilleures estimations obtenues jusqu'à présent concernant AFB garantissent seulement que

$$(3.1.17) \quad \sum_n n \|\xi_n^*\|^2 < +\infty \quad \text{et} \quad \|\xi_n^*\| = o(n^{-1}).$$

Comme souligné dans [103], l'estimation (3.1.17) est moins fine que (3.1.16) selon laquelle on obtient

$$(3.1.18) \quad \min_{0 \leq k \leq n} \|\xi_k^*\| = \mathcal{O}\left(n^{-\frac{3}{2}}\right).$$

Nous montrons que ce résultat reste valable dans le cas particulier de AFB, avec une condition plus restrictive que (3.1.6). Néanmoins, notre algorithme offre quelques degrés de liberté supplémentaires grâce aux paramètres  $\kappa$  et  $\nu_n$  qui peuvent être utilisés pour influencer (accélérer) la convergence des itérés.

### 3.1.5 Organisation du chapitre.

Le plan de ce chapitre est le suivant. Dans la Section 3.2, nous donnons quelques préliminaires sur la méthode numérique. Des reformulations très utiles du schéma (3.1.3)-(3.1.4) sont établies. Le lien entre le schéma proposé et le système dynamique du premier ordre (3.1.13)-(3.1.12) est également discuté. Une inégalité importante est énoncée dans la Section 3.3 en vue d'une analyse de Lyapunov. Dans la Section 3.4, nous établissons quelques estimations sur les suites générées par l'algorithme (3.1.3)-(3.1.4) et nous présentons nos principaux résultats de convergence.

**Remarque 3.1.2.** À partir de maintenant, nous adopterons les notations suivantes :

- Pour une suite  $(u_n)$  donnée, on note  $\dot{u}_n = u_n - u_{n-1}$  et  $\ddot{u}_n = \dot{u}_n - \dot{u}_{n-1}$ .
- Pour un réel  $t$ , on note  $(t)_+ = \max\{0, t\}$ .

## 3.2 Préliminaires

### 3.2.1 Un modèle du second ordre discret

En suivant la méthodologie de [8], nous reformulons le schéma (3.1.3)-(3.1.4) ci-dessus à l'aide de l'opérateur  $M_\mu$  défini pour  $z \in \mathcal{H}$  et  $\mu > 0$  par

$$(3.2.1) \quad M_\mu(z) := \frac{1}{\mu}(z - \text{prox}_{\mu g}(z - \mu \nabla f(z))).$$

**Remarque 3.2.3.** L'opérateur  $M_\mu$  est couramment utilisé comme astuce classique pour évaluer les valeurs des sous-gradients dans les algorithmes proximaux (voir, par exemple, [8, 14]). Rappelons que, pour tout  $z \in \mathcal{H}$ , nous avons les équivalences :

$$M_\mu(z) = 0 \iff z = J_\mu^{\partial g}(z - \nabla f(z)) \iff 0 \in \partial\Theta(z).$$

La proposition suivante sera particulièrement importante pour notre étude.

**Proposition 3.2.1.** *Toutes suites  $\{z_n, x_n, y_n\} \subset \mathcal{H}$  générées par (3.1.3)-(3.1.4), vérifient (pour  $n \geq 1$ )*

$$(3.2.2a) \quad z_n - x_{n+1} = \mu M_\mu(z_n),$$

$$(3.2.2b) \quad M_\mu(z_{n-1}) \in \partial g(x_n) + \nabla f(z_{n-1}),$$

$$(3.2.2c) \quad \dot{y}_{n+1} - \dot{y}_n + (1 - \kappa\alpha_{n-1})\dot{y}_n + (\kappa\mu)M_\mu(z_{n-1}) = 0.$$

**Démonstration.** Les éléments (3.2.2a) et (3.2.2b) sont classiquement obtenus via la formulation proximale de (3.1.3). De plus, d'après (3.1.3c), les suites  $(y_n)$  et  $(z_n)$  générées par (3.1.3)-(3.1.4) vérifient

$$(3.2.3) \quad \dot{y}_{n+1} - (\kappa\alpha_{n-1})\dot{y}_n + (\kappa\mu)M_\mu(z_{n-1}) = 0,$$

ce qui nous donne directement (3.2.2c). ■

## 3.2.2 Une variante discrète d'un modèle continu.

Le but de cette section est de souligner le lien entre la méthode proposée (3.1.3)-(3.1.4) et le système du premier ordre (3.1.13)-(3.1.12).

### 3.2.2.1 D'un modèle du second ordre à un modèle du premier ordre.

La proposition suivante montre l'équivalence des systèmes (3.1.13) et (3.1.10) lorsque  $\Theta$  est supposé être deux fois continûment différentiable (voir [66]).

**Proposition 3.2.2** ([66], Proposition 2.1). *Soit  $\kappa \in ]0, +\infty[$ . Supposons que  $\Theta \in C^2(\mathcal{H})$  et que  $\theta : ]0, +\infty[ \rightarrow ]0, +\infty[$  soit une fonction différentiable. Alors les propositions suivantes sont équivalentes :*

- (i1)  $x$  est une solution classique du système du second ordre (3.1.10);
- (i2)  $(x, y)$  est une solution classique du système du premier ordre (3.1.13).

### 3.2.2.2 Un modèle continu de la méthode proposée.

Les commentaires suivants sont destinés à souligner le lien entre le modèle continu (3.1.13)-(3.1.12) et l'algorithme (3.1.3)-(3.1.4). Ce dernier provient en fait d'une reformulation d'une variante discrète du modèle continu (3.1.13)-(3.1.12). En effet,

une discrétisation temporelle de (3.1.13)-(3.1.12) peut être envisagée via l'algorithme

$$(3.2.4a) \quad z_n = x_n + \theta_n(x_n - y_n),$$

$$(3.2.4b) \quad x_{n+1} = \text{prox}_{\mu g}(z_n - \mu \nabla f(z_n)),$$

$$(3.2.4c) \quad y_{n+1} = (1 - \kappa)y_n + \kappa x_n,$$

où  $\kappa \in ]0, +\infty[$  et où  $\theta_n$  est défini, pour une certaine suite croissante strictement positive  $(\nu_n)$ , par

$$(3.2.5) \quad \theta_n = \frac{\kappa \nu_{n+1} - \dot{\nu}_{n+1}}{\nu_{n+1} + e}.$$

**Remarque 3.2.4.** On peut vérifier que (3.2.4)-(3.2.5) peut s'écrire de manière équivalente sous la forme

$$(3.2.6a) \quad z_n = x_n - \theta_n(y_n - x_n),$$

$$(3.2.6b) \quad \dot{x}_{n+1} + \mu(\partial g(x_{n+1}) + \nabla f(z_n)) + \theta_n(y_n - x_n) \ni 0,$$

$$(3.2.6c) \quad \dot{y}_{n+1} + \kappa(y_n - x_n) = 0.$$

Ainsi, en remplaçant  $\mu$  et  $\kappa$  par  $\mu \Delta t$  et  $\kappa \Delta t$  respectivement dans (3.2.6) (pour une certaine quantité positive  $\Delta t$ ), on observe que (3.2.4)-(3.2.5) peut être vu comme une discrétisation par différences finies du modèle (3.1.13)-(3.1.12). Notons que ce modèle discret fait intervenir le terme explicite  $\nabla f(z_n)$  au lieu du terme  $\nabla f(x_{n+1})$ .

Nous montrons ensuite que (3.1.3)-(3.1.4) entre dans le cas du modèle (3.2.4)-(3.2.5).

**Proposition 3.2.3.** Soient les suites  $\{z_n, x_n, y_n\} \subset \mathcal{H}$  générées par (3.1.3)-(3.1.4). Alors, pour  $n \geq 1$ , on a

$$(3.2.7a) \quad z_n = x_n - \theta_n(y_n - x_n), \quad \text{où} \quad \theta_n = \frac{\kappa \nu_{n+1} - \dot{\nu}_{n+1}}{e + \nu_{n+1}},$$

$$(3.2.7b) \quad x_{n+1} = J_{\mu}^{\partial g}(z_n - \mu \nabla f(z_n)),$$

$$(3.2.7c) \quad y_{n+1} = (1 - \kappa)y_n + \kappa x_n.$$

**Démonstration.** En effet, pour  $n \geq 1$ , d'après (3.1.3c) nous avons

$$\dot{y}_{n+1} = \kappa(x_n - y_n),$$

c'est-à-dire (3.2.7c), tandis que la définition de  $x_n$  dans (3.1.3b) nous donne

$$x_{n+1} = J_\mu^{\text{dg}}(z_n - \mu \nabla f(z_n)),$$

on obtient alors directement (3.2.7b). De plus, en posant  $u_n = y_n - x_n$ , on a, par définition de  $z_n$  (dans (3.1.3a))

$$z_n = y_{n+1} + \alpha_n \dot{y}_{n+1} = x_n + u_n + (1 + \alpha_n) \dot{y}_{n+1}.$$

Donc, comme  $\dot{y}_{n+1} = -\kappa u_n$  (d'après (3.2.7c)), on obtient

$$z_n = x_n + u_n + (1 + \alpha_n)(-\kappa u_n) = x_n + (1 - \kappa - \kappa \alpha_n) u_n.$$

Ainsi, en observant que  $\theta_n = \kappa + \kappa \alpha_n - 1$ , la dernière égalité nous donne (3.2.7a).

■

### 3.3 Résultats généraux pour une analyse de Lyapunov

Dans cette section, nous présentons une suite de type énergie associée à l'algorithme (3.1.3)-(3.1.4). Plus précisément, nous nous concentrons ici sur les itérés générés par la formulation (3.2.4)-(3.2.5) de l'algorithme considéré.

#### 3.3.1 Fonctionnelle de Lyapunov associée à la méthode.

Nous fixons, sous la condition  $(\nu_n)_{n \geq 0} \subset ]0, +\infty[$ , les paramètres suivants

$$(3.3.1) \quad \rho_n := 1 - (1 - \kappa) \frac{\nu_{n+1}}{\nu_n} \quad (\text{pour } n \geq 0) \quad \text{et} \quad \rho_{-1} := \rho_0.$$

En posant  $\bar{\Theta} = \Theta - \min \Theta$ , nous établirons des propriétés de type Lyapunov pour la quantité  $\mathcal{E}_n(s, q)$  définie pour  $(s, q) \in ]0, +\infty[ \times S$  par

$$(3.3.2) \quad \begin{aligned} \mathcal{E}_n(s, q) = & \frac{1}{2} \|s(q - x_n) + \nu_n(y_n - x_n)\|^2 \\ & + \frac{1}{2} s(e - s) \|x_n - q\|^2 + \mu \rho_{n-1}^{-1} (\nu_n + e)^2 \bar{\Theta}(x_n), \end{aligned}$$

Notons que la forme de cette suite de type énergie est bien connue dans la littérature (voir, par exemple, [5, 16, 40, 105]).

Un résultat préliminaire important concernant la suite  $(\mathcal{E}_n(s, q))$  est donné par le lemme suivant :

**Lemme 3.3.1.** *Supposons que les conditions (3.1.2) soient vérifiées. Soient  $\{x_n, y_n\}_{n \geq 0} \subset \mathcal{H}$  des suites générées par (3.2.4)-(3.2.5) avec des paramètres  $\{\kappa, e, \mu\}$  strictement positifs. Soit  $(\nu_n)$  vérifiant les conditions suivantes :*

$$(3.3.3a) \quad (\nu_n)_{n \geq 0} \subset ]\kappa - 1, +\infty[ \text{ est croissante,}$$

$$(3.3.3b) \quad \left( \frac{\nu_{n+1}}{\nu_n} \right)_{n \geq 0} \text{ est décroissante,}$$

$$(3.3.3c) \quad (1 - \kappa) \frac{\nu_{n+1}}{\nu_n} < 1, \quad (\text{pour } n \geq 0).$$

Alors, pour  $(s, q) \in ]0, e] \times S$  et pour  $n \geq 0$ , on a

$$(3.3.4) \quad \begin{aligned} \dot{\mathcal{E}}_{n+1}(s, q) + \mu(s - \beta_n)(e + \nu_{n+1})\bar{\Theta}(x_n) \\ + \frac{1}{2}(1 - L\mu)\rho_n^{-1}(e + \nu_{n+1})^2 \|\dot{x}_{n+1} + \theta_n(y_n - x_n)\|^2 \leq -\mathcal{T}_n, \end{aligned}$$

où  $\rho_n$  et  $\mathcal{E}_n(s, q)$  sont donnés respectivement par (3.3.1) et (3.3.2),  $\bar{\Theta} := \Theta - \min \Theta$ ,  $(\beta_n)_{n \geq 0}$  est défini pour  $n \geq 0$  par

$$(3.3.5) \quad \beta_0 = \frac{2}{\rho_0} \dot{\nu}_1,$$

$$(3.3.6) \quad \beta_n = \frac{2}{\rho_n} \dot{\nu}_{n+1} + \frac{(\kappa-1)_+}{\rho_n \rho_{n-1}} \left( \frac{\nu_n}{\nu_{n-1}} - \frac{\nu_{n+1}}{\nu_n} \right) (e + \nu_n) \quad (\text{si } n \geq 1),$$

et où  $\mathcal{T}_n$  est défini pour  $n \geq 0$  par

$$(3.3.7) \quad \begin{aligned} \mathcal{T}_n = \frac{1}{2}(1 - \kappa)\rho_n \nu_n \nu_{n+1} \left\| y_n - x_n + \left( \frac{1}{\theta_n} - \frac{e-s}{\nu_n \rho_n} \right) \dot{x}_{n+1} \right\|^2 \\ + \frac{1}{2}(e - s) \left( (e + 2\nu_{n+1}) \left( 1 + \frac{(1-\kappa)\nu_{n+1}}{\nu_n \rho_n} \right) + s \frac{(1-\kappa)\nu_{n+1}}{\nu_n \rho_n} \right) \|\dot{x}_{n+1}\|^2. \end{aligned}$$

Ce résultat sera démontré dans la section 3.3.2.3.

### 3.3.2 Démonstration du Lemme 3.3.1.

#### 3.3.2.1 Une égalité importante concernant un modèle plus général.

Afin de démontrer le Lemme 3.3.1, nous établissons une égalité importante tirée d'un schéma plus général qui peut s'avérer utile pour étudier des algorithmes apparentés. Ce schéma consiste en des suites  $(x_n, y_n, \chi_n^*) \in \mathcal{H}^3$  et des paramètres  $\{\kappa, e, \theta_n, \nu_n\} \subset ]0, +\infty[$  vérifiant

$$(3.3.8a) \quad \dot{x}_{n+1} + \chi_n^* + \theta_n(y_n - x_n) = 0,$$

$$(3.3.8b) \quad \dot{y}_{n+1} + \kappa(y_n - x_n) = 0,$$

où  $\theta_n$  est donné par (3.2.5).

En tant qu'élément clé de notre méthodologie, nous associons à (3.3.8) la quantité  $G_n(s, q)$  donnée pour tout  $(s, q) \in [0, +\infty[ \times \mathcal{H}$  par

$$(3.3.9) \quad G_n(s, q) = \frac{1}{2} \|s(q - x_n) + \nu_n(y_n - x_n)\|^2 + \frac{1}{2} s(e - s) \|x_n - q\|^2.$$

Une égalité importante concernant la suite  $(G_n(s, q))$  est établie par la proposition suivante.

**Proposition 3.3.1.** *Soient  $p \in \mathbb{N}$ . Soient  $\{x_n, y_n, \chi_n^*\}_{n \geq p} \subset \mathcal{H}$  des suites vérifiant (3.3.8) pour  $n \geq p$ . Soient  $\{\theta_n, \nu_n\}_{n \geq p} \subset ]0, +\infty[$  et  $\{\kappa, e\} \subset ]0, +\infty[$ . Supposons que la condition suivante soit vérifiée pour  $n \geq p$*

$$(3.3.10) \quad (1 - \kappa) \frac{\nu_{n+1}}{\nu_n} < 1.$$

Alors pour  $(s, q) \in ]0, +\infty[ \times \mathcal{H}$  et pour  $n \geq p$  nous avons

$$(3.3.11) \quad \begin{aligned} \dot{G}_{n+1}(s, q) + \frac{1}{2} \rho_n^{-1} (e + \nu_{n+1})^2 \|\dot{x}_{n+1} + \theta_n(y_n - x_n)\|^2 \\ + s(e + \nu_{n+1}) \langle \chi_n^*, x_{n+1} - q \rangle \\ + (e + \nu_{n+1})^2 \left( \rho_n^{-1} - \frac{s}{e + \nu_{n+1}} \right) \langle \chi_n^*, \dot{x}_{n+1} \rangle = -\mathcal{T}_n, \end{aligned}$$

où  $\rho_n$  est donné par (3.3.1) et où  $\mathcal{T}_n$  est défini par (3.3.7).

La preuve de la Proposition 3.3.1 est donnée en Annexe 3.6.1.

### 3.3.2.2 RIGPA : un cas particulier d'un modèle plus général.

La remarque suivante nous montre le lien entre RIGPA et le modèle plus général introduit précédemment.

**Remarque 3.3.5.** *On observe, à l'aide de la Remarque 3.2.4, que le modèle discret (3.2.4)-(3.2.5) s'écrit de façon équivalente*

$$(3.3.12a) \quad z_n = x_n - \theta_n(y_n - x_n),$$

$$(3.3.12b) \quad \dot{x}_{n+1} + \mu(x_{n+1}^* + \nabla f(z_n)) + \theta_n(y_n - x_n) = 0, \quad \text{avec } x_{n+1}^* \in \partial g(x_{n+1}),$$

$$(3.3.12c) \quad \dot{y}_{n+1} + \kappa(y_n - x_n) = 0.$$

Par conséquent, le modèle (3.2.4)-(3.2.5) entre dans le cadre du modèle général (3.3.8) en prenant

$$(3.3.13) \quad \chi_n^* = \mu(x_{n+1}^* + \nabla f(z_n)).$$

On peut également établir une formulation plus précise de la suite  $(\chi_n^*)_n$ . En effet, par (3.3.12a) et (3.3.12b) nous avons

$$(3.3.14) \quad z_n - x_{n+1} = \mu(x_{n+1}^* + \nabla f(z_n)),$$

ainsi que  $z_n - x_{n+1} = \mu M_\mu(z_n)$  (d'après (3.2.2a)). On en déduit alors que

$$(3.3.15) \quad M_\mu(z_n) = x_{n+1}^* + \nabla f(z_n) \quad \text{et} \quad \chi_n^* = \mu M_\mu(z_n).$$

On obtient, de plus

$$(3.3.16) \quad \dot{x}_{n+1} + \theta_n(y_n - x_n) + \mu M_\mu(z_n) = 0.$$

cette dernière égalité se révélera utile pour nos calculs à venir.

### 3.3.2.3 Démonstration du Lemme 3.3.1.

Nous appliquons la Proposition 3.3.1 dans le cas des suites  $\{z_n, x_n, y_n\}$  générées par (3.2.4)-(3.2.5). De cette façon, et en exploitant les caractéristiques spécifiques de la fonction objectif  $\Theta := f + g$  (comme fait dans [14]), nous établissons des propriétés

de type Lyapunov pour la suite  $\mathcal{E}_n(s, q)$  donnée par (3.3.2).

Rappelons que toutes suites  $\{z_n, x_n, y_n\}$  générées par (3.2.4)-(3.2.5) vérifient également (3.3.8) ainsi que  $\chi_n^* = \mu M_\mu(z_n)$ , d'après la Remarque 3.3.5. Par conséquent, en appliquant la Proposition 3.3.1 et en remarquant que  $(\rho_n)_{n \geq 0} \subset ]0, +\infty[$  (d'après la condition (3.3.3c)), on obtient pour  $(s, q) \in ]0, e] \times S$  et pour  $n \geq 0$

$$(3.3.17) \quad \dot{G}_{n+1}(s, q) + \frac{1}{2} \rho_n^{-1} \tau_n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2 + \mu R_n = -\mathcal{T}_n,$$

où  $\tau_n = \nu_{n+1} + e$ ,  $u_n = y_n - x_n$ , et où  $\theta_n$ ,  $G_n(s, q)$  et  $\mathcal{T}_n$  sont donnés par (3.2.5), (3.3.9) et (3.3.7) respectivement, tandis que  $R_n$  est défini par

$$(3.3.18) \quad R_n := s \tau_n \langle M_\mu(z_n), x_{n+1} - q \rangle + \rho_n^{-1} \tau_n^2 \left(1 - \frac{s \rho_n}{\tau_n}\right) \langle M_\mu(z_n), \dot{x}_{n+1} \rangle.$$

Le reste de la démonstration peut être divisée en plusieurs étapes **(B0)**-**(B3)** :

**(B0)** Montrons que pour  $n \geq 0$ , la quantité  $1 - \frac{s \rho_n}{\tau_n}$  intervenant dans la définition de  $R_n$  (donnée par (3.3.18)) est strictement positive pour  $s \in [0, e]$  et  $\nu_0 > e(\kappa - 1)_+$ .

En effet, pour  $s \in [0, e]$ , on a

$$1 - \frac{s \rho_n}{\tau_n} = 1 - \frac{s \rho_n}{e + \nu_{n+1}} \geq 1 - \frac{e \rho_n}{e + \nu_{n+1}}.$$

On vérifie, d'après la définition de  $\rho_n$  (à savoir  $\rho_n = 1 - (1 - \kappa) \nu_{n+1} \nu_n^{-1}$ ) que le terme du côté droit de l'inégalité précédente est positif si

$$(3.3.19) \quad e + \nu_{n+1} > e \left(1 + (1 - \kappa) \frac{\nu_{n+1}}{\nu_n}\right).$$

Ainsi, comme  $(\nu_n)$  est strictement positive, cette dernière inégalité se réduit à

$$(3.3.20) \quad \nu_n > e(\kappa - 1)_+.$$

On observe alors que l'inégalité (3.3.20) sera satisfaite pour tout  $n \geq 0$  dès lors qu'elle l'est pour  $n = 0$  (car  $(\nu_n)$  est supposée être croissante).

**(B1)** Montrons, par souci d'exhaustivité, que pour tout  $(x, z) \in \mathcal{H}^2$  nous avons

$$(3.3.21) \quad \Theta(z - \mu M_\mu(z)) \leq \Theta(x) + \langle M_\mu(z), z - x \rangle - \mu \left(1 - \frac{L\mu}{2}\right) \|M_\mu(z)\|^2.$$

En effet, posons  $y = z - \mu M_\mu(z)$  et  $v = y - z$  pour  $(x, z) \in \mathcal{H}^2$  donnés. On obtient, en appliquant la formule de Taylor,

$$f(y) - f(z) = \langle \nabla f(z), y - z \rangle + \int_0^1 \langle y - z, \nabla f(z + tv) - \nabla f(z) \rangle dt.$$

Par conséquent, sachant que  $\nabla f$  est  $L$ -lipschitzien et en remarquant que  $\|y - z\| = \|v\| = \mu \|M_\mu(z)\|$ , nous en déduisons que

$$f(y) - f(z) = \langle \nabla f(z), y - z \rangle + \int_0^1 \langle y - z, \nabla f(z + tv) - \nabla f(z) \rangle dt.$$

De plus, l'inégalité convexe nous donne

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

Ainsi, en ajoutant les deux inégalités précédentes, on obtient

$$(3.3.22) \quad f(y) - f(x) \leq \langle \nabla f(z), y - x \rangle + \frac{1}{2} L \mu^2 \|M_\mu(z)\|^2.$$

Par ailleurs, sachant que  $M_\mu(z) = \mu^{-1} (z - \text{prox}_{\mu g}(z - \mu \nabla f(z)))$  et  $y = z - \mu M_\mu(z)$ , il est alors classique de voir que l'on a

$$M_\mu(z) - \nabla f(z) \in \partial g(y).$$

Par conséquent, en posant  $\zeta = M_\mu(z) - \nabla f(z)$  (donc  $\zeta \in \partial g(y)$ ) et en utilisant à nouveau l'inégalité convexe, on obtient

$$g(y) - g(x) \leq \langle \zeta, y - x \rangle.$$

En outre, comme  $y - x = z - x - \mu M_\mu(z)$ , nous obtenons

$$(3.3.23) \quad \begin{aligned} g(y) - g(x) &\leq \langle M_\mu(z), y - x \rangle - \langle \nabla f(z), y - x \rangle \\ &= \langle M_\mu(z), z - x \rangle - \mu \|M_\mu(z)\|^2 - \langle \nabla f(z), y - x \rangle, \end{aligned}$$

En ajoutant (3.3.22) à (3.3.23), on obtient l'inégalité souhaitée.

**(B2)** Montrons que pour  $n \geq 0$ ,

$$(3.3.24) \quad \dot{G}_{n+1}(s, q) + \mu F_n + \frac{1}{2} (1 - L\mu) \rho_n^{-1} \tau_n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2 \leq -\mathcal{T}_n,$$

où  $\tau_n = e + \nu_{n+1}$  et où  $F_n$  est défini par

$$(3.3.25) \quad F_n = s\tau_n (\Theta(x_n) - \Theta(q)) + \tau_n^2 \rho_n^{-1} (\Theta(x_{n+1}) - \Theta(x_n)).$$

En effet, en utilisant **(B1)** et en remarquant que  $x_{n+1} = z_n - \mu M_\mu(z_n)$  (d'après (3.2.2a)), on obtient pour  $n \geq 0$ ,

$$(3.3.26a) \quad \Theta(x_{n+1}) \leq \Theta(q) + \langle M_\mu(z_n), x_{n+1} - q \rangle + \frac{1}{2} L\mu^2 \|M_\mu(z_n)\|^2,$$

$$(3.3.26b) \quad \Theta(x_{n+1}) \leq \Theta(x_n) + \langle M_\mu(z_n), \dot{x}_{n+1} \rangle + \frac{1}{2} L\mu^2 \|M_\mu(z_n)\|^2.$$

Ainsi, d'après (3.3.26a) et (3.3.26b), et en posant  $\bar{\Theta} = \Theta - \min \Theta$ , nous déduisons, par définition de  $R_n$ , que

$$\begin{aligned} R_n &\geq s\tau_n \left( \bar{\Theta}(x_{n+1}) - \frac{L\mu^2}{2} \|M_\mu(z_n)\|^2 \right) \\ &\quad + \tau_n(\rho_n^{-1}\tau_n - s) \left( \bar{\Theta}(x_{n+1}) - \bar{\Theta}(x_n) - \frac{L\mu^2}{2} \|M_\mu(z_n)\|^2 \right) \\ &= s\tau_n \bar{\Theta}(x_{n+1}) + \tau_n(\rho_n^{-1}\tau_n - s) (\bar{\Theta}(x_{n+1}) - \bar{\Theta}(x_n)) - \rho_n^{-1}\tau_n^2 \left( \frac{L\mu^2}{2} \|M_\mu(z_n)\|^2 \right) \\ &= s\tau_n \bar{\Theta}(x_n) + \rho_n^{-1}\tau_n^2 (\bar{\Theta}(x_{n+1}) - \bar{\Theta}(x_n)) - \frac{L\mu^2}{2} \rho_n^{-1}\tau_n^2 \|M_\mu(z_n)\|^2 \\ &= F_n - \frac{L\mu^2}{2} \rho_n^{-1}\tau_n^2 \|M_\mu(z_n)\|^2, \end{aligned}$$

ce qui, avec (3.3.17) et  $\|M_\mu(z_n)\|^2 = \mu^{-2} \|\dot{x}_{n+1} + \theta_n u_n\|^2$  (d'après (3.3.16)), entraîne

$$\begin{aligned} \mathcal{T}_n(s) + \dot{G}_{n+1}(s, q) &= -\frac{1}{2} \rho_n^{-1} \tau_n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2 - \mu R_n \\ &\leq -\frac{1}{2} \rho_n^{-1} \tau_n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2 - \mu F_n + \frac{1}{2} L\mu \rho_n^{-1} \tau_n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2 \\ &= -\mu F_n - \frac{1}{2} (1 - L\mu) \rho_n^{-1} \tau_n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2. \end{aligned}$$

Cela nous donne (3.3.24).

**(B3)** Soit  $q \in S$ . Posons  $\vartheta_n = \rho_n^{-1} (\nu_{n+1} + e)^2$  pour  $n \geq 0$  ainsi que  $\vartheta_{-1} = \rho_0^{-1} (\nu_0 + e)^2$ . Il reste à montrer que la quantité  $F_n$  intervenant dans (3.3.24) vérifie pour  $q \in S$  et  $n \geq 0$

$$(3.3.27) \quad F_n \geq \vartheta_n \bar{\Theta}(x_{n+1}) - \vartheta_{n-1} \bar{\Theta}(x_n) + (s - \beta_n) \tau_n \bar{\Theta}(x_n),$$

où  $\tau_n = e + \nu_{n+1}$  et où  $\beta_n$  est défini dans (3.3.5)-(3.3.6). Rappelons que  $\nu_{n+1}\nu_n^{-1}$  est décroissante pour  $\kappa \leq 1$ . Ainsi, un rapide calcul nous donne

$$(3.3.28) \quad \frac{\rho_{n-1} - \rho_n}{\rho_n \rho_{n-1}} = \frac{(1 - \kappa)}{\rho_n \rho_{n-1}} \left( \frac{\nu_{n+1}}{\nu_n} - \frac{\nu_n}{\nu_{n-1}} \right) \leq \frac{(\kappa - 1)_+}{\rho_n \rho_{n-1}} \left( \frac{\nu_n}{\nu_{n-1}} - \frac{\nu_{n+1}}{\nu_n} \right)_+.$$

D'où, par définition de  $\vartheta_n$  et pour  $n \geq 1$ ,

$$(3.3.29) \quad \begin{aligned} \dot{\vartheta}_n &= \frac{1}{\rho_n} \left( (e + \nu_{n+1})^2 - (e + \nu_n)^2 \right) + \left( \frac{1}{\rho_n} - \frac{1}{\rho_{n-1}} \right) (e + \nu_n)^2 \\ &= \frac{1}{\rho_n} \dot{\nu}_{n+1} (2e + \nu_{n+1} + \nu_n) + \left( \frac{\rho_{n-1} - \rho_n}{\rho_n \rho_{n-1}} \right) (e + \nu_n)^2 \\ &\leq \left( \frac{2}{\rho_n} \dot{\nu}_{n+1} + \frac{(\kappa - 1)_+}{\rho_n \rho_{n-1}} \left( \frac{\nu_n}{\nu_{n-1}} - \frac{\nu_{n+1}}{\nu_n} \right) \right) (e + \nu_n) \tau_n = \beta_n \tau_n. \end{aligned}$$

De plus, comme  $\dot{\vartheta}_0 = \vartheta_0 - \vartheta_{-1}$ , on a

$$(3.3.30) \quad \dot{\vartheta}_0 = \frac{1}{\rho_0} \left( (e + \nu_1)^2 - (e + \nu_0)^2 \right) \leq \frac{2}{\rho_0} \dot{\nu}_1 (e + \nu_1) = \beta_0 \tau_0.$$

Par conséquent, pour  $n \geq 0$ , on obtient en utilisant (3.3.29) et (3.3.30)

$$(3.3.31) \quad \dot{\vartheta}_n \leq \beta_n \tau_n.$$

Ensuite, d'après la définition de  $F_n$  (donnée par (3.3.25)), on obtient (pour  $n \geq 0$ )

$$\begin{aligned} F_n &= s\tau_n \bar{\Theta}(x_n) + \vartheta_n (\bar{\Theta}(x_{n+1}) - \bar{\Theta}(x_n)) \\ &= \vartheta_n \bar{\Theta}(x_{n+1}) - \vartheta_{n-1} \bar{\Theta}(x_n) + (s\tau_n - \dot{\vartheta}_n) \bar{\Theta}(x_n), \end{aligned}$$

ce qui d'après (3.3.31) nous donne (3.3.27). L'inégalité souhaitée découle alors de (3.3.24) et (3.3.27). ■

## 3.4 Analyse de la convergence de la méthode

Dans cette section, nous établissons des estimations et des résultats de convergence concernant l'algorithme (3.1.3)-(3.1.4) avec des paramètres  $(\nu_n)$  généraux et particuliers. La méthodologie repose sur une analyse de Lyapunov associée à la suite  $(\mathcal{E}_n)$  introduite dans (3.3.2).

### 3.4.1 Quelques estimations avec des paramètres généraux.

Un résultat principal de convergence (voir Théorème 3.4.2) est établi concernant l'algorithme (3.1.3)-(3.1.4) avec des paramètres  $(\nu_n)$  généraux. En particulier, des conditions suffisantes pour la convergence de la méthode sont obtenues lorsque  $(\nu_n)$  ne dépasse pas le taux  $\mathcal{O}(n)$ .

Comme établi dans la Proposition 3.2.3, nous rappelons que toutes suites  $\{x_n, y_n, z_n\}$  vérifiant (3.1.3)-(3.1.4) pour  $n \geq 0$  vérifient également (3.2.4)-(3.2.5) pour  $n \geq 1$ . Par conséquent, à l'aide du Lemme 3.3.1, nous estimons des quantités importantes, telles que  $\Theta(x_n) - \min \Theta$ , par rapport à l'énergie  $\mathcal{E}_1(e, q)$ . Un résultat similaire pourrait également être établi pour les itérés donnée par (3.2.4)-(3.2.5) en fonction de l'énergie initiale  $\mathcal{E}_0(e, q)$ , mais nous nous limitons au premier algorithme.

Pour obtenir ces estimations, nous considérons les quantités  $\beta_n$  (introduites dans (3.3.5)-(3.3.6)) et  $\gamma_n$  définies (pour  $n \geq 1$ ) par

$$(3.4.1a) \quad \beta_n = 2\rho_n^{-1}\dot{\nu}_{n+1} + \frac{(\kappa - 1)_+}{\rho_n\rho_{n-1}} \left( \frac{\nu_n}{\nu_{n-1}} - \frac{\nu_{n+1}}{\nu_n} \right) (e + \nu_n),$$

$$(3.4.1b) \quad \gamma_n = (2 - \kappa - L\mu) + (1 - \kappa) \frac{\dot{\nu}_{n+1}}{\nu_n}.$$

Des estimations importantes sont fournies par la proposition suivante.

**Proposition 3.4.1.** *Supposons que (3.1.2) soit vérifié. Soient  $\{\kappa, e, \mu\} \subset ]0, +\infty[$  ainsi que  $(\nu_n)_{n \geq 0}$  vérifiant (3.3.3). Supposons que les conditions suivantes soient satisfaites :*

$$(3.4.2) \quad e - \beta_n \geq 0, \quad \gamma_n \geq 0 \quad (\text{pour } n \geq 1).$$

*Alors, pour toutes suites  $\{z_n, x_n, y_n\}_{n \geq 0} \subset \mathcal{H}$  générées par (3.1.3)-(3.1.4) et pour  $q \in S$ , la suite de type énergie  $(\mathcal{E}_n(e, q))_{n \geq 1}$  (donnée par (3.3.2)) est décroissante. De plus, nous avons les estimations suivantes :*

$$(3.4.3) \quad \Theta(x_n) - \min \Theta \leq \frac{\rho_{n-1}}{\mu(e+\nu_n)^2} \mathcal{E}_1(e, q) \quad (\text{pour } n \geq 2),$$

$$(3.4.4) \quad \sum_{n \geq 1} \gamma_n \rho_n^{-1} (e + \nu_{n+1})^2 \|\dot{x}_{n+1} + \theta_n (y_n - x_n)\|^2 \leq 2\mathcal{E}_1(e, q),$$

$$(3.4.5) \quad \sum_{n \geq 1} (e - \beta_n)(e + \nu_{n+1}) (\Theta(x_n) - \min \Theta) \leq \mu^{-1} \mathcal{E}_1(e, q),$$

$$(3.4.6) \quad \sum_{n \geq 1} \gamma_n \rho_n^{-1} (e + \nu_{n+1})^2 \|M_\mu(z_n)\|^2 \leq 2\mu^{-2} \mathcal{E}_1(e, q),$$

$$(3.4.7) \quad \exists \xi_n^* \in \partial\Theta(x_n) \quad (\text{pour } n \geq 1) \text{ t.q.}$$

$$\sum_{n \geq 1} \gamma_n \rho_n^{-1} (e + \nu_{n+1})^2 \|\xi_n^*\|^2 \leq 2(L + \mu^{-1})^2 \mathcal{E}_1(e, q),$$

où  $M_\mu$ ,  $\rho_n$  et  $\mathcal{E}_1(e, q)$  sont donnés respectivement par (3.2.1), (3.3.1) et (3.3.2).

**Démonstration.** Considérons les itérés  $\{z_n, x_n, y_n\}_{n \geq 0} \subset \mathcal{H}$  générés par (3.1.3)-(3.1.4) et soient  $\{e, \kappa, \mu, \nu_n\}$  des paramètres strictement positifs. Posons  $u_n = y_n - x_n$  et  $\tau_n = e + \nu_{n+1}$ . D'après la Proposition 3.2.3, nous savons que ces itérés vérifient (3.2.4)-(3.2.5) pour  $n \geq 1$ . On applique alors le Lemme 3.3.1 avec  $s = e$  sous les conditions (3.1.2) et (3.3.3). Ainsi, en remarquant que  $\theta_n = \frac{\nu_n \rho_n}{\tau_n}$  et que  $\frac{\nu_{n+1}}{\nu_n} = 1 + \frac{\nu_{n+1}}{\nu_n}$ , on en déduit, pour  $n \geq 1$

$$(3.4.8) \quad \begin{aligned} & \dot{\mathcal{E}}_{n+1}(e, q) + \mu(e - \beta_n) \tau_n \bar{\Theta}(x_n) \\ & \leq -\frac{1}{2} \left( (1 - L\mu) \rho_n^{-1} \tau_n^2 + (1 - \kappa) \frac{\rho_n \nu_n \nu_{n+1}}{\theta_n^2} \right) \|\dot{x}_{n+1} + \theta_n u_n\|^2 \\ & = -\frac{1}{2} \rho_n^{-1} \tau_n^2 \left( (1 - L\mu) + (1 - \kappa) \frac{\nu_{n+1}}{\nu_n} \right) \|\dot{x}_{n+1} + \theta_n u_n\|^2 \\ & = -\frac{1}{2} \rho_n^{-1} \tau_n^2 \left( (2 - \kappa - L\mu) + (1 - \kappa) \frac{\nu_{n+1}}{\nu_n} \right) \|\dot{x}_{n+1} + \theta_n u_n\|^2. \end{aligned}$$

En utilisant la définition de  $\gamma_n$  (donnée par (3.4.1)), l'inégalité précédente se réécrit

$$\mathcal{E}_{n+1}(e, q) - \mathcal{E}_n(e, q) + \frac{1}{2} \gamma_n \rho_n^{-1} \tau_n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2 + \mu(e - \beta_n) \tau_n \bar{\Theta}(x_n) \leq 0.$$

Par conséquent, comme  $\gamma_n \geq 0$  et que  $e - \beta_n \geq 0$  (d'après (3.4.2)), on en déduit que la suite de type énergie  $(\mathcal{E}_n(e, q))_{n \geq 1}$  est décroissante. Ainsi, en sommant l'inégalité précédente pour  $n$  allant de 1 à  $N - 1$  (avec  $N \geq 2$ ), on obtient

$$(3.4.9) \quad \mathcal{E}_N(e, q) + \sum_{n=1}^{N-1} \gamma_n \|\dot{x}_{n+1} + \theta_n u_n\|^2 + \mu \sum_{n=1}^{N-1} \tau_n (e - \beta_n) \bar{\Theta}(x_n) \leq \mathcal{E}_1(e, q).$$

De plus, compte tenu de la positivité de  $\mathcal{E}_n(s, q)$  (d'après (3.3.2)), la dernière inégalité

nous donne (pour  $N \geq 2$ )

$$(3.4.10a) \quad \frac{1}{2} \|e(q - x_N) + \nu_N u_N\|^2 + \mu \frac{(\nu_N + e)^2}{\rho_{N-1}} \bar{\Theta}(x_N) \leq \mathcal{E}_1(e, q),$$

$$(3.4.10b) \quad \sum_{n=1}^{N-1} \gamma_n \rho_n^{-1} \tau_n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2 \leq 2\mathcal{E}_1(e, q),$$

$$(3.4.10c) \quad \mu \sum_{n=1}^{N-1} (e - \beta_n) \tau_n \bar{\Theta}(x_n) \leq \mathcal{E}_1(s, q),$$

Les inégalités ci-dessus conduisent directement aux estimations (3.4.3)-(3.4.5).

Il reste à prouver (3.4.6) et (3.4.7) :

- L'estimation (3.4.6) est immédiatement déduite de la relation  $M_\mu(z_n) = -\mu^{-1}(\dot{x}_{n+1} + \theta_n u_n)$  (d'après (3.3.16)) et de (3.4.4).
- L'égalité (3.3.14), nous donne  $z_n - x_{n+1} = \mu(x_{n+1}^* + \nabla f(z_n))$  (pour un certain  $x_{n+1}^* \in \partial g(x_{n+1})$ ). Ainsi, comme  $\nabla f$  est  $L$ -lipschitzien (d'après la condition (3.1.2)), on obtient

$$\|x_{n+1}^* + \nabla f(x_{n+1})\| \leq \|x_{n+1}^* + \nabla f(z_n)\| + L\|z_n - x_{n+1}\|.$$

Par conséquent, en combinant ces deux résultats, il vient

$$\|x_{n+1}^* + \nabla f(x_{n+1})\| \leq (1 + L\mu) \|x_{n+1}^* + \nabla f(z_n)\|.$$

De plus, comme  $M_\mu(z_n) = x_{n+1}^* + \nabla f(z_n)$  (d'après (3.3.15)), on en déduit que

$$\|x_{n+1}^* + \nabla f(x_{n+1})\|^2 \leq (1 + L\mu)^2 \|M_\mu(z_n)\|^2,$$

ce qui, combiné avec (3.4.6), nous donne (3.4.7). ■

Exposons à présent le résultat principal de cette section.

**Proposition 3.4.2.** *Supposons que (3.1.2) soit vérifié. Soient  $\{x_n, y_n\} \subset \mathcal{H}$  des suites générées par (3.1.3)-(3.1.4) avec  $\kappa \in ]0, 2[$ , ainsi que des paramètres strictement positifs  $\{\mu, e, \nu_n\}$  vérifiant les conditions ci-dessous*

$$(3.4.11a) \quad (\dot{\nu}_n)_{n \geq 0} \subset [\bar{\nu}_m^*, \bar{\nu}_M^*] \quad (\text{avec } \bar{\nu}_m^* > 0 \text{ et } \bar{\nu}_M^* > 0),$$

$$(3.4.11b) \quad (\nu_{n+1}\nu_n^{-1})_{n \geq 0} \text{ est décroissante} \quad (\text{si } \kappa \in ]0, 1]),$$

$$(3.4.11c) \quad (1 - \kappa)\nu_{n+1}\nu_n^{-1} < 1 \quad (\text{pour } n \geq 0)$$

$$(3.4.11d) \quad (\nu_n)_{n \geq 0} \subset ](\kappa - 1)_+e, +\infty[,$$

$$(3.4.11e) \quad e > \sup_{n \geq 1} \beta_n.$$

Supposons en outre que l'une des conditions suivantes soit satisfaite

$$(3.4.12a) \quad \kappa = 1 \quad \text{et} \quad \mu \in ]0, \frac{1}{L}],$$

$$(3.4.12b) \quad \kappa \in ]0, 1[ \quad \text{et} \quad \mu \in ]0, \frac{2-\kappa}{L}[,$$

$$(3.4.12c) \quad \kappa \in ]1, 2[, \mu \in ]0, \frac{2-\kappa}{L}[, \text{ et } \nu_0 > \frac{(\kappa-1)_+ \bar{\nu}_M^*}{2-\kappa-\mu L}.$$

Alors, pour  $q \in S$ , la suite de type énergie  $(\mathcal{E}_n(e, q))_{n \geq 1}$  (donnée par (3.3.2)) est décroissante. De plus, les suites  $(x_n)$  et  $(y_n)$  convergent faiblement vers un élément  $\bar{z} \in S$  et nous avons :

$$(3.4.13a) \quad \Theta(x_n) - \min \Theta = o(n^{-2}), \quad \sum_n n (\Theta(x_n) - \min \Theta) < +\infty,$$

$$(3.4.13b) \quad \sum_n n \|\dot{x}_n\|^2 < +\infty, \quad \|\dot{x}_n\| = o(n^{-1}),$$

$$(3.4.13c) \quad \sum_n n \|\dot{y}_n\|^2 < +\infty, \quad \|\dot{y}_n\| = o(n^{-1}).$$

En particulier, nous obtenons les résultats supplémentaires suivants en fonction de  $\kappa$  et  $\mu$  :

- Si  $\kappa \in ]0, 1]$  et  $\mu \in ]0, \frac{2-\kappa}{L}]$ , alors

$$(3.4.14) \quad \Theta(y_n) - \min \Theta = o(n^{-2}), \quad \sum_n n (\Theta(y_n) - \min \Theta) < +\infty.$$

- Si  $\kappa \in ]0, 2[$  et  $\mu \in ]0, \frac{2-\kappa}{L}[,$  alors

$$(3.4.15a) \quad \exists \xi_n^* \in \partial \Theta(x_n) \quad \text{t.q.} \quad \sum_n n^2 \|\xi_n^*\|^2 < +\infty.$$

**Démonstration.** Nous savons, d'après (3.4.11a), qu'il existe deux constantes stric-

tement positives  $\bar{\nu}_m^*$  et  $\bar{\nu}_M^*$  telles que

$$(3.4.16) \quad \forall n \geq 0, \quad \bar{\nu}_m^* n + \nu_0 \leq \nu_n \leq \bar{\nu}_M^* n + \nu_0.$$

On en déduit alors que

$$(3.4.17a) \quad \nu_n \rightarrow +\infty \quad (\text{lorsque } n \rightarrow +\infty),$$

$$(3.4.17b) \quad \nu_n = \mathcal{O}(n),$$

$$(3.4.17c) \quad \nu_n^{-1} = \mathcal{O}(n^{-1}).$$

De plus, en utilisant à nouveau la condition (3.4.11a), nous avons pour  $n \geq 0$ ,

$$(3.4.18) \quad \frac{\bar{\nu}_m^*}{\nu_n} \leq \frac{\nu_{n+1}}{\nu_n} - 1 \leq \frac{\bar{\nu}_M^*}{\nu_n}.$$

Ainsi en combinant (3.4.17) et (3.4.18) nous obtenons

$$(3.4.19) \quad \lim_{n \rightarrow +\infty} \frac{\nu_{n+1}}{\nu_n} = 1.$$

Par conséquent, les estimations (3.4.17) et (3.4.19), nous donne, par définition de  $(\theta_n)$  et de  $(\rho_n)$  (introduits dans (3.2.5) et (3.3.1)),

$$(3.4.20) \quad \lim_{n \rightarrow +\infty} \theta_n = \lim_{n \rightarrow +\infty} \rho_n = \kappa > 0.$$

Par ailleurs, comme  $\dot{\nu}_n \leq \bar{\nu}_M^*$  (d'après (3.4.11a)), la définition de  $\gamma_n$  nous donne

$$(3.4.21) \quad \gamma_n \geq 2 - \kappa - \mu L - (\kappa - 1)_+ \frac{\bar{\nu}_M^*}{\nu_n}.$$

On déduit alors de cette dernière inégalité que

$$(3.4.22a) \quad \kappa \in ]0, 1] \quad \text{et} \quad \mu \in ]0, \frac{2-\kappa}{L}] \quad \Rightarrow \quad \inf_{n \geq 0} \gamma_n \geq 0,$$

$$(3.4.22b) \quad \kappa \in ]0, 1] \quad \text{et} \quad \mu \in ]0, \frac{2-\kappa}{L}[ \quad \Rightarrow \quad \inf_{n \geq 0} \gamma_n > 0,$$

$$(3.4.22c) \quad \kappa \in ]1, 2[, \mu \in ]0, \frac{2-\kappa}{L}[ \quad \text{et} \quad \nu_0 > \frac{(\kappa-1)_+ \bar{\nu}_M^*}{2-\kappa-\mu L} \quad \Rightarrow \quad \inf_{n \geq 0} \gamma_n > 0.$$

À présent, pour alléger les notations, nous posons  $\bar{\Theta} = \Theta - \min \Theta$ ,  $u_n = y_n - x_n$ , et nous procédons à la démonstration des résultats considérés en distinguant les parties

(C0)-(C4) suivantes :

(C0) Considérons les plages de paramètres suivantes :

$$(3.4.23a) \quad (\mathbf{ha}) : \quad \kappa \in ]0, 1] \quad \text{et} \quad \mu \in ]0, \frac{2-\kappa}{L}],$$

$$(3.4.23b) \quad (\mathbf{hb}) : \quad \kappa \in ]1, 2[, \mu \in ]0, \frac{2-\kappa}{L}[ \quad \text{et} \quad \nu_0 > \frac{(\kappa-1)+\bar{\nu}_M^*}{2-\kappa-\mu L}.$$

Montrons que si l'une des conditions **(ha)** ou **(hb)** est vérifiée alors les propriétés suivantes sont obtenues :

$$(3.4.24a) \quad (\mathcal{E}_n(e, q))_{n \geq 1} \text{ (donnée par (3.3.2)) est décroissante;}$$

$$(3.4.24b) \quad \bar{\Theta}(x_n) = \mathcal{O}(n^{-2});$$

$$(3.4.24c) \quad \sum_n \gamma_n n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2 < +\infty;$$

$$(3.4.24d) \quad \sum_n n \bar{\Theta}(x_n) < +\infty;$$

$$(3.4.24e) \quad \sum_n \gamma_n n^2 \|M_\mu(z_n)\|^2 < +\infty;$$

$$(3.4.24f) \quad \exists \xi_n^* \in \partial\Theta(x_n) \text{ t.q. } \sum_n \gamma_n n^2 \|\xi_n^*\|^2 < +\infty.$$

En effet, sous les conditions **(ha)** et **(hb)**, les éléments (3.4.24a)-(3.4.24f) proviennent directement de la Proposition 3.4.1, ainsi que des conditions (3.4.11),  $\inf_{n \geq 0} \gamma_n \geq 0$  (d'après (3.4.22), et des estimations  $\nu_n^{-1} = \mathcal{O}(n^{-1})$  (d'après (3.4.17) et  $\rho_n \sim \kappa$  (lorsque  $n \rightarrow +\infty$ ).

(C1) Considérons les plages de paramètres suivantes

$$(\mathbf{ha1}) : \quad \kappa \in ]0, 1] \quad \text{et} \quad \mu \in ]0, \frac{2-\kappa}{L}[,$$

$$(\mathbf{hb}) : \quad \kappa \in ]1, 2[, \mu \in ]0, \frac{2-\kappa}{L}[ \quad \text{et} \quad \nu_0 > \frac{(\kappa-1)+\bar{\nu}_M^*}{2-\kappa-\mu L}.$$

Montrons que, sous chacune des conditions **(ha1)** et **(hb)**, nous avons :

$$(3.4.26a) \quad \sum_n n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2 < +\infty;$$

$$(3.4.26b) \quad \sum_n n^2 \|M_\mu(z_n)\|^2 < +\infty;$$

$$(3.4.26c) \quad \exists \xi_n^* \in \partial\Theta(x_n) \text{ t.q. } \sum_n n^2 \|\xi_n^*\|^2 < +\infty.$$

En effet, d'après (3.4.22), on a  $\inf_{n \geq 0} \gamma_n > 0$ . Ainsi, en utilisant **(ha1)**-**(hb)**, ainsi que (3.4.24c), (3.4.24e) et (3.4.24f), nous obtenons directement (3.4.26a), (3.4.26b) et (3.4.26c).

(C2) Considérons les plages de paramètres suivantes :

$$\text{(ha2)} : \quad \kappa = 1 \quad \text{et} \quad \mu \in ]0, \frac{1}{L}],$$

$$\text{(ha3)} : \quad \kappa \in ]0, 1[ \quad \text{et} \quad \mu \in ]0, \frac{2-\kappa}{L}[,$$

$$\text{(hb)} : \quad \kappa \in ]1, 2[, \mu \in ]0, \frac{2-\kappa}{L}[ \quad \text{et} \quad \nu_0 > \frac{(\kappa-1)+\bar{\nu}_M^*}{2-\kappa-\mu L}.$$

Montrons que si l'une des conditions **(ha2)**, **(ha3)** ou **(hb)** est vérifiée, alors :

$$(3.4.28a) \quad (\mathcal{E}_n(s, q))_n \text{ est convergente, pour tout } (s, q) \in [\bar{\beta}, e[ \times S \text{ où } \bar{\beta} = \sup_{j \geq 1} \beta_j;$$

$$(3.4.28b) \quad \sum_n n \|\dot{x}_n\|^2 < +\infty;$$

$$(3.4.28c) \quad \sum_n n \|u_n\|^2 < +\infty \quad \Rightarrow \quad \lim_{n \rightarrow +\infty} n^2 \bar{\Theta}(x_n) = \lim_{n \rightarrow +\infty} n \|u_n\| = 0.$$

En effet, en prenant  $(s, q) \in ]0, e[ \times S$ , et en utilisant le Lemme 3.3.1 sous les conditions (3.4.11) et (3.4.22), on en déduit, pour  $n \geq 1$ , que

$$(3.4.29) \quad \begin{aligned} & \dot{\mathcal{E}}_{n+1}(s, q) + \mu(s - \beta_n) \tau_n \bar{\Theta}(x_n) \\ & \leq -\frac{1}{2}(1 - \kappa) \rho_n \nu_n \nu_{n+1} \left\| u_n + \left( \frac{1}{\theta_n} - \frac{e-s}{\nu \rho_n} \right) \dot{x}_{n+1} \right\|^2 \\ & \quad - \frac{1}{2}(1 - L\mu) \rho_n^{-1} \tau_n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2 - \frac{1}{2}(e - s) h_n(s) \|\dot{x}_{n+1}\|^2, \end{aligned}$$

où  $\tau_n = \nu_{n+1} + e$  et où  $h_n(s)$  est défini par

$$(3.4.30) \quad h_n(s) := (s + 2\nu_{n+1}) \left( 1 + \frac{(1-\kappa)\nu_{n+1}}{\nu_n \rho_n} \right) + s \frac{(1-\kappa)\nu_{n+1}}{\nu_n \rho_n}.$$

Estimons le premier terme à droite de l'inégalité (3.4.29). En rappelant que  $\theta_n = \frac{\nu_n \rho_n}{\tau_n}$ , nous obtenons

$$\begin{aligned} & \rho_n \nu_n \nu_{n+1} \left\| u_n + \left( \frac{1}{\theta_n} - \frac{e-s}{\nu \rho_n} \right) \dot{x}_{n+1} \right\|^2 \\ & \leq 2 \frac{\rho_n \nu_n \nu_{n+1}}{\theta_n^2} \|\dot{x}_{n+1} + \theta_n u_n\|^2 + 2(e - s)^2 \frac{\nu_{n+1}}{\nu_n \rho_n} \|\dot{x}_{n+1}\|^2 \\ & = 2 \rho_n^{-1} \frac{\nu_{n+1}}{\nu_n} \tau_n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2 + 2(e - s)^2 \rho_n^{-1} \frac{\nu_{n+1}}{\nu_n} \|\dot{x}_{n+1}\|^2. \end{aligned}$$

Ce qui d'après (3.4.29) nous donne

$$\begin{aligned}
& \dot{\mathcal{E}}_{n+1}(s, q) + \mu(s - \beta_n) \tau_n \bar{\Theta}(x_n) \\
& \leq \frac{1}{2}(\kappa - 1)_+ \rho_n \nu_n \nu_{n+1} \left\| u_n + \left( \frac{1}{\theta_n} - \frac{e-s}{\nu \rho_n} \right) \dot{x}_{n+1} \right\|^2 \\
& \quad + \frac{1}{2}(L\mu - 1)_+ \rho_n^{-1} \tau_n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2 - \frac{1}{2}(e - s) h_n(s) \|\dot{x}_{n+1}\|^2 \\
& \leq \frac{1}{2}(\kappa - 1)_+ \left( 2\rho_n^{-1} \frac{\nu_{n+1}}{\nu_n} \tau_n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2 + 2(e - s)^2 \rho_n^{-1} \frac{\nu_{n+1}}{\nu_n} \|\dot{x}_{n+1}\|^2 \right) \\
& \quad + \frac{1}{2}(L\mu - 1)_+ \rho_n^{-1} \tau_n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2 - \frac{1}{2}(e - s) h_n(s) \|\dot{x}_{n+1}\|^2 \\
& = \left( (\kappa - 1)_+ \frac{\nu_{n+1}}{\nu_n} + \frac{1}{2}(L\mu - 1)_+ \right) \rho_n^{-1} \tau_n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2 \\
& \quad - \frac{1}{2}(e - s) \left( h_n(s) - 2(\kappa - 1)_+ (e - s) \rho_n^{-1} \frac{\nu_{n+1}}{\nu_n} \right) \|\dot{x}_{n+1}\|^2.
\end{aligned}$$

Introduisons maintenant les nouveaux paramètres suivants :

$$(3.4.31a) \quad \bar{\beta} = \sup_{j \geq 1} \beta_j,$$

$$(3.4.31b) \quad l_n = \left( (\kappa - 1)_+ \frac{\nu_{n+1}}{\nu_n} + \frac{1}{2}(L\mu - 1)_+ \right) \rho_n^{-1} \tau_n^2,$$

$$(3.4.31c) \quad r_n = 1 + 2\rho_n^{-1} \frac{\nu_{n+1}}{\nu_n} (1 - \kappa)_+.$$

On remarque, à partir de la définition de  $h_n(\cdot)$  (donnée par (3.4.30)) que l'on a

$$(3.4.32) \quad h_n(s) - h_n(e) = -(e - s) \left( 1 + 2\rho_n^{-1} \frac{(1 - \kappa)\nu_{n+1}}{\nu_n} \right),$$

ce qui nous donne,

$$\begin{aligned}
(3.4.33) \quad & h_n(s) - h_n(e) - (e - s)(r_n - 1) \\
& = -(e - s) \left( 1 + 2\rho_n^{-1} \frac{\nu_{n+1}}{\nu_n} (1 - \kappa + (\kappa - 1)_+) \right) \\
& = -(e - s) \left( 1 + 2\rho_n^{-1} \frac{\nu_{n+1}}{\nu_n} (1 - \kappa)_+ \right),
\end{aligned}$$

ou, de façon équivalente (à l'aide de (3.4.31c))

$$(3.4.34) \quad h_n(s) = h_n(e) - (e - s)r_n + (e - s)(r_n - 1).$$

Par conséquent, en tenant compte de (3.4.31a) et de (3.4.34), la précédente estimation de  $\dot{\mathcal{E}}_{n+1}(s, q)$  implique que

$$\begin{aligned}
 (3.4.35) \quad & \dot{\mathcal{E}}_{n+1}(s, q) + \mu (s - \bar{\beta}) \tau_n \bar{\Theta}(x_n) \\
 & \leq l_n \|\dot{x}_{n+1} + \theta_n u_n\|^2 - \frac{1}{2} (e - s) (h_n(s) - (e - s)(r_n - 1)) \|\dot{x}_{n+1}\|^2 \\
 & = l_n \|\dot{x}_{n+1} + \theta_n u_n\|^2 - \frac{1}{2} (e - s_0) (h_n(e) - (e - s_0)r_n) \|\dot{x}_{n+1}\|^2.
 \end{aligned}$$

Ainsi, en remarquant que  $\bar{\beta} < e$  (d'après (3.4.11e)) et en prenant  $s = s_0$  avec  $s_0 \in [\bar{\beta}, e[$ , on en déduit que

$$\begin{aligned}
 (3.4.36) \quad & \dot{\mathcal{E}}_{n+1}(s_0, q) + \mu (s_0 - \bar{\beta}) \tau_n \bar{\Theta}(x_n) \\
 & \leq l_n \|\dot{x}_{n+1} + \theta_n u_n\|^2 - \frac{1}{2} (e - s_0) (h_n(e) - (e - s_0)r_n) \|\dot{x}_{n+1}\|^2.
 \end{aligned}$$

On peut également remarquer, à l'aide des définitions de  $h_n(\cdot)$ ,  $l_n$  et  $r_n$  (données par (3.4.30) et (3.4.31)), et par le fait que  $\nu_n \rightarrow +\infty$ ,  $\frac{\nu_{n+1}}{\nu_n} \rightarrow 1$ ,  $\rho_n \rightarrow \kappa$  et  $\nu_n = \mathcal{O}(n)$  lorsque  $n \rightarrow +\infty$  (d'après (3.4.19) et (3.4.17)), que l'on a

$$(3.4.37a) \quad h_n(e) \sim \frac{2}{\kappa} \nu_{n+1} \quad \text{lorsque } n \rightarrow +\infty \quad (\text{d'où } h_n(e) = \mathcal{O}(n)),$$

$$(3.4.37b) \quad r_n \sim 1 + 2\kappa^{-1}(1 - \kappa)_+ \quad \text{lorsque } n \rightarrow +\infty,$$

$$(3.4.37c) \quad l_n = \mathcal{O}(n^2).$$

De plus, concernant le terme  $h_n(e) - (e - s_0)r_n$  intervenant dans le côté droit de (3.4.36), nous avons, en utilisant (3.4.17) et (3.4.37a)-(3.4.37b),

$$(3.4.38) \quad \nu_n^{-1} = \mathcal{O}(n^{-1}) \quad \text{et} \quad h_n(e) - (e - s_0)r_n \sim \frac{2}{\kappa} \nu_{n+1} \quad (\text{lorsque } n \rightarrow +\infty).$$

(a)-(b) Nous sommes maintenant en mesure de démontrer les estimations (3.4.28a)-(3.4.28b) en considérant séparément les deux plages de paramètres suivantes :

- Supposons que **(ha2)** soit vérifié (donc  $l_n = 0$  d'après sa définition).

L'inégalité (3.4.36), nous donne alors

$$(3.4.39) \quad \dot{\mathcal{E}}_{n+1}(s_0, q) \leq -\frac{1}{2} (e - s_0) (h_n(e) - (e - s_0)r_n) \|\dot{x}_{n+1}\|^2.$$

Ainsi, pour  $N \geq 2$  on obtient

$$(3.4.40) \quad \frac{1}{2} (e - s_0) \sum_{n=1}^{N-1} (h_n(e) - (e - s_0)r_n) \|\dot{x}_{n+1}\|^2 \leq \mathcal{E}_1(s_0, q).$$

Par conséquent, on déduit de (3.4.39) et de (3.4.38) que la suite  $(\mathcal{E}_n(s_0, q))$  est convergente. De plus, en utilisant (3.4.38) et (3.4.40), on obtient  $\sum_n n \|\dot{x}_n\|^2 < +\infty$ .

- Supposons que **(ha3)** ou **(hb)** soit vérifié. D'après (3.4.36), nous avons

$$(3.4.41) \quad \dot{\mathcal{E}}_{n+1}(s_0, q) \leq l_n \|\dot{x}_{n+1} + \theta_n u_n\|^2 - \frac{1}{2} (e - s_0) (h_n(e) - (e - s_0) r_n) \|\dot{x}_{n+1}\|^2,$$

et donc, on obtient pour  $N \geq 2$

$$(3.4.42) \quad \begin{aligned} \frac{1}{2} (e - s_0) \sum_{n=1}^{N-1} (h_n(e) - (e - s_0) r_n) \|\dot{x}_{n+1}\|^2 \\ \leq \mathcal{E}_1(s_0, q) + \sum_{n=1}^{N-1} l_n \|\dot{x}_{n+1} + \theta_n u_n\|^2. \end{aligned}$$

De plus, sachant que  $l_n = \mathcal{O}(n^2)$  (d'après (3.4.37c)) et que  $\sum_n n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2 < +\infty$  (d'après (3.4.26a)) on obtient

$$(3.4.43) \quad \sum_n l_n \|\dot{x}_{n+1} + \theta_n u_n\|^2 < +\infty.$$

On déduit alors des estimations (3.4.41) et (3.4.43) que  $(\mathcal{E}_n(s_0, q))$  est convergent. De plus, il résulte de (3.4.38) et de (3.4.42) que l'on a  $\sum_n n \|\dot{x}_n\|^2 < +\infty$

(c) Prouvons (3.4.28c). Pour  $n \geq 1$ , nous avons à partir de (3.3.2)

$$(3.4.44) \quad \mathcal{E}_n(0, q) = \frac{1}{2} \|\nu_n u_n\|^2 + \mu \frac{(\nu_n + e)^2}{\rho_{n-1}} \bar{\Theta}(x_n).$$

De plus, en utilisant (3.4.35) avec  $s \in ]0, e[$  et  $q \in S$ , et en passant à la limite lorsque  $s \rightarrow 0^+$  dans cette inégalité, on obtient

$$(3.4.45) \quad \begin{aligned} \mathcal{E}_{n+1}(0, q) - \mathcal{E}_n(0, q) \\ \leq l_n \|\dot{x}_{n+1} + \theta_n u_n\|^2 + \mu \bar{\beta} \tau_n \bar{\Theta}(x_n) - \frac{1}{2} e (h_n(e) - e r_n) \|\dot{x}_{n+1}\|^2. \end{aligned}$$

Par ailleurs, les estimations de (3.4.37) nous donnent

$$(3.4.46) \quad h_n(e) - e r_n \sim \frac{2}{\kappa} \nu_{n+1} \geq 0 \quad (\text{lorsque } n \rightarrow +\infty).$$

De plus, comme  $\sum_n n \bar{\Theta}(x_n) < +\infty$  (d'après (3.4.24d)),  $\tau_n = \mathcal{O}(n)$  (d'après (3.4.17))

et  $\tau_n = e + \nu_{n+1}$ ), on a aussi

$$(3.4.47) \quad \sum_n \tau_n \bar{\Theta}(x_n) < +\infty,$$

tandis que (comme indiqué dans (3.4.43)) chacune des conditions **(ha3)** ou **(hb)** garantit que

$$(3.4.48) \quad \sum_n l_n \|\dot{x}_{n+1} + \theta_n u_n\|^2 < +\infty.$$

Il en résulte d'après (3.4.45), (3.4.46), (3.4.47) et (3.4.48), que la suite  $(\mathcal{E}_n(0, q))$  converge lorsque  $n \rightarrow +\infty$ . De plus, en supposant que  $\sum_n n \|u_n\|^2 < +\infty$ , on remarque, d'après  $\nu_n = \mathcal{O}(n)$  et la formulation de  $\mathcal{E}_n(0, q)$  ci-dessus, que

$$(3.4.49) \quad \sum_n n^{-1} \mathcal{E}_n(0, q) < +\infty.$$

Par conséquent, sachant que la suite  $(\mathcal{E}_n(0, q))_n$  est convergente et en remarquant que  $\sum_n n^{-1} = +\infty$ , on en déduit que  $\mathcal{E}_n(0, q) \rightarrow 0$  lorsque  $n \rightarrow +\infty$ . Ce dernier résultat associé à  $\dot{y}_{n+1} = -\kappa u_n$  (d'après la Proposition 3.2.3), entraîne immédiatement (3.4.28c).

**(C3)** Prouvons que si l'une des conditions précédentes **(ha2)**, **(ha3)** ou **(hb)** (utilisées dans la partie **(C2)**) est vérifiée, alors nous avons :

$$(3.4.50a) \quad \sum_n n \|y_n - x_n\|^2 < +\infty, \quad \sum_n n \|\dot{y}_n\|^2 < +\infty,$$

$$(3.4.50b) \quad \lim_{n \rightarrow +\infty} n^2 \bar{\Theta}(x_n) = 0,$$

$$(3.4.50c) \quad \lim_{n \rightarrow +\infty} n \|y_n - x_n\| = \lim_{n \rightarrow +\infty} n \|\dot{y}_n\| = 0,$$

$$(3.4.50d) \quad \lim_{n \rightarrow +\infty} n \|\dot{x}_n\| = 0.$$

Pour simplifier, nous posons  $u_n = y_n - x_n$  et nous prouvons chaque élément (a)-(b)-(c)-(d) séparément :

(a) Commençons par prouver la première estimation dans (3.4.50a). En utilisant la convexité de la norme carrée, on a

$$(3.4.51) \quad n \|\theta_n u_n\|^2 \leq 2n \|\theta_n u_n + \dot{x}_{n+1}\|^2 + 2n \|\dot{x}_{n+1}\|^2.$$

Distinguons alors les deux situations suivantes :

- Supposons que **(ha3)** ou **(hb)** soit vérifiée. Alors comme  $\sum_n n \|\theta_n u_n + \dot{x}_{n+1}\|^2 < +\infty$  et  $\sum_n n \|\dot{x}_{n+1}\|^2 < +\infty$  (d'après (3.4.26a) et (3.4.28b) respectivement), on obtient  $\sum_n n \|\theta_n u_n\|^2 < +\infty$ . Ainsi, comme  $\theta_n \rightarrow \kappa$  lorsque  $n \rightarrow +\infty$ , on en déduit que  $\sum_n n \|u_n\|^2 < +\infty$ .

- Supposons que **(ha2)** soit vérifiée (donc  $\kappa = 1$ ). D'après (3.2.7c) de la Proposition 3.2.3, on a  $y_{n+1} = x_n$  pour  $n \geq 1$  (car  $\dot{y}_{n+1} = x_n - y_n$ ). Donc comme  $u_n = y_n - x_n$ , on obtient  $u_n = -(x_n - x_{n-1})$  pour  $n \geq 2$ , ce qui nous donne  $\|u_n\|^2 = \|\dot{x}_n\|^2$ . Ainsi, d'après (3.4.28b) nous obtenons  $\sum_n n \|u_n\|^2 < +\infty$ .

La deuxième estimation de (3.4.50a) est directement déduite de la première avec  $\dot{y}_{n+1} = -\kappa u_n$  (d'après la Proposition 3.2.3).

(b)-(c) Les éléments (3.4.50b) et (3.4.50c) découlent de (3.4.28c), en tenant compte de (3.4.50a) et en rappelant que  $\dot{y}_{n+1} = -\kappa u_n$  (d'après la Proposition 3.2.3).

(d) Pour prouver ce point, nous considérons les deux situations suivantes :

- Supposons que **(ha3)** ou **(hb)** soit vérifié. Alors comme  $n \|\dot{x}_{n+1} + \theta_n u_n\| \rightarrow 0$  (d'après (3.4.26a)) et que  $n \|u_n\| \rightarrow 0$  (d'après (3.4.50c) lorsque  $n \rightarrow +\infty$ , on en déduit,  $n \|\dot{x}_{n+1}\| \rightarrow 0$  lorsque  $n \rightarrow +\infty$  (car  $(\theta_n)$  est borné).

- Supposons que **(ha2)** soit vérifié. Comme  $\kappa = 1$ , on obtient en procédant comme précédemment, que  $u_n = -\dot{x}_n$  (pour  $n \geq 2$ ). Ainsi, on déduit de (3.4.50c), que l'on a  $n \|\dot{x}_n\| \rightarrow 0$  lorsque  $n \rightarrow +\infty$ .

**(C4)** Montrons que toutes suites  $\{x_n, y_n\} \subset \mathcal{H}$  générées par (3.1.3)-(3.1.4) convergent faiblement vers un élément de  $S$  lorsque  $n \rightarrow +\infty$ .

En effet, pour  $(s, q) \in [\bar{\beta}, e[$ , nous savons, d'après (3.4.28a), que  $(\mathcal{E}_n(s, q))$  est convergent lorsque  $n \rightarrow +\infty$ . De plus, par (3.4.50b)-(3.4.50c) et  $\nu_n = \mathcal{O}(n)$  nous avons

$$(3.4.52) \quad \frac{(e + \nu_n)^2}{\rho_{n-1}} \bar{\Theta}(x_n) \rightarrow 0 \quad \text{et} \quad \nu_n \|y_n - x_n\| \rightarrow 0 \quad (\text{lorsque } n \rightarrow +\infty).$$

Par conséquent, on en déduit de la définition de  $\mathcal{E}_n(s, q)$  (donnée par (3.3.2)), que

$$(3.4.53) \quad \lim_{n \rightarrow +\infty} \mathcal{E}_n(s, q) = \left( \frac{s^2}{2} + \left( \frac{1}{2} \right) s(e - s) \right) \lim_{n \rightarrow +\infty} \|x_n - q\|^2.$$

Il en résulte que  $(\|x_n - q\|)$  converge lorsque  $n \rightarrow +\infty$ . De plus, l'estimation (3.4.50b) assure que  $\Theta(x_n) - \Theta(q) \rightarrow 0$  lorsque  $n \rightarrow +\infty$ . Par ailleurs, pour toute sous-suite de

$(x_n)$  (notée encore  $(x_n)$ ) convergeant faiblement vers un certain  $x_* \in \mathcal{H}$ , on obtient, en invoquant la semi-continuité inférieure de  $\Theta$ ,

$$(3.4.54) \quad 0 = \liminf_{n \rightarrow +\infty} (\Theta(x_n) - \Theta(q)) \geq (\Theta(x_*) - \Theta(q)) \geq 0,$$

ce qui implique que  $\Theta(x_*) = \Theta(q)$ . Ceci prouve que  $x_* \in S$ , assurant ainsi que tout point d'accumulation faible de  $(x_n)$  appartient à  $S$ . En appliquant le lemme d'Opial [87] nous concluons que la suite  $(x_n)$  converge faiblement vers un élément  $\bar{z} \in S$ . Par conséquent, comme  $\|x_n - y_n\| \rightarrow 0$  lorsque  $n \rightarrow +\infty$  (d'après (3.4.50c)), nous déduisons notre résultat de convergence.

**(C5) Convergence de  $\Theta(y_n)$ .** Démontrons que, les propriétés suivantes sont obtenues sous chacune des conditions **(ha2)** et **(ha3)**, :

$$(3.4.55a) \quad \sum_n n(\Theta(y_n) - \min \Theta) < +\infty;$$

$$(3.4.55b) \quad \lim_n n^2(\Theta(y_n) - \min \Theta) = 0.$$

- Prouvons (3.4.55a). D'après (3.1.3), nous avons (pour  $n \geq 0$ )

$$(3.4.56) \quad y_{n+1} = (1 - \kappa)y_n + \kappa x_n.$$

Donc, comme  $\kappa \in ]0, 1]$ , en utilisant la convexité de la fonction objective  $\Theta$ , on en déduit que

$$(3.4.57) \quad \Theta(y_{n+1}) \leq (1 - \kappa)\Theta(y_n) + \kappa\Theta(x_n).$$

Cette dernière inégalité peut se réécrire sous la forme

$$(3.4.58) \quad \Gamma_{n+1} \leq (1 - \kappa)\Gamma_n + \kappa\Delta_n,$$

avec  $\Gamma_n = \Theta(y_n) - \min \Theta$  et  $\Delta_n = \Theta(x_n) - \min_{\mathcal{H}} \Theta$ .

En multipliant (3.4.58) par  $(n + 1)$ , on a

$$(3.4.59) \quad (n + 1)\Gamma_{n+1} - n\Gamma_n \leq -(\kappa(n + 1) - 1)\Gamma_n + \kappa(n + 1)\Delta_n.$$

Par conséquent, en utilisant (3.4.58)-(3.4.59) ainsi que  $\sum_n n\Delta_n < +\infty$  (d'après (3.4.24d)), on obtient  $\sum_n n\Gamma_n < +\infty$ , ce qui nous donne (3.4.55a).

- Prouvons (3.4.55b). En multipliant l'inégalité (3.4.58) par  $(n+1)^2$  on obtient

$$(3.4.60) \quad (n+1)^2 \Gamma_{n+1} \leq (1-\kappa)n^2 \Gamma_n + 2(1-\kappa)(n+1)\Gamma_n + \kappa(n+1)^2 \Delta_n.$$

Rappelons à présent le lemme suivant (voir [73, Lemma 4.3] pour une preuve),

**Lemme 3.4.2.** *Soit  $(a_n)$  une suite de réels positifs telle que*

$$(3.4.61) \quad a_{n+1} \leq (1-\gamma_n)a_n + \gamma_n r_n,$$

où  $(r_n)$  est une suite réelle bornée et où  $(\gamma_n) \subset [0, 1]$  vérifie  $\sum_n \gamma_n = +\infty$ .

Alors on a

$$(3.4.62) \quad \limsup_{n \rightarrow +\infty} a_n \leq \limsup_{n \rightarrow +\infty} r_n.$$

Par conséquent, nous obtenons à l'aide du lemme précédent,

$$(3.4.63) \quad \limsup_{n \rightarrow +\infty} n^2 \Gamma_n \leq \limsup_{n \rightarrow +\infty} \kappa^{-1}(2(1-\kappa)(n+1)\Gamma_n + \kappa(n+1)^2 \Delta_n).$$

Ainsi, comme  $\lim_{n \rightarrow +\infty} \nu_n^2 \Delta_n = 0$  (d'après (3.4.50b)) et  $\sum_{n \geq 0} n \Gamma_n < +\infty$  (d'après (3.4.55a)), on déduit de (3.4.63) que  $\lim_{n \rightarrow +\infty} n^2 \Gamma_n = 0$ , c'est à dire (3.4.55b). ■

Quelques inégalités sont données par les résultats suivants afin d'obtenir des estimations sur les itérés  $\{z_n, y_n, x_n\}$  générés par (3.1.3)-(3.1.4).

**Lemme 3.4.3.** *Supposons que (3.1.2) soit vérifiée et soient  $\{z_n, x_n, y_n\} \subset \mathcal{H}$  générées par (3.1.3)-(3.1.4) avec  $\kappa \in ]0, 2[$ ,  $\mu \in ]0, \frac{2-\kappa}{L}[$ ,  $e > 0$ . Soit  $(\nu_n)$  une suite strictement positive. Alors pour  $n \geq 1$ , on a les inégalités suivantes :*

$$(3.4.64a) \quad \|\dot{y}_{n+1} - \dot{y}_n\|^2 \leq 2 \left( \frac{\kappa e + \nu_n}{e + \nu_n} \right)^2 \|\dot{y}_n\|^2 + 2(\kappa\mu)^2 \|M_\mu(z_{n-1})\|^2,$$

$$(3.4.64b) \quad \|z_{n-1} - y_{n+1}\|^2 \leq 2\|\dot{y}_n - \dot{y}_{n+1}\|^2 + 2 \left( \frac{\kappa e + \nu_n}{e + \nu_n} \right)^2 \|\dot{y}_n\|^2,$$

$$(3.4.64c) \quad \|M_\mu(z_n)\|^2 \leq 2\mu^{-2} \|x_{n+1}\|^2 + 2\mu^{-2} \theta_n^2 \|u_n\|^2,$$

$$(3.4.64d) \quad \|M_\mu(y_{n+1})\|^2 \leq 2 \left( \frac{2}{\mu} \right)^2 \|z_n - y_{n+1}\|^2 + 2\|M_\mu(z_n)\|^2,$$

où  $M_\mu$  est défini par (3.2.1).

**Démonstration.** Prouvons séparément chacune des inégalités de (3.4.64) :

En utilisant la relation (3.2.3), on obtient :

$$(3.4.65) \quad \dot{y}_{n+1} - \dot{y}_n + (1 - \kappa\alpha_{n-1})\dot{y}_n + (\kappa\mu)M_\mu(z_{n-1}) = 0.$$

Par ailleurs, un rapide calcul nous donne

$$1 - \kappa\alpha_{n-1} = \frac{\kappa e + \dot{\nu}_n}{e + \nu_n}.$$

On obtient donc classiquement (3.4.64a). De plus, (d'après (3.1.3)), nous avons  $z_{n-1} = y_n + \alpha_{n-1}\dot{y}_n$ , on obtient alors :

$$z_{n-1} - y_{n+1} = \dot{y}_n - \dot{y}_{n+1} - (1 - \alpha_{n-1})\dot{y}_n,$$

ce qui entraîne immédiatement (3.4.64b). L'estimation (3.4.64c) est classiquement obtenue à partir de la relation  $M_\mu(z_n) = \mu^{-1}(\dot{x}_{n+1} + \theta_n u_n)$  donnée par (3.3.16). Enfin, il est bien connu que  $\nabla f$  est  $L^{-1}$ -cocoercif, puisque  $\nabla f$  est  $L$ -lipschitzien. On peut donc vérifier que  $M_\mu$  est  $\mu(1 - \frac{\mu L}{4})$ -cocoercif (voir, par exemple, [8, Lemma B1]), sous la condition  $\mu \in ]0, 2L^{-1}[$ . Il en résulte que l'opérateur  $M_\mu$  est  $(\frac{2}{\mu})$ -lipschitzien. On obtient alors (3.4.64d). ■

**Proposition 3.4.3.** *Supposons que les conditions (3.1.2) soient vérifiées et soient  $\{z_n, x_n, y_n\} \subset \mathcal{H}$  des suites générées par (3.1.3)-(3.1.4). Soit  $e > \sup_{n \geq 0} \beta_n$  et supposons que  $(\nu_n)$  satisfait (3.4.11) avec  $\nu_0 > \frac{(\kappa-1) + \bar{\nu}_M^*}{2 - \kappa - \mu L}$ . Alors, sous chacune des conditions (3.4.12a), (3.4.12b) et (3.4.12c), nous avons*

$$(3.4.66) \quad \sum_n n \|M_\mu(y_n)\|^2 < +\infty, \quad \|M_\mu(y_n)\| = o(n^{-1});$$

et pour  $\kappa \in ]0, 2[$  et  $\mu \in ]0, \frac{2-\kappa}{L}[$ , nous avons aussi

$$(3.4.67) \quad \sum_n n^2 \|\dot{y}_{n+1} - \dot{y}_n\|^2 < +\infty, \quad \sum_n n^2 \|M_\mu(x_n)\|^2 < +\infty.$$

**Démonstration.** Prouvons (3.4.66). D'après l'estimation (3.4.64d) et en remarquant que  $z_n = y_{n+1} + \alpha_n \dot{y}_{n+1}$ , on obtient

$$(3.4.68) \quad \|M_\mu(y_{n+1})\|^2 \leq 2 \left( \frac{2\alpha_n}{\mu} \right)^2 \|\dot{y}_{n+1}\|^2 + 2 \|M_\mu(z_n)\|^2.$$

Ainsi, comme  $\alpha_n$  est bornée et que  $\sum_n n \|\dot{y}_n\|^2 < +\infty$  (d'après (3.4.50a)) ainsi que

$\sum_n n \|M_\mu(z_n)\|^2 < +\infty$ , on obtient  $\sum_n n \|M_\mu(y_n)\|^2 < +\infty$ . Par ailleurs, la condition  $e > \sup_{n \geq 0} \beta_n$  assure que  $n \|\dot{x}_{n+1}\| \rightarrow 0$  et  $n \|u_n\| \rightarrow 0$  (lorsque  $n \rightarrow +\infty$ ), d'après (3.4.50c)-(3.4.50d). Ainsi sachant que  $\theta_n$  est borné (car  $\theta_n = \kappa + \kappa\alpha_n - 1$ ), on déduit de (3.4.64c) que l'on a  $n \|M_\mu(z_n)\| \rightarrow 0$  lorsque  $n \rightarrow +\infty$ . On obtient alors d'après l'estimation (3.4.68), que  $n \|M_\mu(y_n)\| \rightarrow 0$ , car  $\alpha_n$  est borné et  $n \|\dot{y}_n\| \rightarrow 0$  (d'après (3.4.50c)).

Prouvons (3.4.67). En utilisant la relation (3.2.3), on obtient

$$\dot{y}_{n+1} - \dot{y}_n + (1 - \kappa\alpha_{n-1})\dot{y}_n + (\kappa\mu)M_\mu(z_{n-1}) = 0,$$

où  $1 - \kappa\alpha_{n-1} = \frac{\kappa e + \dot{\nu}_n}{e + \nu_n}$ . Donc, comme  $\dot{\nu}_n \leq \beta$ , on obtient

$$\|\dot{y}_{n+1} - \dot{y}_n\|^2 \leq 2 \left( \frac{\kappa e + \beta}{e + \nu_n} \right)^2 \|\dot{y}_n\|^2 + 2(\kappa\mu)^2 \|M_\mu(z_{n-1})\|^2,$$

Par conséquent, la première estimation de (3.4.67) est déduite des deux estimations  $\sum_n n \|\dot{y}_n\|^2 < +\infty$  et  $\sum_n n^2 \|M_\mu(z_n)\|^2 < +\infty$  (d'après (3.4.50a) et (3.4.6)).

Démontrons la deuxième estimation de (3.4.67). L'opérateur  $M_\mu$  étant  $(\frac{2}{\mu})$ -lipschitzien, on a classiquement

$$(3.4.69) \quad \|M_\mu(x_{n+1})\|^2 \leq 2 \|M_\mu(z_n)\|^2 + 2 \left( \frac{2}{\mu} \right)^2 \|z_n - x_{n+1}\|^2.$$

Par ailleurs, la relation (3.2.6a), nous donne  $\|z_n - x_{n+1}\| = \|\dot{x}_{n+1} + \theta_n u_n\|$ , ce qui d'après (3.4.4) entraîne que  $\sum_n n^2 \|z_n - x_{n+1}\|^2 < +\infty$ . De plus, on a  $\sum_n n^2 \|M_\mu(z_n)\|^2 < +\infty$  (d'après (3.4.6)), il résulte alors de (3.4.69) que  $\sum_n n^2 \|M_\mu(x_{n+1})\|^2 < +\infty$ , ce qui entraîne  $\sum_n n^2 \|M_\mu(x_n)\|^2 < +\infty$ . ■

### 3.4.2 Quelques estimations avec des paramètres particuliers.

En tant que conséquence directe de la Proposition (3.4.2), nous présentons deux théorèmes associés à (3.1.3)-(3.1.4) avec des paramètres particuliers  $(\nu_n)$  vérifiant  $\nu_{n+1} - \nu_n \sim 1$  lorsque  $n \rightarrow +\infty$ .

**Remarque 3.4.6.** Notons que, pour toute suite  $(\nu_n)$  vérifiant  $\nu_{n+1} - \nu_n \sim \beta$  lorsque  $n \rightarrow +\infty$  (pour un certain  $\beta > 0$ ), la méthode (3.1.3)-(3.1.4) ne dépend que de  $\beta$  via le terme inertiel  $\alpha_n = \frac{(1-\kappa)e + \nu_n}{\kappa(e + \nu_{n+1})}$  (donné par (3.1.4)). Ce dernier peut, en effet, se réécrire  $\alpha_n = \frac{(1-\kappa)e_* + \bar{\nu}_n}{\kappa(e_* + \bar{\nu}_{n+1})}$  avec  $e_* = (1/\beta)e$  et où  $\bar{\nu}_n = (\nu_n/\beta)$  vérifie  $\bar{\nu}_{n+1} - \bar{\nu}_n \sim 1$ . Ainsi, choisir  $\beta \neq 1$  ne présente aucun intérêt dans notre étude.

Le premier théorème concerne le cas particulier où  $\nu_n = n + \nu_0$  (avec  $\nu_0 > 0$ ).

**Théorème 3.4.1.** *Supposons que la condition (3.1.2) soit vérifiée. Soit  $\nu_n = n + \nu_0$  (pour une certaine valeur positive  $\nu_0$ ) et soient  $\{x_n, y_n\} \subset \mathcal{H}$  des suites générées par (3.1.3)-(3.1.4). Soient  $\kappa \in ]0, 2[$  et  $\{e, \mu\}$  des paramètres strictement positifs vérifiant*

$$(3.4.70) \quad e > 2\kappa^{-1}.$$

*Supposons en outre que l'une des conditions (3.4.71a) et (3.4.71c) soit satisfaite*

$$(3.4.71a) \quad \kappa = 1 \quad \text{et} \quad \mu \in ]0, \frac{1}{L}],$$

$$(3.4.71b) \quad \kappa \in ]0, 1[ \quad \text{et} \quad \mu \in ]0, \frac{2-\kappa}{L}[,$$

$$(3.4.71c) \quad \kappa \in (1, 2) \quad \text{et} \quad \mu \in ]0, \frac{2-\kappa}{L}].$$

*Alors les suites  $(x_n)$  et  $(y_n)$  convergent faiblement vers un élément  $\bar{z} \in S$  et nous avons les estimations suivantes :*

$$(3.4.72a) \quad \Theta(x_n) - \min \Theta = o(n^{-2}), \quad \sum_n n (\Theta(x_n) - \min \Theta) < +\infty,$$

$$(3.4.72b) \quad \sum_n n \|x_{n+1} - x_n\|^2 < +\infty, \quad \|x_{n+1} - x_n\| = o(n^{-1}),$$

$$(3.4.72c) \quad \sum_n n \|y_{n+1} - y_n\|^2 < +\infty, \quad \|y_{n+1} - y_n\| = o(n^{-1}).$$

*En particulier, nous avons les résultats suivants selon les valeurs de  $\kappa$  et  $\mu$  :*

- Si  $\kappa \in ]0, 1]$  et si  $\mu \in ]0, \frac{2-\kappa}{L}]$ , alors

$$(3.4.73) \quad \Theta(y_n) - \min \Theta = o(n^{-2}), \quad \sum_n n (\Theta(y_n) - \min \Theta) < +\infty.$$

- Si  $\kappa \in ]0, 2[$  et si  $\mu \in ]0, \frac{2-\kappa}{L}[,$  alors

$$(3.4.74a) \quad \exists \xi_n^* \in \partial \Theta(x_n) \quad \text{t.q.} \quad \sum_n n^2 \|\xi_n^*\|^2 < +\infty.$$

**Démonstration.** Considérons, pour  $n \geq 1$ , la suite  $(\nu_n)_n$  de la forme  $\nu_n = n + \nu_0$  avec  $\nu_0 > 0$ . Alors en posant  $\rho_n = 1 - (1 - \kappa)\nu_{n+1}\nu^{-1}$ , on obtient

$$\dot{\nu}_n = 1, \quad \nu_{n+1}\nu_n^{-1} = 1 + \nu_n^{-1}, \quad \text{et} \quad \rho_n \sim \kappa \quad \text{lorsque} \quad n \rightarrow +\infty.$$

On déduit de ces estimations que  $\left(\frac{\nu_n}{\nu_{n-1}} - \frac{\nu_{n+1}}{\nu_n}\right)(e + \nu_n) = \frac{(e+\nu_n)}{\nu_{n-1}\nu_n} \sim n^{-1}$  lorsque  $n \rightarrow +\infty$ , et que la suite  $(\beta_n)_n$  introduite dans (3.4.1a) vérifie  $\beta_n \sim 2\kappa^{-1}$  lorsque  $n \rightarrow +\infty$ . Ainsi, pour  $\epsilon \in ]0, 2\kappa^{-1}[$  donné, on sait qu'il va exister un entier  $p$  suffisamment grand

tel que  $2\kappa^{-1} - \epsilon > \beta_n$  (pour  $n \geq p$ ). Par conséquent, la condition  $e > 2\kappa^{-1}$  entraîne que  $e > \epsilon + \beta_n$  (pour  $n \geq p$ ), ce qui implique que  $e \geq \epsilon + \sup_{n \geq p} \beta_n > \sup_{n \geq p} \beta_n$ .

Il en résulte que l'on a, pour  $p$  suffisamment grand

$$\begin{aligned}
 (3.4.75a) \quad & \dot{\nu}_n = 1 \text{ (pour } n \geq 1), \\
 (3.4.75a) \quad & (\nu_{n+1}\nu_n^{-1})_{n \geq p} \text{ est croissante (pour } \kappa > 0), \\
 (3.4.75a) \quad & (\rho_n)_{n \geq p} \subset (0, \infty), \\
 (3.4.75a) \quad & (\nu_n)_{n \geq p} \subset (e(\kappa - 1)_+, \infty), \\
 (3.4.75a) \quad & e > \sup_{n \geq p} \beta_n, \\
 (3.4.75a) \quad & \nu_p > \frac{(\kappa-1)_+}{2-\kappa-\mu L} \text{ si } 2 - \kappa - \mu L > 0.
 \end{aligned}$$

Les conclusions de ce théorème découlent alors de la Proposition 3.4.2. ■

**Remarque 3.4.7.** *Il est clair que le Théorème 3.4.1 couvre les résultats proposés par Attouch-Peypouquet [14] concernant la méthode AFB (3.1.7), à savoir le cas particulier de (3.1.3)-(3.1.4) lorsque  $\kappa = 1$  et  $e = \alpha - 1$  (pour un certain  $\alpha$  strictement positif). Par conséquent, l'application du Théorème 3.4.1 pour  $e > 2$  nous donne les résultats de convergence de [14] qui ont été obtenus pour  $\alpha > 3$ .*

Le second résultat concerne le cas  $\kappa > 1$  et où  $(\nu_n)$  est donné pour  $n \geq 1$  par la formule récursive

$$(3.4.76) \quad \nu_{n+1} = \nu_n \left( 1 + \frac{\kappa + (1 - 2\epsilon_n)\rho_{n-1}}{2\nu_n\rho_{n-1} - (\kappa - 1)} \right),$$

pour une certaine suite  $(\epsilon_n)$  positive décroissant lentement vers 0, et où  $\nu_0$  et  $\nu_1$  sont des paramètres strictement positifs vérifiant

$$(3.4.77a) \quad \nu_1 \geq \nu_0 > \kappa^{-1}(\kappa - 1),$$

$$(3.4.77b) \quad \exists \bar{\epsilon} \in [0, \frac{1}{2}[ \text{ tel que } (\epsilon_n) \subset [0, \bar{\epsilon}] \text{ et } \lim_{n \rightarrow +\infty} \epsilon_n = 0.$$

**Remarque 3.4.8.** *Une accélération de la méthode semble avoir été obtenue concernant la suite récursive  $(\nu_n)$  donnée par (3.4.76) au lieu du choix classique  $\nu_n = n + \nu_0$ . Cependant, nous n'étudions pas cette question d'un point de vue théorique.*

La remarque suivante montre que la formule récursive (3.4.76) est bien définie.

**Remarque 3.4.9.** Rappelons que pour  $\kappa > 1$ , la propriété  $\nu_{n+1} \geq \nu_n > 0$  (pour un certain  $n \geq 0$ ) entraîne  $\rho_n \geq \kappa$ . On peut alors vérifier par récurrence, qu'en prenant  $\nu_0$  et  $\nu_1$  vérifiant (3.4.77), la suite  $(\nu_n)$  donnée par (3.4.76) est bien définie et satisfait, pour  $n \geq 1$ ,

$$(3.4.78) \quad \rho_{n-1} \geq \kappa,$$

$$(3.4.79) \quad 2\nu_n \rho_{n-1} \geq 2\kappa\nu_0 > 2(\kappa - 1),$$

$$(3.4.80) \quad \nu_{n+1} \geq \nu_n.$$

Il en résulte que (3.4.76) génère une suite  $(\nu_n)$  croissante et strictement positive.

Mettons en évidence quelques propriétés concernant la suite  $(\nu_n)$  définie par (3.4.76).

**Proposition 3.4.4.** Pour toute suite  $(\nu_n)$  générée par (3.4.76), ainsi que pour  $(\epsilon_n)$ ,  $\nu_0$  et  $\nu_1$  vérifiant (3.4.77), nous avons les propriétés suivantes

$$(3.4.81) \quad \frac{1-\bar{\epsilon}}{2} \leq \dot{\nu}_n \leq 2 \quad \text{pour } n \geq 2, \quad \text{donc } \nu_n^{-1} = \mathcal{O}(n^{-1}) \text{ et } \nu_n = \mathcal{O}(n),$$

$$(3.4.82) \quad \nu_{n+1} - \nu_n \sim 1 \quad (\text{lorsque } n \rightarrow +\infty).$$

De plus, la suite  $(\nu_n)$  vérifie

$$(3.4.83) \quad 2\dot{\nu}_{n+1} + \frac{\kappa - 1}{\rho_{n-1}} \left( \frac{\nu_n}{\nu_{n-1}} - \frac{\nu_{n+1}}{\nu_n} \right) = 2(1 - \epsilon_n)$$

**Démonstration.** Introduisons l'application  $\varphi_1 : x \mapsto \frac{x}{2x - (\kappa - 1)}$ . On vérifie facilement que  $\varphi_1$  est décroissante sur  $](1/2)(\kappa - 1), +\infty[$  et satisfait

$$(3.4.84) \quad \forall x > \frac{1}{2}(\kappa - 1), \quad \varphi_1(x) \geq \frac{1}{2}.$$

Par ailleurs, d'après (3.4.76) on a

$$(3.4.85) \quad \dot{\nu}_{n+1} = \frac{(\kappa + (1 - 2\epsilon_n)\rho_{n-1})\nu_n}{2\rho_{n-1}\nu_n - (\kappa - 1)}.$$

Ainsi, comme  $\nu_n \rho_{n-1} > \kappa - 1$  (d'après (3.4.79)), on en déduit que

$$(3.4.86) \quad \dot{\nu}_{n+1} \geq \frac{(1 - 2\epsilon_n)\rho_{n-1}\nu_n}{2\rho_{n-1}\nu_n - (\kappa - 1)} = (1 - 2\epsilon_n)\varphi_1(\nu_n \rho_{n-1}) \geq \frac{1 - 2\bar{\epsilon}}{2}.$$

De plus, comme  $\kappa \leq \rho_{n-1}$  (d'après (3.4.78)) et  $\nu_n \rho_{n-1} > \kappa - 1$  (d'après (3.4.79)), on obtient en utilisant (3.4.85)

$$(3.4.87) \quad \begin{aligned} \dot{\nu}_{n+1} &\leq \frac{2(1-\epsilon_n)\nu_n\rho_{n-1}}{2\rho_{n-1}\nu_n - (\kappa-1)} \\ &= 2(1-\epsilon_n)\varphi_1(\nu_n\rho_{n-1}) \leq 2(1-\epsilon_n)\varphi_1(\kappa-1) \leq 2(1-\epsilon_n) \leq 2. \end{aligned}$$

Les inégalités (3.4.86) et (3.4.87) nous donnent alors  $(1/2)(1-2\bar{\epsilon}) \leq \dot{\nu}_{n+1} \leq 2$  et  $\nu_n \rightarrow +\infty$  (lorsque  $n \rightarrow +\infty$ ). On en déduit donc que  $\nu_{n+1}\nu_n^{-1} \rightarrow 1$  (lorsque  $n \rightarrow +\infty$ ). Ceci implique que  $\rho_n \sim \kappa$  (lorsque  $n \rightarrow +\infty$ ), ce qui entraîne, d'après (3.4.85) et  $\epsilon_n \rightarrow 0$ , que

$$(3.4.88) \quad \dot{\nu}_{n+1} \sim 1 \quad (\text{lorsque } n \rightarrow +\infty).$$

Démontrons la relation (3.4.83). L'égalité (3.4.83) est équivalente à

$$\nu_{n+1} \left( 2 - \frac{\kappa-1}{\nu_n\rho_{n-1}} \right) = 2(1-\epsilon_n) + 2\nu_n - \frac{(\kappa-1)}{\rho_{n-1}} \frac{\nu_n}{\nu_{n-1}}.$$

Cette dernière peut se réécrire sous la forme

$$(3.4.89) \quad \nu_{n+1} = \nu_n \left( \frac{2\nu_n\rho_{n-1} + 2(1-\epsilon_n)\rho_{n-1} - (\kappa-1)\nu_n\nu_{n-1}^{-1}}{2\nu_n\rho_{n-1} - (\kappa-1)} \right)$$

$$(3.4.90) \quad = \nu_n \left( 1 + \frac{\kappa-1 + 2(1-\epsilon_n)\rho_{n-1} - (\kappa-1)\nu_n\nu_{n-1}^{-1}}{2\nu_n\rho_{n-1} - (\kappa-1)} \right)$$

Ainsi, en remarquant que  $(\kappa-1)\frac{\nu_n}{\nu_{n-1}} = \rho_{n-1} - 1$  (d'après la définition de  $\rho_n$ ), on obtient

$$(3.4.91) \quad \frac{\nu_{n+1}}{\nu_n} = 1 + \frac{\kappa-1 + 2(1-\epsilon_n)\rho_{n-1} - \rho_{n-1} + 1}{2\rho_{n-1}\nu_n - (\kappa-1)} = 1 + \frac{\kappa + (1-2\epsilon_n)\rho_{n-1}}{2\rho_{n-1}\nu_n - (\kappa-1)}.$$

Ce qui nous donne le résultat souhaité (3.4.76). ■

Exposons à présent le second résultat de cette section.

**Théorème 3.4.2.** *Supposons que les conditions (3.1.2) soient vérifiées. Soient  $\{x_n, y_n\} \subset \mathcal{H}$  des suites générées par (3.1.3)-(3.1.4) avec  $\kappa \in ]1, 2[$ . Soit  $(\nu_n)_{n \geq 0}$  définie de façon récursive par (3.4.76)-(3.4.77b) et soient des paramètres  $\{\nu_0, \nu_1, e, \mu\} \subset$*

$]0, +\infty[$  vérifiant,

$$(3.4.92) \quad e > 2\kappa^{-1}, \quad \mu \in ]0, \frac{2-\kappa}{L}[ \quad \text{et} \quad \nu_1 \geq \nu_0 > e(\kappa - 1).$$

Alors les suites  $(x_n)$  et  $(y_n)$  convergent faiblement vers un élément  $\bar{z} \in S$  et nous obtenons

$$(3.4.93a) \quad \Theta(x_n) - \min \Theta = o(n^{-2}), \quad \sum_n n (\Theta(x_n) - \min \Theta) < +\infty,$$

$$(3.4.93b) \quad \sum_n n \|x_{n+1} - x_n\|^2 < +\infty, \quad \|x_{n+1} - x_n\| = o(n^{-1}),$$

$$(3.4.93c) \quad \sum_n n \|y_{n+1} - y_n\|^2 < +\infty, \quad \|y_{n+1} - y_n\| = o(n^{-1}),$$

$$(3.4.93d) \quad \exists \xi_n^* \in \partial \Theta(x_n) \quad t.q. \quad \sum_n n^2 \|\xi_n^*\|^2 < +\infty.$$

Si, de plus,  $\nu_0$  et  $\nu_1$  vérifient

$$(3.4.94) \quad \nu_1 > \nu_0 > \max \left\{ e(\kappa - 1), \frac{2(\kappa-1)}{2-\kappa-\mu L} \right\},$$

alors la suite de type énergie  $(\mathcal{E}_n(e, q))_{n \geq 1}$  est décroissante.

**Démonstration.** Soit  $(\nu_n)_n$  la suite donnée par (3.4.76) et soient  $\{e, \kappa, \nu_0, \nu_1\} \subset ]0, +\infty[$  vérifiant  $e > 2\kappa^{-1}$  et  $\nu_1 \geq \nu_0 > e(\kappa - 1)$ . Ces deux dernières conditions entraînent  $\nu_1 \geq \nu_0 > 2\kappa^{-1}(\kappa - 1)$ , et donc (3.4.77a) est satisfait. De plus, d'après la définition de  $\beta_n$ , on a pour  $n \geq 1$ ,

$$(3.4.95) \quad \beta_n = \rho_n^{-1} \left( 2\dot{\nu}_{n+1} + \frac{(\kappa-1)}{\rho_{n-1}} \left( \frac{\nu_n}{\nu_{n-1}} - \frac{\nu_{n+1}}{\nu_n} \right) (e + \nu_n) \right).$$

Par conséquent, comme la suite  $(\nu_n)$  est croissante (et donc,  $\rho_n \geq \kappa$ , car  $\kappa > 1$ ), on obtient d'après (3.4.83) que  $\beta_n = 2\rho_n^{-1} \leq 2\kappa^{-1}$ . Ainsi, à partir de la condition  $e > 2\kappa^{-1}$  (d'après (3.4.92)), on en déduit que  $e > 2\kappa^{-1} \geq \sup_{n \geq 1} \beta_n$ . Il en résulte alors que

$$(3.4.96a) \quad (\nu_n)_{n \geq 0} \subset ]e(\kappa - 1), +\infty[,$$

$$(3.4.96b) \quad (\dot{\nu}_n)_{n \geq 2} \subset \left[ \frac{1-2\bar{\epsilon}}{2}, 2 \right],$$

$$(3.4.96c) \quad (\rho_n)_{n \geq 0} \subset ]0, +\infty[,$$

$$(3.4.96d) \quad e > \sup_{n \geq 1} \beta_n.$$

Remarquons que la condition  $\nu_1 > \nu_0$  intervenant dans (3.4.94) assure que

$$(3.4.97) \quad 0 < \min\{\dot{\nu}_1, \frac{1-2\bar{\epsilon}}{2}\} \leq \dot{\nu}_n \leq \max\{\dot{\nu}_1, 2\} \quad (\text{pour } n \geq 1).$$

Les conclusions de ce théorème découlent alors de la Proposition 3.4.2. ■

## 3.5 Simulations numériques

Quelques simulations numériques sont réalisées dans ce chapitre afin d’illustrer le comportement de l’algorithme **RIGPA**, donné par (3.1.3)-(3.1.4), pour trois exemples. Le premier concerne un problème de régression logistique et sera utilisé afin d’illustrer l’influence de certains paramètres sur les performances de **RIGPA**. Le second est consacré à la comparaison des performances de **RIGPA** avec celles de l’algorithme **IGAHD** récemment proposé (voir [18]), qui est également issu d’une discrétisation implicite en temps d’un modèle dynamique du second ordre faisant intervenir la Hessienne.

Dans la suite, nous désignons par  $\|\cdot\|_1$  et  $\|\cdot\|$  respectivement la norme  $L_1$  et la norme  $L_2$  de  $\mathbb{R}^d$ .

### 3.5.1 Exemple 1 : régression logistique

Cet exemple intervient en apprentissage automatique pour la classification binaire. Il s’agit de résoudre le problème de minimisation :

$$(3.5.1) \quad \min_{\mathbb{R}^d} \{\Theta(x) := f(x) + g(x)\},$$

où  $f$  et  $g$  sont deux fonctions définies par

- $g(x) = \lambda \|x\|_1$  (avec  $\lambda > 0$ ),
- $f(x) = \frac{1}{N} \sum_{i=1}^N f_i(x)$ , où  $N$  est un entier donné et où  $f_i(x) = \log(1 + e^{-b_i \langle x, a_i \rangle})$  pour des données d’entraînement  $(a_i, b_i) \in \mathbb{R}^d \times \{+1, -1\}$  ( $i = 1, \dots, N$ ).

On vérifie facilement qu’une constante de lipschitz de  $f_i$  est donnée par  $L_i = \|a_i\|^2/4$ . Ainsi, une constante de lipschitz de  $f$  est  $\bar{L} = \frac{1}{N} \sum_{i=1}^N L_i$ . Dans les simulations, chaque  $a_i$  est divisé par leur norme, afin d’obtenir  $\bar{L} = 1/4$ . Les autres paramètres sont  $d = 8$ ,  $N = 768$  et  $\lambda = 1/768$ . Rappelons que  $e$  doit être tel que  $e > 2\kappa^{-1}$  pour la convergence de **RIGPA**.

Les Figures de 3.5.1 à 3.5.4 montrent la performance des trajectoires générées par RIPGA via l'évolution de  $\Theta(x_n) - \Theta(x^*)$  en fonction du nombre d'itérations, où  $x^* := \operatorname{argmin} \Theta$ . Les conditions initiales sont  $y_0 = y_1 = (0, 0, 0, 0, 0, 0, 0, 0)$ . Sur toutes les figures, les valeurs de  $\Theta(x_n) - \min \Theta$  ont été fixées à  $10^{-13}$  lorsqu'elles sont plus petites que  $10^{-13}$ .

Les Figures 3.5.1, 3.5.2 et 3.5.3 décrivent l'influence de  $\kappa$  sur les trajectoires pour des valeurs fixes de  $e$ . Notons que les oscillations données par RIGPA sont considérablement réduites pour des valeurs croissantes de  $e$ . De plus, la convergence semble être plus rapide pour des valeurs croissantes de  $\kappa$  dans  $]0, 2[$ .

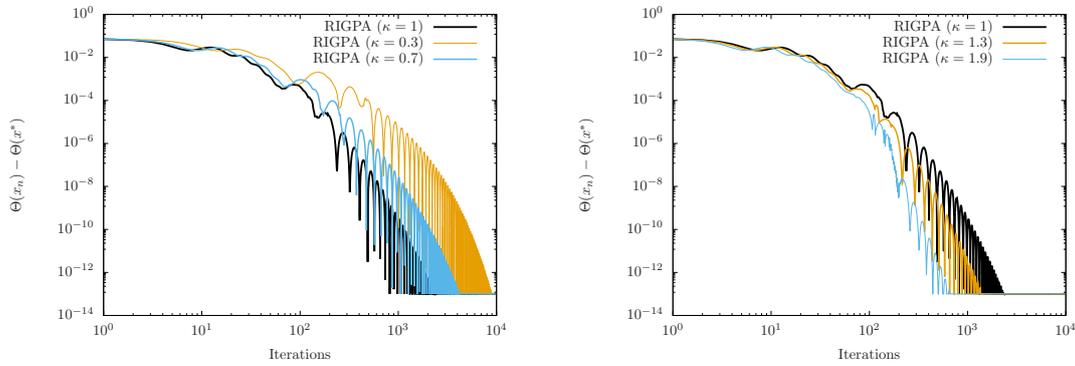


FIGURE 3.5.1 – Évolution de  $\Theta(x_n) - \min \Theta$  pour RIPGA avec  $e = 7$  et diverses valeurs de  $\kappa$ . Les autres paramètres sont  $\mu = 0.35$ ,  $\nu_0 = 144$  et  $\nu_n = n + \nu_0$ .

Sur la Figure 3.5.4, nous comparons RIGPA muni de  $\nu_n = n + \nu_0$  avec RIGPA muni de la suite  $(\nu_n)$  générée par la formule récursive (3.4.76)-(3.4.77). L'accélération de la méthode semble avoir été améliorée pour ce dernier choix de  $(\nu_n)$ .

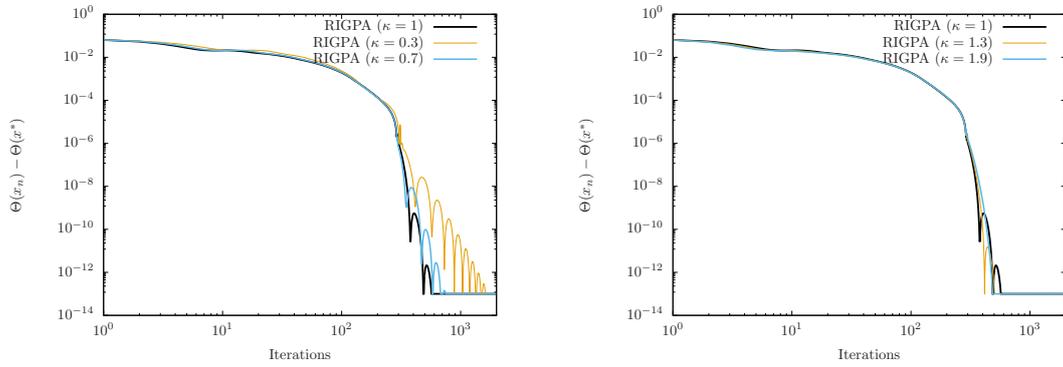


FIGURE 3.5.2 – Évolution de  $\Theta(x_n) - \min \Theta$  pour RIPGA avec  $e = 15$  et diverses valeurs de  $\kappa$ . Les autres paramètres sont les mêmes que pour la Figure 3.5.1.

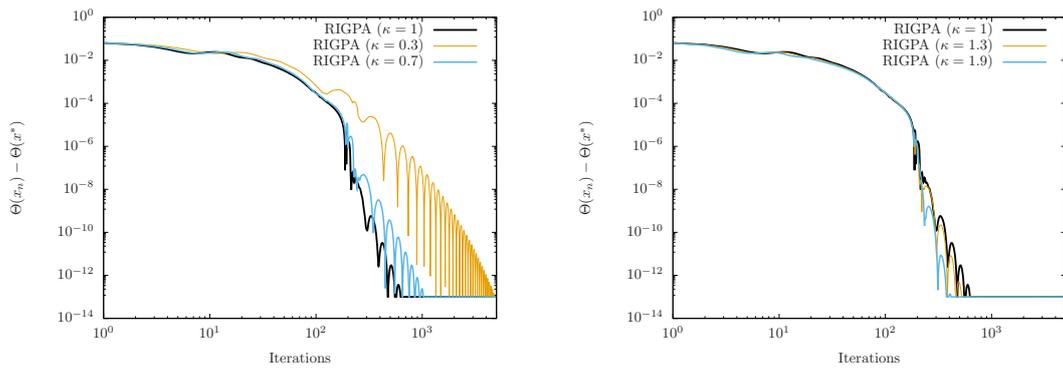


FIGURE 3.5.3 – Évolution de  $\Theta(x_n) - \min_{\mathcal{H}} \Theta$  pour RIPGA avec  $e = 30$  et diverses valeurs de  $\kappa$ . Les autres paramètres sont les mêmes que pour la Figure 3.5.1.

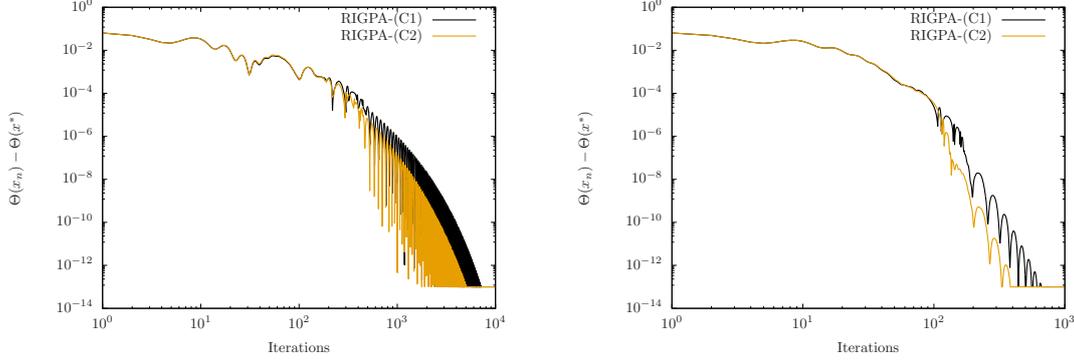


FIGURE 3.5.4 – Évolution de  $\Theta(x_n) - \min_{\mathcal{H}} \Theta$  pour RIGPA avec  $e = 1.5$  (gauche),  $e = 7$  (droite),  $\kappa = 1.9$  et  $\nu_n = n + \nu_0$  pour (C1), la formule récursive de  $(\nu_n)$  pour (C2), avec  $\nu_1 = \nu_0 + 1$  et  $\epsilon_n = (\ln(n + 55))^{-0.5}$ . Les autres paramètres sont les mêmes que pour la Figure 3.5.1.

### 3.5.2 Exemple 2 : comparaison de RIGPA avec IGAHD

Le deuxième exemple est lié à la résolution d'un problème LASSO, intervenant dans de nombreux domaines tels que les statistiques, l'apprentissage automatique et le traitement d'images, pour lequel sont comparés les performances de RIGPA avec celles de IGAHD. Il s'agit de résoudre dans  $\mathcal{H} = \mathbb{R}^d$  :

$$(3.5.2) \quad \min_{x \in \mathbb{R}^d} \{\Theta(x) := f(x) + g(x)\},$$

avec  $f(x) := (1/2)\|y - Ax\|^2$ ,  $g(x) := \|x\|_1$ ,  $A : \mathbb{R}^d \rightarrow \mathbb{R}^d$  un opérateur linéaire, et  $y \in \mathbb{R}^d$  est un vecteur aléatoire.

On vérifie qu'une constante de lipschitz de  $\nabla f$  est donnée par  $L = \|A^*A\|_2$ .

**Remarque 3.5.10.** Rappelons que IGAHD (Inertial Gradient Algorithm with Hessian Driven Damping) a été introduit par Attouch-Chbani-Fadili-Riahi dans [18] pour minimiser une fonction convexe lisse  $\Theta : \mathcal{H} \rightarrow \mathbb{R}$  à gradient  $L_\Theta$ -lipschitzien. Ce schéma est donné pour des valeurs strictement positives  $\{\mu, \alpha, \beta\}$ , par

$$\begin{cases} y_n &= x_n + \left(1 - \frac{\alpha}{n}\right) (x_n - x_{n-1}) - \beta\sqrt{\mu} (\nabla\Theta(x_n) - \nabla\Theta(x_{n-1})) - \frac{\beta\sqrt{\mu}}{n} \nabla\Theta(x_{n-1}), \\ x_{n+1} &= y_n - \mu \nabla\Theta(y_n). \end{cases}$$

La convergence des valeurs de la fonction objectif avec un taux de  $o(n^{-2})$  ainsi que la propriété de convergence rapide vers 0 des gradients (à savoir,  $\sum_n n^2 \|\nabla\Theta(x_n)\|^2 <$

$+\infty$ ) ont été établies sous les conditions

$$(3.5.3) \quad \alpha \geq 3, \quad 0 \leq \beta < 2\sqrt{\mu} \quad \text{et} \quad \mu \leq L_{\Theta}^{-1}.$$

Bien que IGAHD n'a pas encore été étendu au cas de la minimisation structurée, nous l'utilisons pour résoudre le problème (3.5.2) en suivant la même méthodologie que dans [18] pour laquelle nous rappelons les grandes lignes via la Remarque suivante.

**Remarque 3.5.11.** *Pour une matrice symétrique définie positive  $M \in \mathbb{R}^{d \times d}$ , les auteurs de [18] ont montré que  $\operatorname{argmin}_{\mathbb{R}^d} \Theta = \operatorname{Fix}(\operatorname{prox}_{\Theta}^M) = \operatorname{argmin}_{\mathbb{R}^d} \Theta_M$ , où  $\Theta_M$  et  $\operatorname{prox}_{\Theta}^M(x)$  sont respectivement l'enveloppe de Moreau et l'application proximale de  $\Theta$  dans la métrique  $\langle M \cdot, \cdot \rangle$ , définis par*

$$(3.5.4) \quad \Theta_M(x) = \min_{z \in \mathbb{R}^d} \left\{ \frac{1}{2} \|z - x\|_M^2 + \Theta(z) \right\},$$

$$(3.5.5) \quad \operatorname{prox}_{\Theta}^M(x) = \operatorname{argmin}_{z \in \mathbb{R}^d} \left\{ \frac{1}{2} \|z - x\|_M^2 + \Theta(z) \right\}.$$

L'approche consiste ici à minimiser  $\Theta_M$  au lieu de  $\Theta$  pour résoudre (3.5.2). En particulier, pour  $0 < r\|A\|^2 < 1$ , la matrice  $M := r^{-1}I - A^*A$  est symétrique et définie positive, et l'application proximale de  $\Theta$  (dans la métrique  $M$ ) est donné par

$$\operatorname{prox}_{\Theta}^M(x) = \operatorname{prox}_{r_g}(x + rA^*(y - Ax)).$$

De plus,  $\Theta_M$  est une fonction convexe continûment différentiable telle que

$$(3.5.6) \quad \nabla \Theta_M(x) = x - \operatorname{prox}_{\Theta}^M(x) = x - \operatorname{prox}_{r_g}(x + rA^*(y - Ax)).$$

On vérifie par ailleurs que  $\nabla \Theta_M$  est 1-lipschitzien dans la métrique  $M$ .

Considérons ici une matrice réelle aléatoire de taille  $100 \times 100$  ainsi qu'un vecteur aléatoire  $y$  tel que chacune des composantes soit dans  $[-1, 1]$ . La Figure 3.5.5 décrit l'évolution de  $\Theta(x_n) - \min \Theta$  obtenus pour RIPGA (pour diverses valeurs de  $\epsilon$ ) et IGAHD (avec diverses valeurs de  $\alpha$ ). Sur cette Figure, les valeurs de  $\Theta(x_n) - \min \Theta$  ont été fixées à  $10^{-10}$  lorsqu'elles sont inférieures à  $10^{-10}$ . Il semble que IGAHD surpasse RIPGA pour des valeurs de  $\kappa$  dans  $]0, 1]$ . Ceci n'est cependant plus le cas pour des valeurs de  $\kappa$  dans  $]1, 2[$  suffisamment grandes.

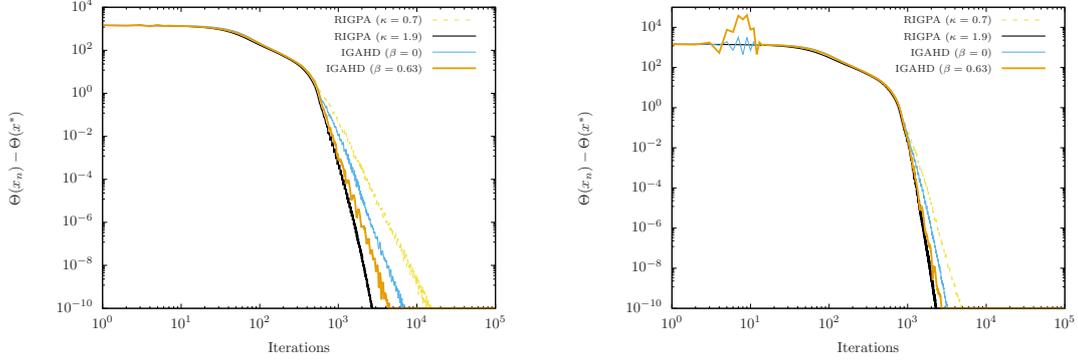


FIGURE 3.5.5 – Profils de  $\Theta(x_n) - \min_{\mathcal{H}} \Theta$  pour RIPGA et IGAHD avec  $(e, \alpha) = (7, 8)$  (gauche) et  $(e, \alpha) = (15, 16)$  (droite). Les conditions initiales sont  $y_0 = y_1 = x_0 = x_1 = (1, 1, \dots, 1)$  avec  $r = (1/2)\|A\|^{-2}$ ,  $\nu_0 = 1$ ,  $\nu_n = n + \nu_0$  et  $\mu = 0.01$ .

### 3.5.3 Application en imagerie : le modèle Rudin-Osher-Fatemi

Dans cette section nous présentons de façon détaillée un exemple d'applications de l'algorithme **RIPGA** dans le cadre d'un problème intervenant dans le traitement de l'image.

#### 3.5.3.1 Préliminaires : Le cas continu

Soit  $\Omega \subset \mathbb{R}^2$  et considérons une image en noir et blanc  $u : \Omega \rightarrow \mathbb{R}$ . En ajoutant un bruit blanc gaussien  $\eta$  à l'image  $u$ , nous obtenons une image bruitée  $u^\diamond$ , cette dernière est donc donnée par la relation  $u^\diamond = u + \eta$ . En 1992, Rudin Osher et Fatemi [99] présentèrent le problème de minimisation suivant afin d'estimer une image débruitée

$$(3.5.7) \quad \min_{u \in \text{BV}(\Omega)} \frac{\lambda}{2} \int_{\Omega} (u(x) - u^\diamond(x))^2 dx + \|u\|_{\text{TV}(\Omega)}.$$

où  $\text{BV}(\Omega)$  désigne l'ensemble des fonctions à variation bornées sur  $\Omega$  et où  $\|u\|_{\text{TV}}$  désigne la semi-norme de la variation totale donnée par (voir [54])

$$(3.5.8) \quad \|u\|_{\text{TV}(\Omega)} := \sup \left\{ \int_{\Omega} u(x) \operatorname{div}(g(x)) dx : g \in C_c^1(\Omega; \mathbb{R}^2), \sqrt{g_1^2 + g_2^2} \leq 1 \right\}$$

Notons que lorsque  $\nabla u \in L^1(\Omega; \mathbb{R}^2)$  existe, nous avons

$$(3.5.9) \quad \|u\|_{\text{TV}} = \int_{\Omega} \|\nabla u(x)\|_2 \, dx$$

### 3.5.3.2 Le cas discret

Dans le cas discret, une image en noir et blanc de taille  $n \times m$  pixels peut être vue sous la forme d’une matrice  $(u_{i,j})$  avec  $1 \leq i \leq n$  et  $1 \leq j \leq m$  ou d’un vecteur  $u \in \mathbb{R}^{n \times m}$  pour lequel les éléments de ce vecteurs sont la concaténation des colonnes de la matrice  $(u_{i,j})$ . Pour  $\lambda > 0$  et en désignant toujours par  $u^\diamond$  l’image issue de  $u$  après ajout d’un bruit blanc, le modèle de ROF discret propose de débruiter l’image  $u^\diamond$  à l’aide du problème de minimisation suivant :

$$(3.5.10) \quad \min_{u \in \mathbb{R}^{n \times m}} \frac{\lambda}{2} \|u - u^\diamond\|_2^2 + \text{TV}(u)$$

où  $\text{TV}(u)$  désigne la variation totale discrète définie par

$$(3.5.11) \quad \text{TV}(u) = \sum_{i=1}^n \sum_{j=1}^m |(Du)_{i,j}| = \sum_{i=1}^n \sum_{j=1}^m \left( (Du)_{i,j,1}^2 + (Du)_{i,j,2}^2 \right)^{\frac{1}{2}}$$

avec  $(Du)_{i,j} = ((Du)_{i,j,1}, (Du)_{i,j,2})$  donnés par l’application  $D : \mathbb{R}^{n \times m} \mapsto \mathbb{R}^{n \times m} \times \mathbb{R}^{n \times m}$  définie par

$$(3.5.12) \quad (Du)_{i,j,1} = \begin{cases} u_{i+1,j} - u_{i,j} & \text{si } 1 \leq i < n \\ 0 & \text{si } i = n \end{cases}$$

$$(3.5.13) \quad (Du)_{i,j,2} = \begin{cases} u_{i,j+1} - u_{i,j} & \text{si } 1 \leq j < m \\ 0 & \text{si } j = m \end{cases}$$

Notons que le problème (3.5.10)-(3.5.11)-(3.5.12)-(3.5.13) peut se mettre sous la forme

$$(3.5.14) \quad \min_{u \in \mathbb{R}^{n \times m}} f(u) + g(u)$$

avec  $f(u) = \frac{\lambda}{2} \|u - u^\diamond\|_2^2$  et  $g(u) = \text{TV}(u)$ . Afin de résoudre le problème (3.5.14), nous allons faire appel à l'algorithme RIGPA. Une difficulté apparaît alors, en effet, bien que les fonctions  $f$  et  $g$  respectent les conditions de régularités nécessaires à la convergence de notre méthode, on observe à l'aide de l'équation (3.5.11) que l'opérateur proximal de la fonction  $g := \text{TV}$  se révèle être difficile à déterminer. Afin de contourner cette difficulté, nous allons nous intéresser au problème dual associé à (3.5.10).

### 3.5.3.3 Une approche duale du problème ROF discret

L'approche duale du problème (3.5.10) proposée dans cette section est classique est en grande partie basée sur celles développées par Chambolle-Pock dans [42] et Li-Han-Wang-Tiande-Guo dans [68] pour laquelle nous détaillons chacune des différentes étapes.

Commençons tout d'abord par présenter le lemme suivant.

**Lemme 3.5.4.** *Soit  $\mathcal{H}$  un espace de Hilbert réel de produit scalaire  $\langle \cdot, \cdot \rangle$  et de norme  $|\cdot| = \sqrt{\langle \cdot, \cdot \rangle}$ . Soient  $(a, b) \in \mathcal{H}^2$  alors on a*

$$|a| = \max_{|b| \leq 1} \langle a, b \rangle$$

**Démonstration.** Soient  $(a, b) \in \mathcal{H}^2$ . Pour  $a = 0$  le résultat est trivialement vérifié. Ensuite, d'une part, en faisant appel à l'inégalité de Cauchy-Schwarz, on obtient  $\max_{|b| \leq 1} \langle a, b \rangle \leq |a|$ , puis en prenant  $b = \frac{a}{|a|}$  pour  $a \neq 0$ , il vient  $\langle a, b \rangle = \frac{|a|^2}{|a|} = |a|$ . ■

À l'aide de la variable duale  $\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2) \in \mathbb{R}^{n \times m} \times \mathbb{R}^{n \times m}$  que l'on peut écrire pixel par pixel sous la forme vectorielle  $\mathbf{p}_{i,j} = (\mathbf{p}_{i,j,1}, \mathbf{p}_{i,j,2})$  pour  $1 \leq i \leq n$  et  $1 \leq j \leq m$ , on peut définir un produit scalaire sur  $\mathbb{R}^{n \times m} \times \mathbb{R}^{n \times m}$  en posant

$$(3.5.15) \quad \langle \mathbf{p}, \mathbf{q} \rangle = \sum_{i=1}^n \sum_{j=1}^m \langle \mathbf{p}_{i,j}, \mathbf{q}_{i,j} \rangle \quad \text{et} \quad \|\mathbf{p}\| = \sqrt{\langle \mathbf{p}, \mathbf{p} \rangle},$$

pour tout  $\mathbf{p}, \mathbf{q} \in \mathbb{R}^{n \times m} \times \mathbb{R}^{n \times m}$ . Ainsi, grâce au Lemme 3.5.4, on peut réécrire l'opérateur TV défini par (3.5.11) sous la forme

$$(3.5.16) \quad \text{TV}(u) = \sum_{i=1}^n \sum_{j=1}^m \max_{|\mathbf{p}_{i,j}| \leq 1} \langle (Du)_{i,j}, \mathbf{p}_{i,j} \rangle = \max_{\mathbf{p} \in \mathcal{P}} \langle Du, \mathbf{p} \rangle,$$

où  $\mathcal{P}$  désigne l'ensemble

$$(3.5.17) \quad \mathcal{P} := \{\mathbf{p} \in \mathbb{R}^{n \times m \times 2}, |\mathbf{p}_{i,j}| \leq 1, \text{ pour tout } 1 \leq i \leq n, 1 \leq j \leq m\},$$

et où  $|\mathbf{p}_{i,j}| = \sqrt{\mathbf{p}_{i,j,1}^2 + \mathbf{p}_{i,j,2}^2}$ . Le problème de minimisation (3.5.10) devient alors

$$(3.5.18) \quad \min_{u \in \mathbb{R}^{n \times m}} \left\{ \frac{\lambda}{2} \|u - u^\diamond\|_2^2 + \max_{\mathbf{p} \in \mathcal{P}} \langle Du, \mathbf{p} \rangle \right\}.$$

On observe que le premier terme de (3.5.18) ne contient pas la variable duale  $\mathbf{p}$ , nous pouvons alors reformuler le problème primal (3.5.10) en un problème min max de la façon suivante

$$(3.5.19) \quad \min_{u \in \mathbb{R}^{n \times m}} \max_{\mathbf{p} \in \mathcal{P}} \left\{ \frac{\lambda}{2} \|u - u^\diamond\|_2^2 + \langle Du, \mathbf{p} \rangle \right\}.$$

De plus, la fonction objectif du problème min max (3.5.19) étant convexe en  $u$  et concave en  $p$ , on déduit alors du théorème du min max de Von Neumann (voir [53]) l'ordre du min et du max peut être échangé afin d'obtenir

$$(3.5.20) \quad \max_{\mathbf{p} \in \mathcal{P}} \min_{u \in \mathbb{R}^{n \times m}} \left\{ \frac{\lambda}{2} \|u - u^\diamond\|_2^2 + \langle Du, \mathbf{p} \rangle \right\}.$$

En outre, en introduisant l'opérateur de divergence discret  $\text{div} : \mathbb{R}^{n \times m \times 2} \rightarrow \mathbb{R}^{n \times m}$  défini comme dans le cas continue par  $\langle Du, \mathbf{p} \rangle = -\langle u, \text{div } \mathbf{p} \rangle$ , nous obtenons alors  $\text{div} = -D^*$  où  $D^*$  désigne l'opérateur adjoint de  $D$ . Par ailleurs, l'opérateur  $\text{div}$  s'écrit explicitement de la façon suivante (voir [39])

$$(\text{div})_{i,j} : \begin{cases} p_{i,j,1} - p_{i-1,j,1} & \text{si } 1 < i < n \\ p_{i,j,1} & \text{si } i = 1 \\ -p_{i,j,1} & \text{si } i = n \end{cases} + \begin{cases} p_{i,j,2} - p_{i,j-1,2} & \text{si } 1 < j < m \\ p_{i,j,2} & \text{si } j = 1 \\ -p_{i,j,2} & \text{si } j = m \end{cases}.$$

Ainsi, l'équation (3.5.20) nous donne

$$(3.5.21) \quad \max_{\mathbf{p} \in \mathcal{P}} \min_{u \in \mathbb{R}^{n \times m}} \left\{ \frac{\lambda}{2} \|u - u^\diamond\|_2^2 - \langle u, \text{div } \mathbf{p} \rangle \right\}.$$

Comme le problème de minimisation  $\min_{u \in \mathbb{R}^{n \times m}} \left\{ \frac{\lambda}{2} \|u - u^\diamond\|_2^2 - \langle u, \text{div } \mathbf{p} \rangle \right\}$  ne concerne que la variable  $u$  et est sans contrainte, la condition d'optimalité du premier ordre

nous donne

$$(3.5.22) \quad u = u^\diamond + \frac{1}{\lambda} \operatorname{div} \mathbf{p}.$$

Ainsi, en remplaçant la valeur de  $u$  obtenue à partir de (3.5.22) dans l'équation (3.5.21), nous obtenons

$$(3.5.23) \quad \max_{\mathbf{p} \in \mathcal{P}} \left\{ -\frac{1}{2\lambda} \|\operatorname{div} \mathbf{p}\|_2^2 - \langle u^\diamond, \operatorname{div} \mathbf{p} \rangle \right\}.$$

En complétant le carré et en supprimant le terme constant en  $u^\diamond$ , les solutions de (3.5.23) sont données par les solutions du problème suivant

$$(3.5.24) \quad \min_{\mathbf{p} \in \mathcal{P}} \left\{ \frac{1}{2} \|\operatorname{div} \mathbf{p} + \lambda u^\diamond\|_2^2 \right\}.$$

Enfin, cette dernière équation peut se réécrire sous la forme d'un problème de minimisation sans contrainte en définissant  $\delta_{\mathcal{P}}$  comme étant la fonction indicatrice de  $\mathcal{P}$  où  $\mathcal{P}$  est défini par (3.5.17) :

$$(3.5.25) \quad \min_{\mathbf{p} \in \mathbb{R}^{n \times m \times 2}} \left\{ \frac{1}{2} \|\operatorname{div} \mathbf{p} + \lambda u^\diamond\|_2^2 + \delta_{\mathcal{P}}(\mathbf{p}) \right\}.$$

avec

$$(3.5.26) \quad \delta_{\mathcal{P}}(p) = \begin{cases} 0 & \text{si } p \in \mathcal{P} \\ +\infty & \text{sinon} \end{cases}.$$

On observe alors que le problème (3.5.25) est bien de la forme (3.5.14) et que cette fois-ci, l'opérateur proximal de l'application  $\delta_{\mathcal{P}}$  de paramètre  $\tau > 0$  se calcule facilement. En effet, ce dernier n'est autre que la projection orthogonal sur l'ensemble  $\mathcal{P}$ , et peut être donné pixel par pixel par la formule suivante (voir [42, 68])

$$(3.5.27) \quad \hat{\mathbf{p}} = \operatorname{prox}_{\tau \delta_{\mathcal{P}}}(\tilde{\mathbf{p}}) \iff \hat{\mathbf{p}}_{i,j} = \frac{\tilde{\mathbf{p}}_{i,j}}{\max\{1, |\tilde{\mathbf{p}}_{i,j}|\}}.$$

### 3.5.3.4 Un exemple numérique

Observons que le problème dual (3.5.25) s'écrit pour  $\mathbf{p}^* \in \mathbb{R}^{n \times m} \times \mathbb{R}^{n \times m}$

$$(3.5.28) \quad \mathbf{p}^* = \underset{\mathbf{p} \in \mathbb{R}^{n \times m \times 2}}{\operatorname{argmin}} \{f(\mathbf{p}) + g(\mathbf{p})\}.$$

avec  $f(\mathbf{p}) = \frac{1}{2} \|\operatorname{div} \mathbf{p} + \lambda u^\diamond\|_2^2$  et  $g(\mathbf{p}) = \delta_{\mathcal{P}}(\mathbf{p})$  où  $\delta_{\mathcal{P}}$  est défini par (3.5.26). La fonction  $f$  est convexe et différentiable et on a  $\nabla f(\mathbf{p}) = -D \circ (\operatorname{div} \mathbf{p} + \lambda u^\diamond)$  où  $D$  est défini par (3.5.12)-(3.5.13). De plus, les opérateurs  $D$  et  $\operatorname{div}$  sont linéaires et vérifient  $\|D\|^2 = \|\operatorname{div}\|^2 \leq 8$  (voir [39]), où  $\|\cdot\|$  désigne la norme d'opérateurs. Ceci nous montre que pour tout  $\mathbf{p}, \mathbf{q} \in \mathbb{R}^{n \times m} \times \mathbb{R}^{n \times m}$ ,

$$(3.5.29) \quad \|\nabla f(\mathbf{p}) - \nabla f(\mathbf{q})\| \leq 8\|\mathbf{p} - \mathbf{q}\|.$$

Ainsi l'application  $\nabla f$  est 8-lipschitzienne et de plus l'application proximale de  $\delta_{\mathcal{P}}$  est simple à calculer et est donnée par (3.5.27), nous pouvons donc appliquer notre algorithme RIGPA pour résoudre le problème (3.5.25)-(3.5.26).

Considérons pour cela une image en noir et blanc  $u \in \mathbb{R}^{720 \times 540}$  de dimension  $720 \times 540$ . Chaque pixel  $u_{i,j}$  (avec  $1 \leq i \leq 720$  et  $1 \leq j \leq 540$ ) de cette image prend une valeur entre 0 (noir) et 255 (blanc), que l'on ramène à une valeur dans  $[0, 1]$  en divisant par 255. Nous altérons ensuite cette image en lui appliquant un bruit gaussien de moyenne nulle et de variance  $\sigma^2 = 0.01$  (voir Figure 3.5.6).

Afin de quantifier la qualité des images obtenues, nous comparons le PSNR (Peak Signal to Noise Ratio) (voir [108]) en fonction du nombre d'itérations pour chacun des algorithmes RIGPA, Accelerated Forward-Backward (AFB) et Forward-Backward (FB). Rappelons que le PSNR est une mesure dérivée de l'erreur quadratique moyenne et indique le rapport entre l'intensité maximale du pixel et la puissance de la distorsion. Rappelons enfin que bien que ces deux mesures soient simple à calculer, elle peuvent ne pas correspondre à la qualité visuelle observée, il est donc de plus nécessaire de comparer visuellement la qualité des images obtenues.



(b) Image bruitée



(a) Image originale

FIGURE 3.5.6 – (a) Image originale de taille  $720 \times 540$  pixels; (b) Image obtenue à partir de (a) après ajout d'un bruit blanc gaussien de moyenne nulle et de variance  $\sigma^2 = 0.01$  ;

Les paramètres utilisés pour chacun des algorithmes **RIGPA**, Accelerated Forward-Backward (AFB) et Forward-Backward (FB) sont les suivants :  $\lambda$  et  $\tau$  ont été pris égaux à 12 et 0.0013 respectivement, les autres paramètres sont

- $x_0 = y_0 = 0.5 \times (1, \dots, 1) \in \mathbb{R}^{720 \times 540}$ ,  $\kappa = 1.9$ ,  $e = 4$ ,  $\nu_n = n + 10$  pour **RIGPA**,
- $x_1 = x_0 = 0.5 \times (1, \dots, 1) \in \mathbb{R}^{720 \times 540}$ ,  $\alpha = 5$ , pour AFB,
- $x_0 = 0.5 \times (1, \dots, 1) \in \mathbb{R}^{720 \times 540}$  pour FB.

La Figure 3.5.7 montre l'évolution du PSNR (Peak Signal to Noise Ratio) pour chacun des algorithmes **RIGPA**, Accelerated Forward-Backward (AFB) et Forward-Backward (FB). On observe que les performances de **RIGPA** sont meilleures que celles de AFB jusqu'à environ 200 itérations. Après 200 itérations les performances de **RIGPA** et de AFB semblent être identiques. Les performances de FB sont quant à elles, inférieures à celles de **RIGPA** et de AFB.

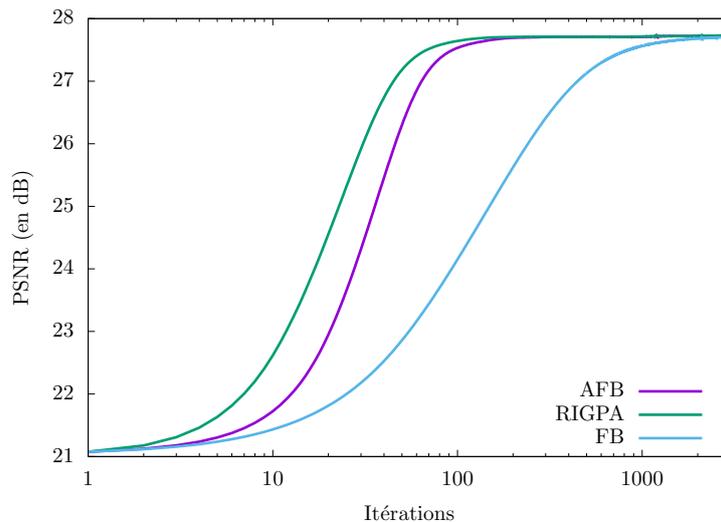


FIGURE 3.5.7 – Évolution du PSNR en fonctions du nombre d'itérations avec les algorithmes **RIGPA**, AFB et FB.

Les Figures 3.5.8, 3.5.9 et 3.5.10, montrent l'évolution de la reconstruction de l'image bruitée de la Figure 3.5.3.4 avec les algorithmes **RIGPA**, AFB et FB après 20, 250 et 1000 itérations. On observe qu'après 30 itérations, la qualité de l'image obtenue avec **RIGPA** est légèrement meilleure que celle de AFB (notamment en ce qui concerne le fond noir) et nettement meilleure que FB tandis qu'après 250 et 1000 itérations la qualité de l'image obtenue avec **RIGPA** et AFB semble être identique alors que FB apparaît toujours bruitée en particulier au niveau du visage.



FIGURE 3.5.8 – Représentation des images obtenues avec RIGPA (gauche), AFB (milieu) et FB (droite) après 20 itérations.



FIGURE 3.5.9 – Représentation des images obtenues avec RIGPA (gauche), AFB (milieu) et FB (droite) après 250 itérations.



FIGURE 3.5.10 – Représentation des images obtenues avec RIGPA (gauche), AFB (milieu) et FB (droite) après 1000 itérations.

## 3.6 Annexe

### 3.6.1 Démonstration de la Proposition 3.3.1

Soient  $\{\kappa, e, \nu_n\}$  des paramètres strictement positifs. Posons  $\varrho = 1 - \kappa$ ,  $\tau_n = e + \nu_{n+1}$  et  $u_n = y_n - x_n$ . Il est alors facile de voir que (3.3.8) peut être réécrit (pour  $n \geq p$ ) sous la forme

$$(3.6.1a) \quad \theta_n = \frac{1}{\tau_n} (\nu_n - \varrho \nu_{n+1}),$$

$$(3.6.1b) \quad \dot{x}_{n+1} + \chi_n^* + \theta_n u_n = 0,$$

$$(3.6.1c) \quad \dot{y}_{n+1} + \kappa u_n = 0.$$

Le reste de la démonstration peut être divisé en plusieurs parties **(r1)**-**(r4)** :

**(r1) Une estimation de la partie inertielle de la méthode.**

Commençons par montrer que la dérivée discrète  $\dot{G}_{n+1}(s, q)$  vérifie, pour  $(s, q) \in [0, +\infty[ \times \mathcal{H}$ ,

$$(3.6.2) \quad \begin{aligned} \dot{G}_{n+1}(s, q) + s\tau_n \langle \chi_n^*, x_{n+1} - q \rangle = \\ - (s\nu_n + \varrho \nu_{n+1}^2) \langle \dot{x}_{n+1}, u_n \rangle \\ - \frac{1}{2} (\nu_n^2 - \varrho^2 \nu_{n+1}^2) \|u_n\|^2 - \frac{1}{2} (se - \nu_{n+1}^2) \|\dot{x}_{n+1}\|^2. \end{aligned}$$

Afin d'obtenir ce résultat, remarquons que  $\dot{G}_{n+1}(s, q)$  peut être formulé comme suit

$$(3.6.3) \quad \dot{G}_{n+1}(s, q) = s(\dot{\nu}_{n+1}a_n + \nu_{n+1}\dot{a}_{n+1}) + se\dot{b}_{n+1} + \nu_{n+1}^2\dot{c}_{n+1} + c_n(\nu_{n+1}^2 - \nu_n^2),$$

où  $a_n := \langle q - x_n, u_n \rangle$ ,  $b_n := (1/2)\|x_n - q\|^2$  et où  $c_n := (1/2)\|u_n\|^2$ . Pour simplifier les notations, fixons les quantités suivantes

$$P_n := \langle q - x_{n+1}, \dot{x}_{n+1} \rangle, \quad R_n := \langle q - x_{n+1}, \dot{y}_{n+1} \rangle \quad \text{et} \quad W_n := \langle \chi_n^*, x_{n+1} - q \rangle.$$

Comme  $a_n = \langle q - x_n, u_n \rangle$  et  $u_n = -\frac{1}{\kappa}\dot{y}_{n+1}$  (d'après (3.6.1c)) on obtient

$$(3.6.4) \quad a_n = \langle q - x_{n+1}, u_n \rangle + \langle \dot{x}_{n+1}, u_n \rangle = -\frac{1}{\kappa}R_n + \langle \dot{x}_{n+1}, u_n \rangle.$$

En utilisant à nouveau la relation  $a_n = \langle q - x_n, u_n \rangle$  et en remarquant que  $\dot{u}_{n+1} =$

$\dot{y}_{n+1} - \dot{x}_{n+1}$  (car  $u_n = y_n - x_n$ ), on en déduit que

$$(3.6.5) \quad \dot{a}_{n+1} = \langle -\dot{x}_{n+1}, u_n \rangle + \langle q - x_{n+1}, \dot{u}_{n+1} \rangle = -\langle \dot{x}_{n+1}, u_n \rangle - P_n + R_n.$$

En prenant le produit scalaire de (3.6.1b) avec  $q - \dot{x}_{n+1}$  et en observant que  $u_n = -\frac{1}{\kappa}\dot{y}_{n+1}$ , on obtient  $P_n - W_n = \kappa^{-1}\theta_n R_n$ . Cette dernière égalité, combinée avec  $\theta_n = \tau_n^{-1}(\nu_n - \varrho\nu_{n+1})$  (d'après (3.6.1a)) nous donne

$$(3.6.6) \quad (\nu_n - \varrho\nu_{n+1})R_n = \kappa\tau_n(P_n - W_n).$$

On déduit alors de (3.6.4), (3.6.5) et (3.6.6)

$$(3.6.7) \quad \begin{aligned} & \dot{\nu}_{n+1}a_n + \nu_{n+1}\dot{a}_{n+1} \\ &= \dot{\nu}_{n+1} \left( \langle \dot{x}_{n+1}, u_n \rangle - \frac{1}{\kappa}R_n \right) + \nu_{n+1} \left( -\langle \dot{x}_{n+1}, u_n \rangle - P_n + R_n \right) \\ &= (\dot{\nu}_{n+1} - \nu_{n+1})\langle \dot{x}_{n+1}, u_n \rangle - \nu_{n+1}P_n + \left( \nu_{n+1} - \frac{1}{\kappa}\dot{\nu}_{n+1} \right) R_n \\ &= -\nu_n\langle \dot{x}_{n+1}, u_n \rangle - \nu_{n+1}P_n + \frac{1}{\kappa}(\nu_n - \varrho\nu_{n+1})R_n \\ &= -\nu_n\langle \dot{x}_{n+1}, u_n \rangle - \nu_{n+1}P_n + \tau_n(P_n - W_n). \end{aligned}$$

De plus, comme  $b_{n+1} = (1/2)\|x_{n+1} - q\|^2$ , on obtient

$$(3.6.8) \quad \dot{b}_{n+1} = \frac{1}{2}\langle \dot{x}_{n+1}, x_{n+1} - q \rangle + \frac{1}{2}\langle x_n - q, \dot{x}_{n+1} \rangle = -P_n - \frac{1}{2}\|\dot{x}_{n+1}\|^2.$$

Par ailleurs, comme  $c_{n+1} = (1/2)\|u_{n+1}\|^2$  et  $\dot{u}_{n+1} = -\kappa u_n - \dot{x}_{n+1}$  (car  $u_n = y_n - x_n$  et  $\dot{y}_{n+1} = -\kappa u_n$ ), on en déduit que

$$(3.6.9) \quad \begin{aligned} \dot{c}_{n+1} &= \frac{1}{2}\langle \dot{u}_{n+1}, u_{n+1} + u_n \rangle \\ &= \langle -\dot{u}_{n+1}, -\frac{1}{2}\dot{u}_{n+1} - u_n \rangle \\ &= \langle \kappa u_n + \dot{x}_{n+1}, \left(\frac{\kappa}{2} - 1\right)u_n + \frac{1}{2}\dot{x}_{n+1} \rangle \\ &= \frac{1}{2}\|\dot{x}_{n+1}\|^2 - \kappa\left(1 - \frac{\kappa}{2}\right)\|u_n\|^2 - \varrho\langle \dot{x}_{n+1}, u_n \rangle. \end{aligned}$$

Il résulte alors de (3.6.3) (3.6.7), (3.6.8) et (3.6.9) que

$$\begin{aligned} \dot{G}_{n+1}(s, q) &= s(-\nu_n\langle \dot{x}_{n+1}, u_n \rangle - \nu_{n+1}P_n + \tau_n(P_n - W_n)) + se\left(-P_n - \frac{1}{2}\|\dot{x}_{n+1}\|^2\right) \\ &\quad + \nu_{n+1}^2\left(\frac{1}{2}\|\dot{x}_{n+1}\|^2 - \kappa\left(1 - \frac{\kappa}{2}\right)\|u_n\|^2 - \varrho\langle \dot{x}_{n+1}, u_n \rangle\right) \\ &\quad + \frac{1}{2}(\nu_{n+1}^2 - \nu_n^2)\|u_n\|^2 \\ &= -\left(s\nu_n + \varrho\nu_{n+1}^2\right)\langle \dot{x}_{n+1}, u_n \rangle - \frac{1}{2}\left(se - \nu_{n+1}^2\right)\|\dot{x}_{n+1}\|^2 \\ &\quad - \bar{\eta}_n\|u_n\|^2 - s\tau_n W_n, \end{aligned}$$

où  $\bar{\eta}_n$  est donné par

$$\begin{aligned}\bar{\eta}_n &= \kappa \left(1 - \frac{\kappa}{2}\right) \nu_{n+1}^2 - \frac{1}{2}(\nu_{n+1}^2 - \nu_n^2) \\ &= \frac{1}{2} (\nu_n^2 - \nu_{n+1}^2 (1 - \kappa)^2) \quad (\text{car } \kappa \left(1 - \frac{\kappa}{2}\right) = \frac{1}{2} - \frac{1}{2}(1 - \kappa)^2).\end{aligned}$$

Ceci nous donne le résultat voulu.

**(r2) Une estimation de la partie proximale de la méthode.**

Montrons que, pour tout  $\xi_n \neq 1$ , on a

$$(3.6.10) \quad \begin{aligned}\xi_n \langle \chi_n^*, \dot{x}_{n+1} \rangle + \frac{1}{2} \|\dot{x}_{n+1} + \theta_n u_n\|^2 \\ = \theta_n (1 - \xi_n) \langle u_n, \dot{x}_{n+1} \rangle + \frac{1}{2} \theta_n^2 \|u_n\|^2 - \left(\xi_n - \frac{1}{2}\right) \|\dot{x}_{n+1}\|^2.\end{aligned}$$

En effet, d'après (3.6.1b), nous avons  $\dot{x}_{n+1} = -\theta_n u_n - \chi_n^*$ . Ce qui, peut être réécrit, pour tout  $\xi_n \neq 1$ , sous la forme

$$(3.6.11) \quad \xi_n \dot{x}_{n+1} = -(1 - \xi_n) (\dot{x}_{n+1} + (1 - \xi_n)^{-1} \theta_n u_n) - \chi_n^* = -(1 - \xi_n) H_n - \chi_n^*,$$

avec  $H_n = \dot{x}_{n+1} + (1 - \xi_n)^{-1} \theta_n u_n$ . De plus, comme  $-\chi_n^* = \dot{x}_{n+1} + \theta_n u_n$  (d'après (3.6.1b)), on obtient, en notant  $Q_n = \langle \dot{x}_{n+1}, u_n \rangle$

$$(3.6.12) \quad \begin{aligned}\langle (-\chi_n^*), H_n \rangle &= \langle \dot{x}_{n+1} + \theta_n u_n, \dot{x}_{n+1} + (1 - \xi_n)^{-1} \theta_n u_n \rangle \\ &= \|\dot{x}_{n+1}\|^2 + (1 - \xi_n)^{-1} \theta_n^2 \|u_n\|^2 + \frac{2 - \xi_n}{(1 - \xi_n)} \theta_n Q_n.\end{aligned}$$

Par conséquent, en prenant le produit scalaire du côté gauche de (3.6.11) avec  $\chi_n^*$  et en ajoutant  $(1/2)\|\chi_n^*\|^2$ , on obtient d'après (3.6.12) et  $\|\chi_n^*\|^2 = \|\dot{x}_{n+1} + \theta_n u_n\|^2$ ,

$$\begin{aligned}\xi_n \langle \chi_n^*, \dot{x}_{n+1} \rangle + \frac{1}{2} \|\chi_n^*\|^2 &= (1 - \xi_n) \langle (-\chi_n^*), H_n \rangle - \frac{1}{2} \|\chi_n^*\|^2 \\ &= (1 - \xi_n) \left( \|\dot{x}_{n+1}\|^2 + \frac{\theta_n^2}{(1 - \xi_n)} \|u_n\|^2 + \frac{2 - \xi_n}{(1 - \xi_n)} \theta_n Q_n \right) \\ &\quad - \frac{1}{2} (\|\dot{x}_{n+1}\|^2 + \theta_n^2 \|u_n\|^2 + 2\theta_n Q_n) \\ &= (1 - \xi_n) \theta_n Q_n + \frac{1}{2} \theta_n^2 \|u_n\|^2 + \left(\frac{1}{2} - \xi_n\right) \|\dot{x}_{n+1}\|^2.\end{aligned}$$

Ceci nous donne (3.6.10).

**(r3) Association des effets proximaux et inertiels.**

Supposons que la condition (3.3.10) soit vérifiée et posons  $\rho_n = 1 - (1 - \kappa) \frac{\nu_{n+1}}{\nu_n}$ . Remarquons que le terme  $\theta_n$  intervenant dans (3.6.1) peut être exprimé simplement

comme suit

$$(3.6.13) \quad \theta_n = \frac{\nu_n \rho_n}{\tau_n} \quad \text{où } \tau_n = e + \nu_{n+1}.$$

Ainsi, comme  $\rho_n > 0$  (d'après la condition (3.3.3a)), on déduit de (3.6.13) que  $\theta_n$  est une suite strictement positive.

Introduisons à présent la suite réelle  $(\bar{\gamma}_n)$  définie par

$$(3.6.14) \quad \bar{\gamma}_n = 1 - \frac{s\rho_n}{\tau_n} \quad (\text{avec } s > 0).$$

D'après (3.6.14), et comme  $\tau_n > 0$  (car  $\tau_n := e + \nu_{n+1}$ ) et  $\rho_n > 0$ , nous avons

$$(3.6.15) \quad \bar{\gamma}_n < 1 \quad (\text{pour tout } s > 0).$$

Ensuite, pour  $s > 0$  donné, nous montrons que les itérés générés par (3.3.8) (ou, de manière équivalente, par (3.6.1)) vérifient

$$(3.6.16) \quad \begin{aligned} \dot{G}_{n+1}(s, q) + \frac{1}{2}\rho_n^{-1}\tau_n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2 \\ + (s\tau_n)\langle \chi_n^*, x_{n+1} - q \rangle + \bar{\gamma}_n \rho_n^{-1} \tau_n^2 \langle \chi_n^*, \dot{x}_{n+1} \rangle = -T_n(u_n, \dot{x}_{n+1}), \end{aligned}$$

où  $T_n(u, x)$  est défini pour tout  $(u, x) \in \mathcal{H}^2$  par

$$(3.6.17) \quad T_n(u, x) = w_n \langle u, x \rangle + \eta_n \|u\|^2 + \sigma_n \|x\|^2,$$

avec les paramètres

$$(3.6.18a) \quad w_n = \varrho (\nu_{n+1}^2 + s\nu_{n+1}), \quad \eta_n = \frac{1}{2}\varrho\rho_n\nu_n\nu_{n+1},$$

$$(3.6.18b) \quad \sigma_n = \frac{1}{2}(se - \nu_{n+1}^2 + \rho_n^{-1}\tau_n^2(2\bar{\gamma}_n - 1)).$$

En effet, en posant  $Q = \langle \dot{x}_{n+1}, u_n \rangle$ , nous savons d'après (3.6.2) que

$$(3.6.19) \quad \begin{aligned} \dot{G}_{n+1}(s, q) + (s\tau_n)\langle \chi_n^*, x_{n+1} - q \rangle = \\ - (s\nu_n + \varrho\nu_{n+1}^2) Q_n - \frac{1}{2}(\nu_n^2 - \varrho^2\nu_{n+1}^2) \|u_n\|^2 - \frac{1}{2}(se - \nu_{n+1}^2) \|\dot{x}_{n+1}\|^2. \end{aligned}$$

De plus, comme  $\bar{\gamma}_n \neq 1$  (d'après (3.6.15)), on obtient en utilisant (3.6.10) avec la

valeur  $\xi_n = \bar{\gamma}_n := 1 - \frac{s\rho_n}{\tau_n}$  et en rappelant que  $\theta_n = \frac{\nu_n\rho_n}{\tau_n}$ ,

$$(3.6.20) \quad \begin{aligned} \bar{\gamma}_n \langle \chi_n^*, \dot{x}_{n+1} \rangle + \frac{1}{2} \|\dot{x}_{n+1} + \theta_n u_n\|^2 \\ = s \frac{\nu_n \rho_n^2}{\tau_n^2} Q_n + \frac{1}{2} \frac{\nu_n^2 \rho_n^2}{\tau_n^2} \|u_n\|^2 - \left(\bar{\gamma}_n - \frac{1}{2}\right) \|\dot{x}_{n+1}\|^2. \end{aligned}$$

Puis, en multipliant l'égalité (3.6.20) par  $\rho_n^{-1}\tau_n^2$ , et en ajoutant à (3.6.19) l'égalité obtenue, on obtient

$$\begin{aligned} \dot{G}_{n+1}(s, q) + (s\tau_n) \langle \chi_n^*, x_{n+1} - q \rangle \\ + \bar{\gamma}_n \rho_n^{-1} \tau_n^2 \langle \chi_n^*, \dot{x}_{n+1} \rangle + \frac{1}{2} \rho_n^{-1} \tau_n^2 \|\dot{x}_{n+1} + \theta_n u_n\|^2 \\ = \left(- (s\nu_n + \varrho\nu_{n+1}^2) + s\nu_n\rho_n\right) Q_n \\ + \left(-\frac{1}{2} (\nu_n^2 - \varrho^2\nu_{n+1}^2) + \frac{1}{2}\nu_n^2\rho_n\right) \|u_n\|^2 \\ + \left(\frac{1}{2} (se - \nu_{n+1}^2) + \rho_n^{-1}\tau_n^2 (\bar{\gamma}_n - \frac{1}{2})\right) \|\dot{x}_{n+1}\|^2. \end{aligned}$$

Par conséquent, en remarquant que  $\nu_n\rho_n = \nu_n - \varrho\nu_{n+1}$ , nous déduisons que (3.6.16)-(3.6.17) est vérifié avec les paramètres

$$\begin{aligned} w_n &= (s\nu_n + \varrho\nu_{n+1}^2) - s(\nu_n - \varrho\nu_{n+1}) = \varrho(\nu_{n+1}^2 + s\nu_{n+1}), \\ \eta_n &= \frac{1}{2} (\nu_n^2 - \nu_{n+1}^2\varrho^2) - \frac{1}{2} (\nu_n^2 - \varrho\nu_n\nu_{n+1}) = \frac{1}{2}\varrho\nu_{n+1}(\nu_n - \nu_{n+1}\varrho) = \frac{1}{2}\varrho\nu_{n+1}\nu_n\rho_n, \\ \sigma_n &= \frac{1}{2} (se - \nu_{n+1}^2) + \rho_n^{-1}\tau_n^2 (\bar{\gamma}_n - \frac{1}{2}) = \frac{1}{2} (se - \nu_{n+1}^2 + \rho_n^{-1}\tau_n^2 (2\bar{\gamma}_n - 1)). \end{aligned}$$

**(r4)** Enfin, nous donnons une formulation alternative de la quantité  $T_n(u, x)$  donnée par (3.6.17)-(3.6.18). Pour ce faire, nous allons reformuler la suite  $(\sigma_n)$ . D'après les définitions de  $\tau_n := e + \nu_{n+1}$  et de  $\bar{\gamma}_n := 1 - s\frac{\rho_n}{\tau_n}$ , on obtient

$$\begin{aligned} \rho_n^{-1}\tau_n^2(2\gamma_{s,n} - 1) &= \frac{(e+\nu_{n+1})^2}{\rho_n} \left(1 - 2s\frac{\rho_n}{(e+\nu_{n+1})}\right) \\ &= \frac{1}{\rho_n} (e^2 + 2e\nu_{n+1} + (\nu_{n+1})^2) - 2s(e + \nu_{n+1}) \\ &= e\left(\frac{e}{\rho_n} - s\right) - se + 2\nu_{n+1}\left(\frac{e}{\rho_n} - s\right) + \frac{(\nu_{n+1})^2}{\rho_n} \\ &= (e + 2\nu_{n+1})\left(\frac{e}{\rho_n} - s\right) - se + \frac{(\nu_{n+1})^2}{\rho_n} \\ &= \tau_{n,e}(e\rho_n^{-1} - s) - se + \rho_n^{-1}(\nu_{n+1})^2, \end{aligned}$$

où  $\tau_{n,t} = t + 2\nu_{n+1}$  (pour  $t \geq 0$ ). Par conséquent, en utilisant la définition précédente de  $\sigma_n$  (donnée par (3.6.18)), on a

$$(3.6.21) \quad \begin{aligned} 2\sigma_n &= (\rho_n^{-1} - 1)(\nu_{n+1})^2 + \tau_{n,e}(e\rho_n^{-1} - s) \\ &= ((\nu_{n+1})^2 + e\tau_{n,e})(\rho_n^{-1} - 1) + \tau_{n,e}(e - s). \end{aligned}$$

Considérons alors les deux situations suivantes selon les valeurs de  $\kappa$  :

- Dans le cas particulier où  $\kappa = 1$  (donc  $\rho_n = 1$  et  $\rho_n^{-1} = 1$ ), on a clairement  $w_n = 0$  et  $\eta_n = 0$ . La définition de  $T_n$  (donnée par (3.6.17)) avec  $\sigma_n = \frac{(e-s)}{2}\tau_{n,e}$  (d'après (3.6.21)) donne alors pour  $(u, x) \in \mathcal{H}^2$

$$(3.6.22) \quad T_n(u, x) = \frac{(e-s)}{2}\tau_{n,e}\|x\|^2.$$

- Pour  $\kappa \in (0, 1) \cup (1, \infty)$  (donc  $\eta_n \neq 0$ ), en posant  $\varsigma_n := \frac{w_n}{2\eta_n}$ , et  $\psi_n := 4\sigma_n\eta_n - w_n^2$ , on obtient par définition de  $T_n$

$$(3.6.23) \quad T_n(u, x) = \eta_n\|u + \varsigma_n x\|^2 + \frac{\psi_n}{4\eta_n}\|x\|^2.$$

D'une part, comme  $w_n = \varrho\nu_{n+1}(\nu_{n+1} + s)$  (d'après (3.6.18)) et que  $\tau_{n,s} = s + 2\nu_{n+1}$ , on a

$$w_n^2 = (\varrho\nu_{n+1})^2((\nu_{n+1})^2 + s\tau_{n,s}).$$

En posant  $S_n = \varrho\rho_n\nu_n\nu_{n+1}$  (de sorte que  $S_n = 2\eta_n$  et  $\psi_n = 2\sigma_n S_n - w_n^2$ ), on obtient en utilisant (3.6.21) et la définition de  $\psi_n$

$$\psi_n = S_n((\nu_{n+1})^2 + e\tau_{n,e})(\rho_n^{-1} - 1) + S_n\tau_{n,e}(e - s) - (\varrho\nu_{n+1})^2((\nu_{n+1})^2 + s\tau_{n,s}).$$

Par ailleurs, on vérifie aisément que  $S_n(\rho_n^{-1} - 1) = (\varrho\nu_{n+1})^2$ , ce qui, en utilisant l'égalité précédente, donne

$$(3.6.24) \quad \psi_n = S_n\tau_{n,e}(e - s) + (\varrho\nu_{n+1})^2(e\tau_{n,e} - s\tau_{n,s}).$$

Ainsi, en remarquant que  $e\tau_{n,e} - s\tau_{n,s} = (e - s)\tau_{n,e+s}$  (car  $\tau_{n,t} := t + 2\nu_{n+1}$ , pour tout  $t \geq 0$ ), on en déduit que

$$\psi_n = (e - s)(S_n\tau_{n,e} + (\varrho\nu_{n+1})^2\tau_{n,e+s}),$$

ce qui par (3.6.23) entraîne que

$$T_n(u, x) = \frac{1}{2}S_n\|u + \varsigma_n x\|^2 + \frac{(e - s)}{2S_n}(S_n\tau_{n,e} + (\varrho\nu_{n+1})^2\tau_{n,e+s})\|x\|^2.$$

D'autre part, on a  $\varsigma_n = \frac{w_n}{S_n}$  (car  $S_n = 2\eta_n$ ), ainsi que  $w_n = \varrho\nu_{n+1}(\nu_{n+1} + s)$ ,

$S_n = \varrho \rho_n \nu_n \nu_{n+1}$  et  $\frac{1}{\theta_n} = \frac{e+\nu_{n+1}}{\rho_n \nu_n}$  (d'après (3.6.13)). On obtient alors

$$\varsigma_n = \frac{\nu_{n+1}+s}{\rho_n \nu_n} = \frac{(e+\nu_{n+1})-(e-s)}{\rho_n \nu_n} = \frac{1}{\theta_n} - \frac{(e-s)}{\nu_n \rho_n}.$$

En combinant ces deux derniers résultats, nous obtenons

$$T_n(u, x) = \frac{1}{2} S_n \left\| u + \left( \frac{1}{\theta_n} - \frac{e-s}{\nu_n \rho_n} \right) x \right\|^2 + \frac{(e-s)}{2} \left( \tau_{n,e} + \frac{(\varrho \nu_{n+1})^2}{S_n} \tau_{n,e+s} \right) \|x\|^2.$$

Ceci achève la preuve . ■

# Chapitre 4

## Algorithmes forward-backward accélérés pour la résolution d'inclusions monotones structurées

---

### 4.1 Introduction

Dans ce chapitre,  $\mathcal{H}$  désigne un espace de Hilbert réel muni d'un produit scalaire noté  $\langle \cdot, \cdot \rangle$  et de la norme induite notée  $\| \cdot \|$ . L'objectif de ce chapitre est de concevoir des méthodes proximales rapides pour la résolution d'inclusions monotones structurées. Nous nous concentrons, dans un premier temps, au plus simple de ces problèmes qui s'écrit :

$$(4.1.1) \quad \text{trouver } \bar{x} \in \mathcal{H} \text{ tel que } 0 \in A\bar{x} + B\bar{x},$$

sous les conditions :

$$(4.1.2a) \quad A : \mathcal{H} \rightarrow 2^{\mathcal{H}} \text{ est un opérateur maximal monotone,}$$

$$(4.1.2b) \quad B : \mathcal{H} \rightarrow \mathcal{H} \text{ est un opérateur maximal monotone et } \lambda\text{-cocoercif,}$$

$$(4.1.2c) \quad S := (A + B)^{-1}(0) \neq \emptyset.$$

**Définition 4.1.1.** *On rappelle qu'un opérateur  $B : \mathcal{H} \rightarrow \mathcal{H}$  est dit  $\lambda$ -cocoercif (pour un certain  $\lambda > 0$ ) si pour tout  $(x, y) \in \mathcal{H}^2$ ,  $\langle Bx - By, x - y \rangle \geq \lambda \|Bx - By\|^2$ .*

Une méthode classique pour résoudre (4.1.1) est l'algorithme Forward-Backward (FB, en abrégé) qui s'écrit

$$(4.1.3) \quad x_{n+1} = J_{\mu_n}^A(x_n - \mu_n Bx_n),$$

où  $(\mu_n) \subset ]0, 2\lambda[$  et où  $J_{\mu_n}^A := (I + \mu_n A)^{-1}$  est la résolvante de  $A$  d'indice  $\mu_n$  qui est bien connue pour être univoque et partout définie (voir [31, 58, 97])

Il est bien connu que (4.1.1)-(4.1.2) inclut le problème de minimisation convexe structuré non lisse

$$(4.1.4) \quad \min_{\mathcal{H}} \{\Theta := f + g\},$$

où  $f : \mathcal{H} \rightarrow ]-\infty, +\infty[$  et  $g : \mathcal{H} \rightarrow ]-\infty, +\infty[$  sont telles que

$$(4.1.5a) \quad f \text{ est convexe, différentiable et son gradient est lipschitzien,}$$

$$(4.1.5b) \quad g \text{ est convexe propre et semi-continue inférieurement,}$$

$$(4.1.5c) \quad S := \operatorname{argmin}(f + g) \neq \emptyset.$$

En effet, rappelons que le gradient d'une fonction convexe et Fréchet-différentiable est cocoercif dès lors qu'il est lipschitzien (voir [22, Corollary 18.17]). Par conséquent, le problème (4.1.4) se trouve être un cas particulier de (4.1.1) lorsque  $B = \nabla f$  et  $A = \partial g$  ( $\nabla f$  étant le gradient de  $f$  et  $\partial g$  le sous-différentiel de Fenchel de  $g$ ). Ainsi, dans ce cadre, l'opérateur résolvante  $J_{\mu}^A$  se réduit à l'opérateur proximal de  $g$  d'indice  $\mu > 0$  défini pour  $x \in \mathcal{H}$  par

$$(4.1.6) \quad \operatorname{prox}_{\mu g}(x) := \operatorname{argmin}_{y \in \mathcal{H}} (g(y) + (2\mu)^{-1} \|x - y\|^2).$$

Il a été montré (voir chapitre 1, Section 1.2.3) que lorsque la méthode Forward-Backward (FB) est appliquée au problème de minimisation (4.1.4), celle-ci génère des suites  $(x_n)$  faiblement convergentes vérifiant  $\Theta(x_n) - \inf_{\mathcal{H}} \Theta = \mathcal{O}(n^{-1})$  (pour la convergence en valeur) ainsi que  $\|x_{n+1} - x_n\| = \mathcal{O}(n^{-1})$  (concernant la vitesse discrète). Ces estimations ont été considérablement améliorées via l'algorithme AFB (décrit dans le chapitre 1, Section 1.2.4). Ce dernier produit en effet des suites  $(x_n)$  vérifiant les taux de convergences  $\Theta(x_n) - \inf_{\mathcal{H}} \Theta = o(n^{-2})$  et  $\|x_{n+1} - x_n\| = o(n^{-1})$ . Par la suite, diverses variantes de AFB (généralisées à des opérateurs maximaux

monotones  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  et  $B : \mathcal{H} \rightarrow \mathcal{H}$ ) ont été abondamment étudiées dans le but d'obtenir des taux de convergence similaires à ceux obtenus dans le cas de la minimisation convexe. Cependant, la plupart de ces travaux concernent des résultats empiriques.

Le but de ce chapitre est d'étendre, en termes de vitesse discrète et de résidus de point fixe, les propriétés de convergence obtenues ci-dessus concernant la minimisation convexe structurée dans le cadre d'opérateurs maximaux monotones. Plus précisément, nous introduisons un algorithme de type Forward-Backward inertiel relaxé que nous noterons **ALG1** consistant en des suites  $\{x_n, y_n, z_n\}$  générées par le schéma suivant

### ALG1

► **Initialisation** : Choisir  $\{z_{-1}, x_0, y_0\} \subset \mathcal{H}$ .

► **Étape principale** : Pour  $\{z_{n-1}, x_n, y_n\} \subset \mathcal{H}$  donnés, on effectue

$$(4.1.7a) \quad z_n = x_n - \theta_n(y_n - x_n) - \gamma_n(x_n - z_{n-1}),$$

$$(4.1.7b) \quad x_{n+1} = \left(1 - \frac{t}{\mu}\right) z_n + \left(\frac{t}{\mu}\right) J_{\mu}^A(z_n - \mu B(z_n)),$$

$$(4.1.7c) \quad y_{n+1} = (1 - \kappa)y_n + \kappa x_n.$$

Avec  $\{\kappa, t, \mu\} \subset ]0, +\infty[$  et où les suites  $(\theta_n)_n$  et  $\gamma_n$  sont définies par :

$$(4.1.8a) \quad \theta_n := \frac{\nu_n - (1 - \kappa)\nu_{n+1}}{e + \nu_{n+1}} \text{ et } \gamma_n := 1 - \frac{s_0}{e + \nu_{n+1}} \left(1 - (1 - \kappa)^{\frac{\nu_{n+1}}{\nu_n}}\right).$$

Les autres paramètres sont :

$$(4.1.8b) \quad e > 0, \quad s_0 \in ]0, e[ \text{ et } (\nu_n)_{n \geq 0} \subset ]0, +\infty[.$$

En tant qu'élément clé de notre méthodologie, nous introduisons l'opérateur  $M_{\mu}(\cdot)$  (voir par exemple [8, 13]) défini pour  $z \in \mathcal{H}$  par

$$(4.1.9) \quad M_{\mu}(z) = \frac{1}{\mu}(z - J_{\mu}^A(z - \mu Bz)).$$

Rappelons que  $M_{\mu}(z) = 0$  est équivalent à  $z \in S$ . En particulier, en utilisant des paramètres appropriés, nous établissons, parmi d'autres résultats, la convergence faible des itérés  $(x_n)$  ainsi que les taux  $\|x_{n+1} - x_n\| = o(n^{-1})$  (pour la vitesse discrète) et  $\|M_{\mu}(z)\| = o(n^{-1})$  (pour les résidus des points fixes).

Notons que lorsque  $A = \partial g$ ,  $B = \nabla f$ ,  $\gamma_n = 0$  et  $t = \mu$  dans [ALG1](#), nous retrouvons l'algorithme [RIGPA](#) étudié au chapitre 3 dans le cadre de la minimisation convexe structurée. Un algorithme similaire à [ALG1](#) a été discuté par Maingé [\[71\]](#) dans le cas particulier où  $\kappa = 1$ . Nous proposons de généraliser ce travail pour  $\kappa \in ]0, 2[$  car ce terme agit comme un paramètre d'extrapolation, en particulier lorsque  $\kappa$  se trouve dans la plage de sur-relaxation  $]1, 2[$ , comme observé dans les simulations numériques de l'algorithme [RIGPA](#) (voir Section [3.5](#)).

**Remarque 4.1.12.** *Faisons les deux observations suivantes.*

- (1) *Comme établi dans la section [4.6.1](#), l'algorithme [ALG1](#) peut se réécrire pour  $n \geq 1$ , sous la forme plus classique :*

$$(4.1.10a) \quad z_{n-1} = y_n + \kappa^{-1}(\xi_n - \gamma_n)(y_n - y_{n-1}) - \gamma_{n-1}(y_{n-1} - z_{n-2}),$$

$$(4.1.10b) \quad x_n = \left(1 - \frac{t}{\mu}\right) z_{n-1} + \left(\frac{t}{\mu}\right) J_{\mu}^A(z_{n-1} - \mu B(z_{n-1})),$$

$$(4.1.10c) \quad y_{n+1} = (1 - \kappa)y_n + \kappa x_n,$$

où  $(\xi_n)$  est une suite strictement positive dépendant de  $(\theta_n)$ .

- (2) *En suivant la même méthodologie que Maingé [\[71\]](#) (voir aussi Raguet-Fadili-Peyré [\[94\]](#)), nous adapterons les techniques d'accélération proposées dans ce chapitre au problème d'inclusion plus général :*

$$(4.1.11) \quad \text{trouver } \bar{x} \in \mathcal{H} \text{ tel que } 0 \in B\bar{x} + \sum_{k=1}^p A_k \bar{x},$$

où  $B : \mathcal{H} \rightarrow \mathcal{H}$  est  $\lambda$ -cocoercif et où  $(A_i)_{i=1}^p : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  est une famille de  $p$  opérateurs maximaux monotones. Nous proposerons un algorithme de type Forward-backward généralisé (que nous appellerons [ALG2](#)) présentant des similitudes avec l'algorithme FB. En effet, [ALG2](#) consiste en une étape « forward » explicite, suivie d'une étape implicite (« backward ») dans laquelle la résolvante de chaque  $A_i$  est calculée en parallèle.

## 4.1.1 Rappels sur les méthodes d'éclatement et les techniques d'accélération

### 4.1.1.1 Quelques méthodes d'éclatements classique

Les méthodes d'éclatement (pour le calcul des zéros de la somme de deux opérateurs maximaux monotones) peuvent être divisées en deux grandes catégories :

(a) Les méthodes d'éclatement qui procèdent essentiellement en des étapes « backward » (voir Peaceman-Rachford et Douglas-Rachford ([51, 90]).

(b) Les méthodes d'éclatement qui combinent à la fois des étapes « backward » et « forward ».

- (b1) Un exemple classique de ces méthodes est l'algorithme « forward-backward » (FB) donné par (4.1.3). En particulier, lorsque  $B$  est  $\beta$ -cocoercif, il a été démontré que la méthode FB génère des suites faiblement convergentes dès lors que  $(\mu_n) \subset ]0, 2\beta[$ .

- (b2) Un autre exemple de méthode lorsque  $B$  est univoque est l'algorithme « forward-backward-forward » proposé par Tseng [106]. Cet algorithme génère des suites faiblement convergentes lorsque  $B$  est  $L$ -lipschitzien dès lors que  $(\mu_n) \subset ]0, 1/L[$ .

- (b3) Par la suite, des variantes plus générales des méthodes forward-backward et forward-backward-forward ont été adaptés à des problèmes d'inclusions monotones plus générales tels que ceux de la forme (4.1.11) (voir [43, 94]).

### 4.1.1.2 Des variantes accélérées

Notons que, dans le cadre de la minimisation convexe, la méthode forward-backward 4.1.3 devient

$$(4.1.12) \quad x_{n+1} = \text{prox}_{\mu_n}(x_n - \mu_n \nabla f(x_n)).$$

Lorsque  $\nabla f$  est  $L$ -lipschitzien et que  $(\mu_n) \subset ]0, L^{-1}[$ , cette méthode est bien connue pour générer des suites  $(x_n)$  faiblement convergentes vérifiant le taux sous-linéaire  $(f + g)(x_n) - \inf_{\mathcal{H}}(f + g) = \mathcal{O}(n^{-1})$  (voir chapitre 1, section 1.2.3).

Cette méthode fut, par la suite, améliorée via notamment l'algorithme « Fast Iterative Thresholding Algorithm » (FISTA) proposé par Beck-Teboulle [24] qui est basé sur les techniques d'accélération de Güler et Nesterov. Il a été démontré que FISTA génère des suites qui garantissent le taux  $(f + g)(x_n) - \inf_{\mathcal{H}}(f + g) = \mathcal{O}(n^{-2})$ , cependant, la convergence de ces itérés n'a pas été établie.

Ce défaut fut corrigé par la variante suivante de FISTA, récemment introduite par Chambolle-Dossal [40] (voir aussi Attouch-Peypouquet [14]) donnée par

$$(4.1.13) \quad \begin{aligned} z_n &= x_n + \frac{n-1}{n+\alpha-1}(x_n - x_{n-1}), \\ x_{n+1} &= \text{prox}_\mu(x_n - \mu \nabla f(x_n)), \end{aligned}$$

où  $\mu \in ]0, L^{-1}[$  et  $\alpha > 0$ . Pour  $\alpha > 3$ , les auteurs de [14]) ont démontré que (4.1.13) génère des suites  $(x_n)$  faiblement convergentes qui vérifient le taux de convergence en valeur  $(f + g)(x_n) - \inf_{\mathcal{H}}(f + g) = o(n^{-2})$  au lieu des taux  $\mathcal{O}(n^{-1})$  et  $\mathcal{O}(n^{-2})$  pour (4.1.3) et FISTA respectivement.

Dans le cas d'un opérateur maximal monotone  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  quelconque, des variantes accélérées de la méthode PPA ont été proposées via notamment les algorithmes « RIPA » (Regularized Inertial Proximal Algorithm) et « PRINAM » (Proximal Regularized Inertial Newton Algorithm) étudiés respectivement par Attouch-Peypouquet [13] et Attouch-László [12]. Malgré des propriétés asymptotiques intéressantes, ces algorithmes requièrent un indice proximal non borné et ne peuvent donc pas être étendus au cadre des méthodes forward-backward.

Dans le même contexte, une méthode du point proximal accélérée utilisant un indice proximal constant a été proposée par Kim [65]. Cette dernière est basée sur l'approche *Performance Estimation Problem* (PEP) introduite par Drori-Teboulle [52]. Cette méthode fournit le taux de convergence  $\mathcal{O}(n^{-2})$  pour les résidus de points fixes. La convergence des itérés n'a, cependant, pas été démontrée.

À notre connaissance, il n'existe pas de résultats de convergence qui permettent d'étendre les estimations obtenues pour (4.1.13) (concernant la minimisation convexe) au problème d'inclusion monotone générale (4.1.1)-(4.1.2). Excepté l'algorithme étudié par Attouch-Cabot dans [8], seules quelques méthodes accélérées empiriques basées sur des techniques inertielles et de relaxation ont été proposées. Rappelons quelques-uns des ces algorithmes

(c1) Des variantes inertielles de (4.1.3) avec un opérateur  $B$  cocoercif, ont été étudié par Moudafi-Oliny [78]

$$(4.1.14) \quad z_n = x_n + \alpha_n(x_n - x_{n-1}), \quad x_{n+1} = J_{\mu_n}^A(z_n - \mu_n B(x_n)),$$

et Lorenz-Pock [70]

$$(4.1.15) \quad z_n = x_n + \alpha_n(x_n - x_{n-1}), \quad x_{n+1} = J_{\mu_n}^A(z_n - \mu_n B(z_n)),$$

où  $(\alpha_n)$  et  $(\mu_n)$  sont des suites strictement positives. Notons que dans (4.1.15), l'évaluation de l'opérateur  $B$  se fait en la variable  $z_n$  (au lieu de  $x_n$  pour (4.1.14)).

(c2) Une variante de (4.1.3) de type réflexion pour un opérateur lipschitzien  $B$  a été étudié pour Cevher-Vu [38] :

$$(4.1.16) \quad y_n = 2x_n - x_{n-1}, \quad x_{n+1} = J_{\mu}^A(x_n - \mu B y_n).$$

(c3) Une variante inertielle relaxée de (4.1.3) pour un opérateur  $B$   $\beta$ -cocoercif a été étudié par Attouch-Cabot [8]

$$(4.1.17) \quad z_n = x_n + \alpha_n(x_n - x_{n-1}), \quad x_{n+1} = (1 - w_n)z_n + w_n J_{\mu_n}^A(z_n - \mu_n B(z_n)),$$

où  $\{\alpha_n, w_n, \mu_n\}$  sont des suites bornées strictement positives. Sous diverses conditions portant sur ces derniers paramètres, les auteurs ont établi la convergence faible des itérés  $(x_n)$  ainsi que des taux de convergence concernant la vitesse discrète et de résidus de points fixes  $\|M_{\mu}(z_n)\|$  (où  $M_{\mu}$  est défini par (4.1.9). En particulier, la méthode (4.1.17) fait intervenir des techniques d'accélération de Nesterov via le choix  $\alpha_n := 1 - \alpha n^{-1}$  avec  $\alpha > 2$  (pour le coefficient inertielle),  $\mu_n = \mu \in ]0, 2\beta[$  (pour l'indice proximal) et  $w_n = 1 - \rho n^{-2}$  où  $0 < \rho < \alpha(\alpha - 2)(1 - \frac{\mu}{4\beta})$  (concernant le terme de relaxation). Avec ces paramètres, les auteurs ont démontré (voir [9, Corollary 4.9]) les estimations  $\|x_{n+1} - x_n\| = \mathcal{O}(n^{-1})$  et  $\sum_n n \|x_{n+1} - x_n\|^2 < +\infty$  (pour la vitesse discrète) ainsi que  $\sum_n n^{-1} \|M_{\mu}(x_n)\|^2 < +\infty$  et  $\lim_{n \rightarrow +\infty} \|M_{\mu}(x_n)\| = 0$  (concernant les résidus de points fixes).

Grâce au terme correcteur  $\gamma_n$  de ALG1, nous améliorons ces derniers taux de convergence en  $\|x_{n+1} - x_n\| = o(n^{-1})$  et  $\|M_{\mu}(x_n)\| = o(n^{-1})$ .

## 4.1.2 Esquisse des principaux résultats

Un nouvelle classe d'algorithme (donnée par (4.1.7)-(4.1.8)) est proposée pour résoudre des problèmes d'inclusions monotones générales. Plusieurs résultats de convergence sont établis concernant la méthode proposée dans le cas de paramètres

généraux (Théorème 4.3.1) ou particuliers (Théorème 4.3.2). Notre approche est basée sur les propriétés de Lyapunov d'une suite de type énergie. Plus précisément, nous prouvons, sous certaines conditions sur les paramètres  $\{e, s_0, t, \mu, \nu_n\}$ , que les suites  $\{x_n, z_n, y_n\}$  produites par **ALG1** convergent faiblement un élément de  $S := (A + B)^{-1}(0)$ .

En particulier, pour  $\nu_n \sim n$  (lorsque  $n \rightarrow +\infty$ ) nous obtenons les estimations  $\|x_{n+1} - x_n\| = o(n^{-1})$  (pour la vitesse discrète) et  $\|M_\mu(z_n)\| = o(n^{-1})$  (concernant les résidus de points fixes), ce qui améliore grandement les estimations obtenues jusqu'à présent. De plus, nous exhibons une suite  $\xi_n$  convergeant faiblement vers un élément de  $S$  pour laquelle nous démontrons l'existence d'une suite  $\xi_n^* \in (A + B)(\xi_n)$  vérifiant

$$(4.1.18) \quad \|\xi_n^*\| = o(n^{-1}) \quad \text{et} \quad \sum_n n \|\xi_n^*\|^2 < +\infty.$$

Notons que ces estimations sont similaires à celles obtenues pour des sous-gradients par l'algorithme « Accelerated Forward-Backward » dans le cadre de la minimisation convexe structurée non lisse (voir chapitre 3, Section 3.1.4).

## 4.2 Propriétés de type Lyapunov

Pour des itérés  $\{z_n, x_n, y_n\} \subset \mathcal{H}$  générés par **ALG1**, on associe la suite  $(\mathcal{E}_n(s, q))$  définie pour  $(s, q) \in ]0, +\infty[ \times \mathcal{H}$  et pour  $n \geq 1$  par

$$(4.2.1) \quad \begin{aligned} \mathcal{E}_n(s, q) = & \frac{1}{2} \|s(q - x_n) + \nu_n(y_n - x_n)\|^2 \\ & + \left(\frac{1}{2}\right) s(e - s) \|x_n - q\|^2 + st(e + \nu_n) \langle M_\mu(z_{n-1}), x_n - q \rangle. \end{aligned}$$

Des propriétés de Lyapunov pour cette suite seront établies plus loin, exposons tout d'abord quelques résultats préliminaires.

### 4.2.1 Préliminaires.

#### 4.2.1.1 Cocoercivité de $M_\mu$ et quelques estimations.

Afin d'établir nos principaux résultats, nous rappelons une certaine propriété de cocoercivité concernant l'opérateur  $M_\mu$  dont nous tirons quelques résultats préliminaires. Les deux remarques suivantes se révéleront utiles pour notre analyse.

**Remarque 4.2.13.** On observe que les itérés  $\{z_n, x_n, y_n\} \subset \mathcal{H}$  générés par(4.1.7)-(4.1.8) vérifient

$$(4.2.2) \quad \forall n \geq 0, \quad x_{n+1} - z_n = -tM_\mu(z_n).$$

**Remarque 4.2.14.** Supposons que l'opérateur  $B$  soit  $\lambda$ -cocoercif (pour une certaine constante  $\lambda$  strictement positive). On peut alors vérifier, sous la condition  $\mu \in ]0, 2\lambda[$  (voir, par exemple, [8, Lemma B1]) que l'opérateur  $M_\mu$  est  $\sigma$ -cocoercif avec

$$(4.2.3) \quad \sigma = \mu \left(1 - \frac{\mu}{4\lambda}\right).$$

Ces deux remarques nous permettent d'énoncer quelques estimations utiles concernant les itérés produits par l'algorithme(4.1.7)-(4.1.8).

**Lemme 4.2.5.** Supposons que les conditions (4.1.2) soient vérifiées. Soient  $\mu \in ]0, 2\lambda[$  et  $t \in ]0, \sigma[$ . Soient  $\{z_n, x_n, y_n\} \subset \mathcal{H}$  les suites générées par(4.1.7)-(4.1.8). Alors, pour  $n \geq 1$ , nous avons les propriétés suivantes :

$$(4.2.4a) \quad \langle M_\mu(z_{n-1}), x_n - q \rangle \geq (\sigma - t)\|M_\mu(z_{n-1})\|^2, \quad \text{pour } q \in S,$$

$$(4.2.4b) \quad \langle M_\mu(z_n) - M_\mu(z_{n-1}), x_{n+1} - x_n \rangle \geq (\sigma - t)\|M_\mu(z_n) - M_\mu(z_{n-1})\|^2.$$

**Démonstration.** La Remarque 4.2.13, nous donne  $x_n = z_{n-1} - tM_\mu(z_{n-1})$  (pour  $n \geq 1$ ). On obtient alors, pour  $q \in S$ ,

$$(4.2.5) \quad \langle M_\mu(z_{n-1}), x_n - q \rangle = \langle M_\mu(z_{n-1}), z_{n-1} - q \rangle - t\|M_\mu(z_{n-1})\|^2.$$

Par conséquent, comme  $M_\mu(q) = 0$  et que  $M_\mu$  est  $\sigma$ -cocoercif, on en déduit immédiatement (4.2.4a). En utilisant à nouveau la Remarque 4.2.13 pour  $n \geq 1$ , on a

$$x_{n+1} - x_n = z_n - z_{n-1} - t(M_\mu(z_n) - M_\mu(z_{n-1})).$$

On en déduit que

$$\begin{aligned} \langle M_\mu(z_n) - M_\mu(z_{n-1}), x_{n+1} - x_n \rangle \\ = \langle M_\mu(z_n) - M_\mu(z_{n-1}), z_n - z_{n-1} \rangle - t\|M_\mu(z_n) - M_\mu(z_{n-1})\|^2. \end{aligned}$$

L'inégalité (4.2.4b) découle alors de l'égalité précédente et de la  $\sigma$ -cocoercivité de  $M_\mu$ . ■

## 4.2.2 Un résultat général pour une analyse de Lyapunov.

**Remarque 4.2.15.** À partir de maintenant, nous adopterons la notation suivante : pour une suite quelconque  $(u_n)$ , on note  $\dot{u}_n = u_n - u_{n-1}$ .

Le lemme suivant constitue un résultat important pour notre étude du comportement asymptotique des suites générées par l'algorithme (4.1.7)-(4.1.8).

Les quantités  $\rho_n$  et  $T_n(s)$  (pour une certaine valeur  $s > 0$ ) définies ci-dessous jouent un rôle crucial dans notre méthodologie :

$$(4.2.6) \quad \rho_n := 1 - (1 - \kappa) \frac{\nu_{n+1}}{\nu_n},$$

$$(4.2.7) \quad T_n(s) := \frac{1}{2}(1 - \kappa)\rho_n\nu_n\nu_{n+1} \left\| y_n - x_n + \left( \frac{1}{\theta_n} - \frac{e-s}{\nu_n\rho_n} \right) \dot{x}_{n+1} \right\|^2 + \frac{1}{2}(e-s) \left( (e + 2\nu_{n+1})\rho_n^{-1} + s(\rho_n^{-1} - 1) \right) \|\dot{x}_{n+1}\|^2.$$

Énonçons à présent le résultat principal de cette section.

**Lemme 4.2.6.** Soient  $\{z_n, x_n, y_n\} \subset \mathcal{H}$  des suites générées par (4.1.7)-(4.1.8) avec des paramètres  $\{e, \kappa, s_0, t, \mu, \nu_n\} \subset ]0, +\infty[$ . Alors nous avons les propriétés suivantes

(1) S'il existe un entier naturel  $n_1$  tel que

$$(4.2.8) \quad (1 - \kappa) \frac{\nu_{n+1}}{\nu_n} < 1 \quad (\text{pour } n \geq n_1),$$

alors pour tout  $(s, q) \in ]0, +\infty[ \times \mathcal{H}$ , on a (pour  $n \geq n_1$ )

$$(4.2.9) \quad \begin{aligned} & \dot{\mathcal{E}}_{n+1}(s, q) + s(s_0\rho_n - \dot{\nu}_{n+1}) \langle tM_\mu(z_{n-1}), x_n - q \rangle \\ & + \frac{1}{2}\rho_n^{-1}(e + \nu_{n+1})^2 \|\dot{x}_{n+1} + \theta_n(y_n - x_n)\|^2 \\ & + \left( 1 - s \frac{\rho_n}{e + \nu_{n+1}} \right) \rho_n^{-1}(e + \nu_{n+1})^2 \langle tM_\mu(z_n) - tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle \\ & + (s_0 - s)(e + \nu_{n+1}) \langle tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle + T_n(s) = 0, \end{aligned}$$

où  $\mathcal{E}_n(s, q)$ ,  $\rho_n$  et  $T_n(s)$  sont respectivement donnés par (4.2.1), (4.2.6) et (4.2.7).

(2) Si les paramètres  $\{e, \kappa, s_0, t, \mu, \nu_n\} \subset ]0, +\infty[$  vérifient

$$(4.2.10) \quad \kappa \in ]0, 2[, \quad s_0 \in ]0, e[ \quad \text{et} \quad \nu_{n+1} \sim \nu_n \quad (\text{lorsque } n \rightarrow +\infty),$$

et si pour  $\kappa \neq 1$ , il existe  $\delta \in ]0, \frac{1}{2}[$  et un entier naturel  $n_1$  tels que pour  $n \geq n_1$ ,

$$(4.2.11) \quad \nu_n \geq (e - s_0) \frac{\delta + (1-\delta)|1-\kappa|}{\delta(1-|1-\kappa|)} - s_0(1 - \kappa),$$

alors on a (pour  $n \geq n_1$ )

$$(4.2.12) \quad \begin{aligned} & \dot{\mathcal{E}}_{n+1}(s_0, q) + s_0 (s_0 \rho_n - \dot{\nu}_{n+1}) \langle tM_\mu(z_{n-1}), x_n - q \rangle \\ & + (\gamma_n \rho_n^{-1})(e + \nu_{n+1})^2 \langle tM_\mu(z_n) - tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle \\ & + c \rho_n^{-1} (e + \nu_{n+1})^2 \|\dot{x}_{n+1} + \theta_n(y_n - x_n)\|^2 \\ & + \frac{1}{2}(e - s_0) \rho_n^{-1} (e + \nu_{n+1}) \|\dot{x}_{n+1}\|^2 \leq 0. \end{aligned}$$

où  $c$  est une constante strictement positive définie par

$$(4.2.13) \quad c := \begin{cases} \frac{1}{2}(2 - \delta)(1 - |1 - \kappa|) & \text{si } \kappa \neq 1, \\ \frac{1}{2} & \text{si } \kappa = 1. \end{cases}$$

#### 4.2.2.1 Démonstration du Lemme 4.2.6

(a)- **Partie I : estimations à partir d'un modèle plus général.** Afin de démontrer le Lemme 4.2.6, nous rappelons une égalité importante concernant le cadre plus général des suites  $\{x_n, y_n, d_n\} \subset \mathcal{H}$  vérifiant

$$(4.2.14a) \quad (e + \nu_{n+1})\theta_n = \nu_n - (1 - \kappa)\nu_{n+1},$$

$$(4.2.14b) \quad \dot{x}_{n+1} + d_n + \theta_n(y_n - x_n) = 0,$$

$$(4.2.14c) \quad y_{n+1} - y_n = -\kappa(y_n - x_n),$$

où  $\{\kappa, \theta_n, e, \nu_n\}$  sont des paramètres strictement positifs. À cette fin, nous associons à (4.2.14) la quantité  $G_n(s, q)$  définie pour  $(s, q) \in [0, +\infty[ \times \mathcal{H}$  par

$$(4.2.15) \quad G_n(s, q) = \frac{1}{2} \|s(q - x_n) + \nu_n(y_n - x_n)\|^2 + \frac{1}{2} s(e - s) \|x_n - q\|^2.$$

La proposition suivante a été établie par la Proposition 3.3.1 du chapitre 3.

**Proposition 4.2.1.** Soient  $\{x_n, y_n, d_n\} \subset \mathcal{H}$  et  $\{\theta_n, \nu_n\} \subset ]0, +\infty[$  vérifiant (4.2.14) pour  $n \geq 0$ . Soient  $\{\kappa, e\} \subset ]0, +\infty[$ , et supposons que la condition (4.2.8) soit vérifiée. Alors pour  $(s, q) \in ]0, +\infty[ \times \mathcal{H}$  et pour  $n \geq 0$  nous avons

$$(4.2.16) \quad \begin{aligned} & \dot{G}_{n+1}(s, q) + \frac{1}{2}\rho_n^{-1}(e + \nu_{n+1})^2 \|\dot{x}_{n+1} + \theta_n(y_n - x_n)\|^2 \\ & + s(e + \nu_{n+1}) \langle d_n, x_{n+1} - q \rangle \\ & + \left(1 - s \frac{\rho_n}{e + \nu_{n+1}}\right) \rho_n^{-1}(e + \nu_{n+1})^2 \langle d_n, \dot{x}_{n+1} \rangle = -T_n(s), \end{aligned}$$

où  $\rho_n$  et  $T_n(s)$  sont définis respectivement par (4.2.6) et (4.2.7).

**(b)- Partie II :** La remarque suivante met en lumière le lien entre le schéma(4.1.7)-(4.1.8) et le modèle général (4.2.14)

**Remarque 4.2.16.** Pour  $n \geq 1$ , on a  $x_n - z_{n-1} = -tM_\mu(z_{n-1})$  d'après (4.2.2). Par conséquent, l'égalité (4.1.7a) s'écrit

$$(4.2.17) \quad z_n = x_n - \theta_n(y_n - x_n) + t\gamma_n M_\mu(z_{n-1}),$$

De plus, en utilisant à nouveau la relation (4.2.2), on a  $z_n = x_{n+1} + tM_\mu(z_n)$ . En combinant ces deux résultats, on obtient

$$(4.2.18) \quad \dot{x}_{n+1} + \theta_n(y_n - x_n) + t(M_\mu(z_n) - \gamma_n M_\mu(z_{n-1})) = 0.$$

On observe alors que l'algorithme (4.1.7)-(4.1.8) entre dans le modèle général (4.2.14) en prenant  $d_n = t(M_\mu(z_n) - \gamma_n M_\mu(z_{n-1}))$ .

Le lemme suivant est une conséquence directe de la Proposition (4.2.1) pour une suite  $\gamma_n$  positive quelconque :

**Lemme 4.2.7.** Soient  $\{z_n, x_n, y_n\} \subset \mathcal{H}$  des suites générées par (4.1.7)-(4.1.8) avec  $\gamma_n \geq 0$  et des paramètres  $\{e, \kappa, t, \mu, \nu_n\} \subset ]0, +\infty[$  vérifiant (4.2.8). Alors, pour tout  $(s, q) \in ]0, +\infty[ \times \mathcal{H}$ , on a (pour  $n \geq 1$ )

$$(4.2.19) \quad \begin{aligned} & \dot{\mathcal{E}}_{n+1}(s, q) + s \left( (e + \nu_n) - \gamma_n(e + \nu_{n+1}) \right) \langle tM_\mu(z_{n-1}), x_n - q \rangle \\ & + \frac{1}{2}\rho_n^{-1}(e + \nu_{n+1})^2 \|\dot{x}_{n+1} + \theta_n(y_n - x_n)\|^2 \\ & + \rho_n^{-1}(e + \nu_{n+1})^2 \left( (\vartheta_n - \gamma_n)Q_n + \vartheta_n Z_n \right) = -T_n(s), \end{aligned}$$

où  $\mathcal{E}_n(s, q)$ ,  $\rho_n$  et  $T_n(s)$  sont donnés respectivement par (4.2.1) (4.2.6) et (4.2.7),  $\vartheta_n := 1 - s \frac{\rho_n}{e + \nu_{n+1}}$  et où

$$Q_n = \langle tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle, \quad Z_n = \langle tM_\mu(z_n) - tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle.$$

La démonstration du Lemme 4.2.7 est donnée en annexe (section 4.6.2).

**(c)- Partie III :** Démontrons (4.2.9). Les définitions de  $\gamma_n := 1 - s\rho_n(e + \nu_{n+1})^{-1}$  et  $\vartheta_n := 1 - s_0\rho_n(e + \nu_{n+1})^{-1}$  donnent (pour  $n \geq 1$ )

$$\vartheta_n - \gamma_n = s\rho_n(e + \nu_{n+1})^{-1} - s_0\rho_n(e + \nu_{n+1})^{-1} = (s - s_0)\rho_n(e + \nu_{n+1})^{-1},$$

$$(e + \nu_n) - (e + \nu_{n+1})\gamma_n = (e + \nu_n) - (e + \nu_{n+1}) + s_0\rho_n = s\rho_n - \dot{\nu}_{n+1}.$$

Par conséquent, à l'aide de ces deux résultats, l'équation (4.2.19) nous donne

$$(4.2.20) \quad \begin{aligned} & \dot{\mathcal{E}}_{n+1}(s, q) + s(s_0\rho_n - \dot{\nu}_{n+1}) \langle tM_\mu(z_{n-1}), x_n - q \rangle \\ & + \frac{1}{2}\rho_n^{-1}(e + \nu_{n+1})^2 \|\dot{x}_{n+1} + \theta_n(y_n - x_n)\|^2 \\ & + (s_0 - s)(e + \nu_{n+1})Q_n \\ & + \vartheta_n\rho_n^{-1}(e + \nu_{n+1})^2 Z_n = -T_n(s), \end{aligned}$$

où  $Z_n = \langle tM_\mu(z_n) - tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle$ . Ce qui nous donne (4.2.9).

Démontrons (4.2.12). En prenant  $s = s_0$  (et donc  $\vartheta_n = \gamma_n$ ) dans (4.2.20), on obtient

$$(4.2.21) \quad \begin{aligned} & \dot{\mathcal{E}}_{n+1}(s_0, q) + s_0(s_0\rho_n - \dot{\nu}_{n+1}) \langle tM_\mu(z_{n-1}), x_n - q \rangle \\ & + \frac{1}{2}\rho_n^{-1}(e + \nu_{n+1})^2 \|\dot{x}_{n+1} + \theta_n(y_n - x_n)\|^2 \\ & + \gamma_n\rho_n^{-1}(e + \nu_{n+1})^2 \langle tM_\mu(z_n) - tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle = -T_n(s_0). \end{aligned}$$

Estimons, à présent, le terme du côté droit de l'égalité précédente. Par définition de  $T_n$  (pour  $n \geq 1$ ) :

$$(4.2.22) \quad -T_n(s_0) \leq \frac{1}{2} \left| 1 - \kappa|\rho_n\nu_n\nu_{n+1}| \left\| y_n - x_n + \left( \frac{1}{\theta_n} - \frac{e-s_0}{\nu_n\rho_n} \right) \dot{x}_{n+1} \right\|^2 \right. \\ \left. - \frac{1}{2}(e - s_0) \left( (e + 2\nu_{n+1})\rho_n^{-1} + s_0(\rho_n^{-1} - 1) \right) \|\dot{x}_{n+1}\|^2 \right.$$

Par ailleurs, pour  $\varepsilon > 0$  donné, l'inégalité de Young nous donne

$$\begin{aligned}
 & \left\| y_n - x_n + \left( \frac{1}{\theta_n} - \frac{e-s_0}{\nu_n \rho_n} \right) \dot{x}_{n+1} \right\|^2 \\
 (4.2.23) \quad &= \left\| \frac{1}{\theta_n} (\theta_n (y_n - x_n) + \dot{x}_{n+1}) - \frac{e-s_0}{\nu_n \rho_n} \dot{x}_{n+1} \right\|^2 \\
 &\leq (1 + 2\varepsilon) \left\| \frac{1}{\theta_n} (\theta_n (y_n - x_n) + \dot{x}_{n+1}) \right\|^2 + \left(1 + \frac{1}{2\varepsilon}\right) \left\| \frac{e-s_0}{\nu_n \rho_n} \dot{x}_{n+1} \right\|^2 \\
 &= (1 + 2\varepsilon) \frac{1}{\theta_n^2} \|\dot{x}_{n+1} + \theta_n (y_n - x_n)\|^2 + \left(1 + \frac{1}{2\varepsilon}\right) \left(\frac{e-s_0}{\nu_n \rho_n}\right)^2 \|\dot{x}_{n+1}\|^2.
 \end{aligned}$$

Ainsi, en remarquant que l'on a  $\theta_n = \frac{\nu_n \rho_n}{\tau_n}$ , on obtient

$$\begin{aligned}
 & \frac{1}{2} \rho_n \nu_n \nu_{n+1} \left\| y_n - x_n + \left( \frac{1}{\theta_n} - \frac{e-s_0}{\nu_n \rho_n} \right) \dot{x}_{n+1} \right\|^2 \\
 (4.2.24) \quad &\leq \left(\frac{1}{2} + \varepsilon\right) \rho_n \nu_n \nu_{n+1} \frac{1}{\theta_n^2} \|\dot{x}_{n+1} + \theta_n (y_n - x_n)\|^2 \\
 &\quad + \rho_n \nu_n \nu_{n+1} \left(\frac{1}{2} + \frac{1}{4\varepsilon}\right) \left(\frac{e-s_0}{\nu_n \rho_n}\right)^2 \|\dot{x}_{n+1}\|^2 \\
 &= \left(\frac{1}{2} + \varepsilon\right) \rho_n^{-1} \frac{\nu_{n+1}}{\nu_n} (e + \nu_{n+1})^2 \|\dot{x}_{n+1} + \theta_n (y_n - x_n)\|^2 \\
 &\quad + \left(\frac{1}{2} + \frac{1}{4\varepsilon}\right) (e - s_0)^2 \rho_n^{-1} \frac{\nu_{n+1}}{\nu_n} \|\dot{x}_{n+1}\|^2.
 \end{aligned}$$

Par conséquent, en combinant ce dernier résultat avec (4.2.22), on obtient

$$\begin{aligned}
 -T_n(s_0) &\leq \left(\frac{1}{2} + \varepsilon\right) |1 - \kappa| \rho_n^{-1} \frac{\nu_{n+1}}{\nu_n} (e + \nu_{n+1})^2 \|\dot{x}_{n+1} + \theta_n (y_n - x_n)\|^2 \\
 &\quad + \left(\frac{1}{2} + \frac{1}{4\varepsilon}\right) |1 - \kappa| (e - s_0)^2 \rho_n^{-1} \frac{\nu_{n+1}}{\nu_n} \|\dot{x}_{n+1}\|^2 \\
 &\quad - \frac{1}{2} (e - s_0) \rho_n^{-1} (e + 2\nu_{n+1} + s_0(1 - \rho_n)) \|\dot{x}_{n+1}\|^2.
 \end{aligned}$$

En outre, la définition de  $\rho_n$  nous donne  $1 - \rho_n = (1 - \kappa) \nu_{n+1} \nu_n^{-1}$ . Ainsi l'inégalité précédente se réduit à

$$\begin{aligned}
 & -T_n(s_0) \\
 &\leq \left(\frac{1}{2} + \varepsilon\right) |1 - \kappa| \rho_n^{-1} \frac{\nu_{n+1}}{\nu_n} (e + \nu_{n+1})^2 \|\dot{x}_{n+1} + \theta_n (y_n - x_n)\|^2 \\
 &+ \frac{1}{2} (e - s_0) \rho_n^{-1} \left( e + 2\nu_{n+1} + s_0(1 - \kappa) \frac{\nu_{n+1}}{\nu_n} - 2 \left(\frac{1}{2} + \frac{1}{4\varepsilon}\right) |1 - \kappa| (e - s_0) \frac{\nu_{n+1}}{\nu_n} \right) \|\dot{x}_{n+1}\|^2 \\
 &= -\frac{1}{2} (e - s_0) \rho_n^{-1} \left( e + 2\nu_{n+1} + (s_0(1 - \kappa) - (e - s_0) \left(1 + \frac{1}{2\varepsilon}\right) |1 - \kappa|) \frac{\nu_{n+1}}{\nu_n} \right) \|\dot{x}_{n+1}\|^2,
 \end{aligned}$$

ou, de façon équivalente

$$(4.2.25) \quad -T_n(s_0) \leq \frac{1}{2}C_1(\varepsilon, \kappa)\rho_n^{-1}\frac{\nu_{n+1}}{\nu_n}(e + \nu_{n+1})^2\|\dot{x}_{n+1} + \theta_n(y_n - x_n)\|^2 - \frac{1}{2}(e - s_0)\rho_n^{-1}\left(e + 2\nu_{n+1} - C_2(\varepsilon, \kappa)\frac{\nu_{n+1}}{\nu_n}\right)\|\dot{x}_{n+1}\|^2,$$

où les quantités  $C_1(\varepsilon, \kappa)$  et  $C_2(\varepsilon, \kappa)$  sont définies par

$$(4.2.26a) \quad C_1(\varepsilon, \kappa) = (1 + 2\varepsilon)|1 - \kappa|,$$

$$(4.2.26b) \quad C_2(\varepsilon, \kappa) = (e - s_0)\left(1 + \frac{1}{2\varepsilon}\right)|1 - \kappa| - s_0(1 - \kappa).$$

Observons que pour  $\kappa \in ]0, 2[$ , on a clairement

$$(4.2.27) \quad |1 - \kappa| < 1.$$

Notons par ailleurs que pour  $\kappa = 1$ , on a  $C_1(\varepsilon, \kappa) = C_2(\varepsilon, \kappa) = 0$ . Pour  $\kappa \neq 1$  et en considérant  $\delta \in ]0, \frac{1}{2}[$  tel que

$$(4.2.28) \quad \varepsilon_0 = \frac{1}{2}\delta\frac{1 - |1 - \kappa|}{|1 - \kappa|},$$

on obtient

$$(4.2.29a) \quad C_1(\varepsilon_0, \kappa) = |1 - \kappa| + \delta(1 - |1 - \kappa|),$$

$$(4.2.29b) \quad C_2(\varepsilon_0, \kappa) = (e - s_0)\frac{\delta + (1 - \delta)|1 - \kappa|}{\delta(1 - |1 - \kappa|) - s_0(1 - \kappa)} - s_0(1 - \kappa).$$

L'inégalité (4.2.25) nous donne alors pour  $\kappa \in ]0, 2[$  et pour  $n \geq n_1$ ,

$$(4.2.30) \quad -T_n(s_0) \leq \frac{1}{2}\bar{C}_1\rho_n^{-1}\frac{\nu_{n+1}}{\nu_n}(e + \nu_{n+1})^2\|\dot{x}_{n+1} + \theta_n(y_n - x_n)\|^2 - \frac{1}{2}(e - s_0)\rho_n^{-1}\left(e + 2\nu_{n+1} - \bar{C}_2\frac{\nu_{n+1}}{\nu_n}\right)\|\dot{x}_{n+1}\|^2,$$

où  $\bar{C}_1$  et  $\bar{C}_2$  sont donnés par

$$(4.2.31a) \quad \bar{C}_1 = C_1(\varepsilon_0, \kappa) \text{ si } \kappa \neq 1 \text{ et } \bar{C}_1 = 0 \text{ si } \kappa = 1,$$

$$(4.2.31b) \quad \bar{C}_2 = C_2(\varepsilon_0, \kappa) \text{ si } \kappa \neq 1 \text{ et } \bar{C}_2 = 0 \text{ si } \kappa = 1.$$

Par conséquent, en combinant (4.2.21) et (4.2.30) et en rappelant que  $s_0 \in ]0, e[$ , on en déduit que (pour  $n \geq n_1$ )

$$(4.2.32) \quad \begin{aligned} & \dot{\mathcal{E}}_{n+1}(s_0, q) + s_0 (s_0 \rho_n - \dot{\nu}_{n+1}) \langle tM_\mu(z_{n-1}), x_n - q \rangle \\ & + \frac{1}{2} \left( 1 - \bar{C}_1 \frac{\nu_{n+1}}{\nu_n} \right) \rho_n^{-1} (e + \nu_{n+1})^2 \|\dot{x}_{n+1} + \theta_n(y_n - x_n)\|^2 \\ & + \gamma_n \rho_n^{-1} (e + \nu_{n+1})^2 \langle tM_\mu(z_n) - tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle \\ & + \frac{1}{2} (e - s_0) \rho_n^{-1} \left( e + 2\nu_{n+1} - \bar{C}_2 \frac{\nu_{n+1}}{\nu_n} \right) \|\dot{x}_{n+1}\|^2 \leq 0. \end{aligned}$$

Par ailleurs, on vérifie aisément les équivalences suivantes

$$(4.2.33) \quad \begin{aligned} & 1 + \frac{\delta(1-|1-\kappa|)}{\delta+(1-\delta)|1-\kappa|} > \frac{\nu_{n+1}}{\nu_n} \\ & \Leftrightarrow 2\delta + (1-2\delta)|1-\kappa| > (\delta + (1-\delta)|1-\kappa|) \frac{\nu_{n+1}}{\nu_n} = C_1(\varepsilon_0, \kappa) \frac{\nu_{n+1}}{\nu_n} \\ & \Leftrightarrow 1 - C_1(\varepsilon_0, \kappa) \frac{\nu_{n+1}}{\nu_n} > 1 - 2\delta - (1-2\delta)|1-\kappa| = (1-2\delta)(1-|1-\kappa|). \end{aligned}$$

De plus, la condition (4.2.11) peut se réécrire pour  $n \geq n_1$  ( $n_1$  étant un entier naturel suffisamment grand) sous la forme

$$(4.2.34) \quad \nu_{n+1} - C_2(\varepsilon_0, \kappa) \frac{\nu_{n+1}}{\nu_n} \geq 0.$$

Observons, en outre, que la condition  $\nu_{n+1} \sim \nu_n$  (lorsque  $n \rightarrow +\infty$ ) nous donne (pour  $\kappa \in ]0, 2[$ )

$$(4.2.35) \quad \lim_{n \rightarrow +\infty} \frac{\nu_{n+1}}{\nu_n} < 1 + \frac{\delta(1-|1-\kappa|)}{\delta+(1-\delta)|1-\kappa|} \quad (\text{pour tout } \delta \in ]0, \frac{1}{2}[).$$

Ainsi, pour  $n \geq n_1$  ( $n_1$  étant un entier naturel suffisamment grand), on obtient

$$(4.2.36) \quad \frac{\nu_{n+1}}{\nu_n} < 1 + \frac{\delta(1-|1-\kappa|)}{\delta+(1-\delta)|1-\kappa|},$$

ce qui d'après (4.2.33) entraîne

$$(4.2.37) \quad 1 - C_1(\varepsilon_0, \kappa) \frac{\nu_{n+1}}{\nu_n} > (1-2\delta)(1-|1-\kappa|).$$

Par conséquent, en combinant les arguments précédents avec (4.2.32) et en considérant la quantité  $c$  définie par (4.2.13), on obtient pour  $\kappa \in ]0, 2[$  et pour  $n \geq n_1$

$$\begin{aligned}
 (4.2.38) \quad & \dot{\mathcal{E}}_{n+1}(s_0, q) + s_0 (s_0 \rho_n - \dot{\nu}_{n+1}) \langle tM_\mu(z_{n-1}), x_n - q \rangle \\
 & + (\gamma_n \rho_n^{-1})(e + \nu_{n+1})^2 \langle tM_\mu(z_n) - tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle \\
 & + c \rho_n^{-1} (e + \nu_{n+1})_n^2 \|\dot{x}_{n+1} + \theta_n(y_n - x_n)\|^2 \\
 & + \frac{1}{2} (e - s_0) \rho_n^{-1} (e + \nu_{n+1}) \|\dot{x}_{n+1}\|^2 \leq 0,
 \end{aligned}$$

ce qui n'est autre que (4.2.12). ■

### 4.3 Analyse de la convergence de la méthode.

Dans cette section, nous démontrons la convergence faible des itérés générés par [ALG1](#) et nous établissons des conditions suffisantes sur les paramètres pour garantir la convergence de la méthode.

#### 4.3.1 Quelques estimations préliminaires

À partir de maintenant, nous ferons les hypothèses suivantes concernant les paramètres  $\{\kappa, t, \mu, s_0, \nu_n\}$  de l'algorithme [ALG1](#) :

$$(4.3.1a) \quad \mu \in ]0, 2\lambda[ \text{ et } 0 < t < \mu(1 - \frac{\mu}{4\lambda}) \text{ où } \lambda \text{ est donné par (4.1.2),}$$

$$(4.3.1c) \quad \lim_{n \rightarrow +\infty} \frac{\nu_{n+1}}{\nu_n} = 1,$$

$$(4.3.1d) \quad (\nu_n) \text{ est croissante,}$$

$$(4.3.1e) \quad \limsup_{n \rightarrow +\infty} \dot{\nu}_{n+1} < \frac{1}{2} s_0 \kappa,$$

$$(4.3.1b) \quad \text{Si } \kappa \neq 1, \quad \sup_{n \geq 0} \nu_n \geq (e - s_0) \frac{\delta + (1-\delta)|1-\kappa|}{\delta(1-|1-\kappa|)} - s_0(1 - \kappa) \text{ où } \delta \in ]0, \frac{1}{2}[.$$

Quelques observations sont données par la remarque suivante concernant le comportement asymptotique des paramètres  $\{\rho_n, \theta_n\}$ .

**Remarque 4.3.17.** *À l'aide des définitions de  $\rho_n = 1 - (1 - \kappa)\nu_{n+1}\nu_n^{-1}$  et de  $\theta_n = \nu_n \rho_n (e + \nu_{n+1})^{-1}$ , on observe que*

$$(4.3.2) \quad \rho_n \sim \kappa \text{ et } \theta_n \sim \kappa \text{ (lorsque } n \rightarrow +\infty).$$

De plus que la suite  $(\rho_n)_n$  peut se réécrire sous la forme

$$(4.3.3) \quad \rho_n = \kappa - (1 - \kappa) \frac{\dot{\nu}_{n+1}}{\nu_n}.$$

Une première série d’estimations est donnée par le lemme suivant dans le but de démontrer la convergence de l’algorithme [ALG1](#).

**Lemme 4.3.8.** *Soient  $\{z_n, x_n, y_n\} \subset \mathcal{H}$  les suites générées par [ALG1](#). Alors, sous les conditions [\(4.1.2\)](#), [\(4.3.1\)](#),  $\kappa \in ]0, 2[$  et  $s_0 \in ]0, e[$ , nous avons :*

$$\begin{aligned}
 (4.3.4a) \quad & (\mathcal{E}_n(s_0, q))_n \text{ est convergent,} \\
 (4.3.4b) \quad & (x_n) \text{ est borné,} \\
 (4.3.4c) \quad & \sum_n \langle M_\mu(z_{n-1}), x_n - q \rangle < +\infty, \\
 (4.3.4d) \quad & \sum_n \nu_n^2 \|M_\mu(z_n) - M_\mu(z_{n-1})\|^2 < +\infty, \\
 (4.3.4e) \quad & \sum_n \nu_n \|\dot{x}_{n+1}\|^2 < +\infty, \\
 (4.3.4f) \quad & \sum_n \nu_n^2 \|\dot{x}_{n+1} + \theta_n(y_n - x_n)\|^2 < +\infty, \\
 (4.3.4g) \quad & \sum_n \nu_n \|y_n - x_n\|^2 < +\infty.
 \end{aligned}$$

**Démonstration.** Soient  $q \in S$  et  $\kappa \in ]0, 2[$ . En utilisant l’inégalité [\(4.2.12\)](#) du Lemme [4.2.6](#) pour  $n \geq n_1$  (pour un certain entier  $n_1$  suffisamment grand), on obtient (pour  $q \in S$ )

$$(4.3.5) \quad \dot{\mathcal{E}}_{n+1}(s_0, q) + s_0(s_0\rho_n - \dot{\nu}_{n+1})U_n + (\gamma_n\rho_n^{-1})(e + \nu_{n+1})^2 Z_n + c\rho_n^{-1}(e + \nu_{n+1})^2 W_n + \frac{1}{2}(e - s_0)\rho_n^{-1}(e + \nu_{n+1})\|\dot{x}_{n+1}\|^2 \leq 0,$$

où  $U_n = \langle tM_\mu(z_{n-1}), x_n - q \rangle$ ,  $Z_n = \langle tM_\mu(z_n) - tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle$ ,  $c$  est une constante strictement positive et où  $W_n = \|\dot{x}_{n+1} + \theta_n(y_n - x_n)\|^2$ . Observons que le Lemme [4.2.5](#) garantit, sous les conditions [\(4.3.1\)](#), que les quantités  $U_n$  et  $Z_n$  sont positives. Par ailleurs, concernant le second terme du côté gauche de [\(4.3.5\)](#), on obtient d’après [\(4.3.1e\)](#) et pour un entier  $n_1$  suffisamment grand

$$(4.3.6) \quad \inf_{n \geq n_1} (s_0\rho_n - \dot{\nu}_{n+1}) > 0.$$

Par conséquent, il résulte de [\(4.3.5\)](#) et des arguments précédents que la suite positive  $(\mathcal{E}_{n+1}(s_0, q))_{n \geq n_1}$  est décroissante, et donc qu’elle converge. Ceci nous donne [\(4.3.4a\)](#). En outre,  $\mathcal{E}_n(s_0, q)$  s’écrit

$$\mathcal{E}_n(s_0, q) = \frac{1}{2} \|s_0(q - x_n) + \nu_n(y_n - x_n)\|^2 + \frac{1}{2} s_0(e - s_0) \|x_n - q\|^2 + s_0(e + \nu_n)U_n.$$

On en déduit donc que la suite  $(x_n)$  est bornée. Cela nous conduit à (4.3.4b). Par ailleurs, en sommant l'inégalité (4.3.5) pour  $n$  allant de  $n_1$  à  $N$  (pour un certain  $N \geq n_1$ ), on a

$$(4.3.7) \quad \begin{aligned} & \mathcal{E}_{N+1}(s_0, q) + s_0 \sum_{n=n_1}^N (s_0 \rho_n - \dot{\nu}_{n+1}) U_n \\ & + \sum_{n=n_1}^N (\gamma_n \rho_n^{-1}) (e + \nu_{n+1})^2 Z_n + c \sum_{n=n_1}^N \rho_n^{-1} (e + \nu_{n+1})^2 W_n \\ & + \frac{1}{2} (e - s_0) \sum_{n=n_1}^N \rho_n^{-1} (e + \nu_{n+1}) \|\dot{x}_{n+1}\|^2 \leq \mathcal{E}_{n_1}(s_0, q). \end{aligned}$$

On obtient alors les estimations (4.3.4c)-(4.3.4f) en utilisant le Lemme 4.2.5 et en remarquant que  $e + \nu_{n+1} \sim \nu_n$  et que  $(\gamma_n \rho_n^{-1}) \sim \kappa^{-1}$  (lorsque  $n \rightarrow +\infty$ ). Par ailleurs, en notant  $u_n = y_n - x_n$ , on a classiquement

$$(4.3.8) \quad \|\theta_n u_n\|^2 \leq 2\|\dot{x}_{n+1}\|^2 + 2\|\dot{x}_{n+1} + \theta_n u_n\|^2.$$

Ainsi, sachant que  $\sum_n \nu_n \|\dot{x}_{n+1}\|^2 < +\infty$  et  $\sum_n \nu_n \|\dot{x}_{n+1} + \theta_n u_n\|^2 < +\infty$  (d'après (4.3.4e)) et (4.3.4f) respectivement), on en déduit que  $\sum_n \theta_n^2 \nu_n \|u_n\|^2 < +\infty$ . Il en résulte alors que  $\sum_n \nu_n \|u_n\|^2 < +\infty$  (car  $\theta_n \sim \kappa$ , lorsque  $n \rightarrow +\infty$ ), ce qui nous donne (4.3.4g). ■

**Remarque 4.3.18.** *En notant  $u_n := y_n - x_n$  pour  $n \geq 0$ , on obtient  $x_{n+1} = -u_{n+1} + y_{n+1}$ . Ainsi comme  $y_{n+1} = y_n - \kappa u_n$  (d'après (4.1.7)), on obtient*

$$(4.3.9) \quad x_{n+1} = -u_{n+1} + y_{n+1} = -u_{n+1} + y_n - \kappa u_n.$$

*En utilisant à nouveau la relation  $u_n := y_n - x_n$ , ce dernier résultat peut se réécrire sous la forme*

$$(4.3.10) \quad x_{n+1} - x_n = -u_{n+1} + u_n - \kappa u_n = -u_{n+1} + (1 - \kappa)u_n.$$

À l'aide de la remarque précédente, nous établissons une deuxième série d'estimations via le lemme suivant.

**Lemme 4.3.9.** *Sous les conditions du Lemme 4.3.8, nous avons :*

$$(4.3.11a) \quad \sum_n \nu_n \|\dot{x}_n + tM_\mu(z_{n-1}) + (\kappa - 1)(y_{n-1} - x_{n-1})\|^2 < +\infty,$$

$$(4.3.11b) \quad \|\dot{x}_n + tM_\mu(z_{n-1}) + (\kappa - 1)(y_{n-1} - x_{n-1})\| = o(\nu_n^{-1}),$$

$$(4.3.11c) \quad \sum_n \nu_n \|M_\mu(z_{n-1})\|^2 < +\infty,$$

$$(4.3.11d) \quad \sum_n \nu_n |\langle M_\mu(z_{n-1}), \dot{x}_{n+1} \rangle| < +\infty,$$

$$(4.3.11e) \quad \|\dot{x}_n\| = o(\nu_n^{-1}),$$

$$(4.3.11f) \quad \|y_n - x_n\| = o(\nu_n^{-1}),$$

$$(4.3.11g) \quad \|M_\mu(z_n)\| = o(\nu_n^{-1}),$$

$$(4.3.11h) \quad \sum_n \nu_n \|M_\mu(y_n)\|^2 < +\infty,$$

$$(4.3.11i) \quad \|M_\mu(y_n)\| = o(\nu_n^{-1}).$$

**Démonstration.** Pour démontrer les éléments (4.3.11a) et (4.3.11b), nous allons procéder à une reformulation de la méthode(4.1.7)-(4.1.8). Afin d'alléger les notations, nous introduisons la suite  $(u_n)$  définie pour  $n \geq 0$  par  $u_n := y_n - x_n$ . Par ailleurs, d'après (4.2.18), nous avons aussi (pour  $n \geq 1$ ),

$$\dot{x}_{n+1} + tM_\mu(z_n) - \gamma_n(\dot{x}_n + tM_\mu(z_{n-1})) = -\gamma_n \dot{x}_n - \theta_n u_n.$$

Ainsi, comme  $\dot{x}_n = -u_n - (\kappa - 1)u_{n-1}$  (d'après (4.3.10)), l'équation précédente nous donne

$$(4.3.12) \quad \begin{aligned} & \dot{x}_{n+1} + tM_\mu(z_n) - \gamma_n(\dot{x}_n + tM_\mu(z_{n-1})) \\ &= (\gamma_n - \theta_n)u_n + \gamma_n(\kappa - 1)u_{n-1} \\ &= (\gamma_n - \theta_n + \kappa - 1)u_n - (\kappa - 1)u_n + \gamma_n(\kappa - 1)u_{n-1}, \end{aligned}$$

où les paramètres  $\gamma_n$  et  $\theta_n$  sont définis par

$$(4.3.13) \quad \gamma_n = 1 - s_0 \rho_n (e + \nu_{n+1})^{-1} \quad \text{et} \quad \theta_n = \nu_n \rho_n (e + \nu_{n+1})^{-1}.$$

La relation (4.3.12) nous amène à considérer la suite  $H_n$  définie pour  $n \geq 1$  par

$$(4.3.14) \quad H_n = \dot{x}_n + tM_\mu(z_{n-1}) + (\kappa - 1)u_{n-1},$$

on obtient alors immédiatement, pour  $n \geq 1$

$$(4.3.15) \quad H_{n+1} = \gamma_n H_n + (\gamma_n - \theta_n + \kappa - 1)u_n.$$

Établissons à présent des estimations sur la suite  $H_n$  à l'aide de cette dernière égalité. En utilisant les paramètres de (4.3.13) ainsi que la relation  $\rho_n = \kappa - (1 - \kappa)\dot{\nu}_{n+1}\nu_n^{-1}$  (d'après (4.3.3)), on obtient

$$\begin{aligned} & \gamma_n - \theta_n + \kappa - 1 \\ &= -s_0\rho_n(e + \nu_{n+1})^{-1} - \nu_n(e + \nu_{n+1})^{-1} \left( \kappa - (1 - \kappa)\frac{\dot{\nu}_{n+1}}{\nu_n} \right) + \kappa \\ &= -s_0\rho_n(e + \nu_{n+1})^{-1} + (e + \nu_{n+1})^{-1} (-\kappa\nu_n + (1 - \kappa)\dot{\nu}_{n+1} + \kappa e + \kappa\nu_{n+1}) \\ &= -s_0\rho_n(e + \nu_{n+1})^{-1} + (e + \nu_{n+1})^{-1} (\dot{\nu}_{n+1} + \kappa e) \\ &= (e + \nu_{n+1})^{-1} (-s_0\rho_n + \dot{\nu}_{n+1} + \kappa e). \end{aligned}$$

Par conséquent, on déduit de (4.3.15) et de l'égalité précédente que

$$(4.3.16) \quad H_{n+1} = \gamma_n H_n + (1 - \gamma_n) \frac{(e + \nu_{n+1})^{-1} (\dot{\nu}_{n+1} + \kappa e - s_0\rho_n)}{1 - \gamma_n} u_n.$$

De plus, la convexité de la norme au carré nous donne

$$(4.3.17) \quad \begin{aligned} \|H_{n+1}\|^2 &\leq \gamma_n \|H_n\|^2 + (1 - \gamma_n) \left( \frac{(e + \nu_{n+1})^{-1} (\dot{\nu}_{n+1} + \kappa e - s_0\rho_n)}{1 - \gamma_n} \right)^2 \|u_n\|^2 \\ &= \gamma_n \|H_n\|^2 + \frac{(s_0\rho_n - \dot{\nu}_{n+1} - \kappa e)^2}{1 - \gamma_n} (e + \nu_{n+1})^{-2} \|u_n\|^2. \end{aligned}$$

Ainsi, sachant que  $\gamma_n = 1 - s_0\rho_n(e + \nu_{n+1})^{-1}$ , il en résulte que

$$(4.3.18) \quad \begin{aligned} \|H_{n+1}\|^2 &\leq \left( 1 - s_0\rho_n(e + \nu_{n+1})^{-1} \right) \|H_n\|^2 \\ &\quad + (s_0\rho_n)^{-1} (s_0\rho_n - \dot{\nu}_{n+1} - \kappa e)^2 (e + \nu_{n+1})^{-1} \|u_n\|^2. \end{aligned}$$

En multipliant cette dernière inégalité par  $(e + \nu_{n+1})^2$ , on a

$$(4.3.19) \quad \begin{aligned} (e + \nu_{n+1})^2 \|H_{n+1}\|^2 &\leq \left( (e + \nu_{n+1})^2 - s_0\rho_n(e + \nu_{n+1}) \right) \|H_n\|^2 \\ &\quad + (s_0\rho_n)^{-1} (s_0\rho_n - \dot{\nu}_{n+1} - \kappa e)^2 (e + \nu_{n+1}) \|u_n\|^2, \end{aligned}$$

tandis qu'un rapide calcul nous donne

$$(4.3.20) \quad (e + \nu_{n+1})^2 \leq (e + \nu_n)^2 + 2\dot{\nu}_{n+1}(e + \nu_{n+1}).$$

En combinant ces deux derniers résultats, nous obtenons

$$(4.3.21) \quad \begin{aligned} & (e + \nu_{n+1})^2 \|H_{n+1}\|^2 - (e + \nu_n)^2 \|H_n\|^2 \\ & + (s_0 \rho_n - 2\dot{\nu}_{n+1})(e + \nu_{n+1}) \|H_n\|^2 \\ & \leq (s_0 \rho_n)^{-1} (s_0 \rho_n - \dot{\nu}_{n+1} - \kappa e)^2 (e + \nu_{n+1}) \|u_n\|^2. \end{aligned}$$

Ainsi, comme  $\inf_{n \geq n_1} (s_0 \rho_n - 2\dot{\nu}_{n+1})(e + \nu_n) > 0$  (pour un certain entier  $n_1$  suffisamment grand),  $\sum_n \nu_n \|u_n\|^2 < +\infty$  (d'après (4.3.4g)) et que la quantité  $(e \rho_n - \dot{\nu}_{n+1} - \kappa e)$  est bornée, on en déduit que  $\sum_n (e + \nu_{n+1}) \|H_n\|^2 < +\infty$ . Cette dernière estimation nous donne (4.3.11a). De plus, les arguments précédents impliquent l'existence d'un  $l_1 \geq 0$  vérifiant  $\lim_{n \rightarrow +\infty} (e + \nu_n)^2 \|H_n\|^2 = l_1$ . Remarquons que nous avons aussi  $\lim_{n \rightarrow +\infty} (e + \nu_{n+1})^2 \|H_n\|^2 = l_1$  (car  $\frac{(e + \nu_n)}{(e + \nu_{n+1})} \rightarrow 1$  lorsque  $n \rightarrow +\infty$ ). Par conséquent, comme  $\sum_n (e + \nu_{n+1})^{-1} = +\infty$  (car  $\dot{\nu}_n$  est borné d'après (4.3.1e)), on en déduit que  $l_1 = 0$ , ce qui nous donne (4.3.11b).

Démontrons à présent les éléments (4.3.11c) et (4.3.11d). Un rapide calcul nous donne

$$(4.3.22) \quad \begin{aligned} & \|tM_\mu(z_{n-1})\|^2 \\ & \leq 2 \|tM_\mu(z_{n-1}) + \dot{x}_n + (\kappa - 1)u_{n-1}\|^2 + 2 \|\dot{x}_n + (\kappa - 1)u_{n-1}\|^2 \\ & \leq 2 \|tM_\mu(z_{n-1}) + \dot{x}_n + (\kappa - 1)u_{n-1}\|^2 + 4 \|\dot{x}_n\|^2 + 4(\kappa - 1)^2 \|u_{n-1}\|^2. \end{aligned}$$

Donc, en utilisant les estimations  $\sum_n \nu_n \|\dot{x}_n\|^2 < +\infty$ ,  $\sum_n \nu_n \|u_{n-1}\|^2 < +\infty$  et  $\sum_n \nu_n \|tM_\mu(z_{n-1}) + \dot{x}_n + (\kappa - 1)u_{n-1}\|^2 < +\infty$  (d'après (4.3.4e), (4.3.4g) et (4.3.11a) respectivement), on en déduit (4.3.11c). L'estimation (4.3.11d) provient quant à elle des estimations  $\sum_n \nu_n \|\dot{x}_n\|^2 < +\infty$  et  $\sum_n \nu_n \|M_\mu(z_{n-1})\|^2 < +\infty$  (d'après (4.3.4e) et (4.3.11c)).

Par ailleurs, en utilisant le Lemme 4.2.6 et en passant à la limite lorsque  $s \rightarrow 0^+$  dans (4.2.9), nous obtenons

$$\begin{aligned} & \nu_{n+1}^2 \|\dot{x}_{n+1}\|^2 - \nu_n^2 \|\dot{x}_n\|^2 + \frac{1}{2} \rho_n^{-1} (e + \nu_{n+1})^2 \|\dot{x}_{n+1} + \theta_n (y_n - x_n)\|^2 \\ & + (e + \nu_{n+1}) \rho_n^{-1} (e + \nu_{n+1})^2 \langle tM_\mu(z_n) - tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle \\ & + s_0 (e + \nu_{n+1}) \langle tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle \\ & + \frac{1}{2} (1 - \kappa) \rho_n \nu_n \nu_{n+1} \left\| y_n - x_n + \left( \frac{1}{\theta_n} - \frac{e}{\nu_n \rho_n} \right) \dot{x}_{n+1} \right\|^2 \\ & + \frac{1}{2} e (e + 2\nu_{n+1}) \rho_n^{-1} \|\dot{x}_{n+1}\|^2 = 0, \end{aligned}$$

ce qui d'après le Lemme 4.2.5 entraîne

$$\begin{aligned}
& \nu_{n+1}^2 \|\dot{x}_{n+1}\|^2 - \nu_n^2 \|\dot{x}_n\|^2 \\
& \leq -s_0(e + \nu_{n+1}) \langle tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle \\
& \quad - \frac{1}{2}(1 - \kappa)\rho_n \nu_n \nu_{n+1} \left\| y_n - x_n + \left( \frac{1}{\theta_n} - \frac{e}{\nu_n \rho_n} \right) \dot{x}_{n+1} \right\|^2 \\
& \leq s_0(e + \nu_{n+1}) |\langle tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle| \\
& \quad + |1 - \kappa| \rho_n \nu_{n+1}^2 \left\| y_n - x_n + \frac{1}{\theta_n} \dot{x}_{n+1} \right\|^2 + e^2 |1 - \kappa| \frac{\nu_{n+1}}{\nu_n} \rho_n^{-1} \|\dot{x}_{n+1}\|^2 \\
& = s_0(e + \nu_{n+1}) |\langle tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle| \\
& \quad + \frac{|1 - \kappa| \rho_n}{\theta_n^2} \nu_{n+1}^2 \|\dot{x}_{n+1} + \theta_n(y_n - x_n)\|^2 + e^2 |1 - \kappa| \frac{\nu_{n+1}}{\nu_n} \rho_n^{-1} \|\dot{x}_{n+1}\|^2.
\end{aligned}$$

Observons que la condition (4.3.1c) et la Remarque 4.3.17 nous donne

$$(4.3.23) \quad e + \nu_{n+1} \sim \nu_n, \quad \frac{\rho_n}{\theta_n^2} \nu_{n+1}^2 \sim \frac{1}{\kappa} \nu_n^2, \quad \frac{\nu_{n+1}}{\nu_n} \rho_n^{-1} \sim \kappa^{-1}, \quad \text{lorsque } n \rightarrow +\infty.$$

Par conséquent, en utilisant les estimations  $\sum_n \nu_n |\langle M_\mu(z_{n-1}), \dot{x}_{n+1} \rangle| < +\infty$ ,  $\sum_n \nu_n \|\dot{x}_{n+1}\|^2 < +\infty$  et  $\sum_n \nu_n^2 \|\dot{x}_{n+1} + \theta_n(y_n - x_n)\|^2 < +\infty$  (d'après (4.3.11d), (4.3.4e) et (4.3.4f) respectivement), on déduit des arguments précédents qu'il existe  $l_2 \geq 0$  tel que  $\lim_{n \rightarrow +\infty} \nu_n^2 \|\dot{x}_n\|^2 = l_2$ . De plus, en utilisant à nouveau l'estimation  $\sum_n \nu_n \|\dot{x}_{n+1}\|^2 < +\infty$  (d'après (4.3.4e)), et en rappelant que l'on a  $\sum_n (e + \nu_{n+1})^{-1} = +\infty$ , on obtient  $\liminf_{n \rightarrow +\infty} \nu_n^2 \|\dot{x}_n\|^2 = 0$ . Il en résulte que  $l_2 = 0$ , autrement dit (4.3.11e). Par ailleurs, l'estimation (4.3.4f) nous donne directement  $\|\dot{x}_{n+1} + \theta_n(y_n - x_n)\| = o(\nu_n^{-1})$ , on en déduit alors que  $\|y_n - x_n\| = o(\nu_n^{-1})$  (car  $\theta_n \sim \kappa$  lorsque  $n \rightarrow +\infty$ ), ce qui nous donne (4.3.11f). En combinant ce dernier résultat avec  $\|\dot{x}_n + tM_\mu(z_{n-1}) + (\kappa - 1)(y_{n-1} - x_{n-1})\| = o(\nu_n^{-1})$  (d'après (4.3.11b)), on obtient  $\|M_\mu(z_n)\| = o(\nu_n^{-1})$ , ce qui n'est autre que (4.3.11g).

Finalement, comme  $B$  est supposé être  $\lambda$ -cocoercif, nous savons que  $M_\mu$  est  $\mu(1 - \frac{\mu}{4\lambda})$ -cocoercif sous la condition  $\mu \in ]0, 2\lambda[$  (voir [8, Lemma B1]). Il en résulte, par un calcul rapide, que l'opérateur  $M_\mu$  est  $(\frac{2}{\mu})$ -Lipschitzien. De plus, comme  $x_{n+1} - z_n = -tM_\mu(z_n)$  (d'après la Remarque 4.2.13), on obtient

$$\begin{aligned}
 & \|M_\mu(y_{n+1})\|^2 \\
 (4.3.24) \quad & \leq 2 \left(\frac{2}{\mu}\right)^2 \|y_{n+1} - z_n\|^2 + 2\|M_\mu(z_n)\|^2 \\
 & \leq 4 \left(\frac{2}{\mu}\right)^2 \|y_{n+1} - x_{n+1}\|^2 + 4 \left(\frac{2}{\mu}\right)^2 \|x_{n+1} - z_n\|^2 + 2\|M_\mu(z_n)\|^2 \\
 & = 4 \left(\frac{2}{\mu}\right)^2 \|y_{n+1} - x_{n+1}\|^2 + \left(4 \left(\frac{2}{\mu}\right)^2 t^2 + 2\right) \|M_\mu(z_n)\|^2.
 \end{aligned}$$

Par conséquent, comme  $\sum_n \nu_n \|y_n - x_n\|^2 < +\infty$ , et  $\sum_n \nu_n \|M_\mu(z_n)\|^2 < +\infty$  (d'après (4.3.4g) et (4.3.11c)), on en déduit que  $\sum_n \nu_n \|M_\mu(y_n)\|^2 < +\infty$ , à savoir (4.3.11h). De plus, (4.3.24) montre que  $\nu_n \|M_\mu(y_n)\| \rightarrow 0$  lorsque  $n \rightarrow +\infty$ , car  $\nu_n \|y_n - x_n\| \rightarrow 0$  et  $\nu_n \|M_\mu(z_n)\| \rightarrow 0$  lorsque  $n \rightarrow +\infty$  (d'après (4.3.11f) et (4.3.11g) respectivement). On obtient alors (4.3.11i). ■

Énonçons à présent le résultat principal de cette section.

**Théorème 4.3.1.** *Soient  $\kappa \in ]0, 2[$ ,  $e > 0$  et  $s_0 \in ]0, e[$ . Supposons que les suites  $\{z_n, x_n, y_n\} \subset \mathcal{H}$  soient générées par *ALG1*, sous les conditions (4.1.2) et (4.3.1). Alors les suites  $(x_n)_n$ ,  $(y_n)_n$  et  $(z_n)_n$  convergent faiblement vers un élément  $\bar{x} \in S$  et nous avons les estimations suivantes :*

$$\begin{aligned}
 (4.3.25a) \quad & \|x_{n+1} - x_n\| = o(\nu_n^{-1}), \quad \|y_n - x_n\| = o(\nu_n^{-1}), \quad \|M_\mu(z_n)\| = o(\nu_n^{-1}), \\
 (4.3.25b) \quad & \sum_n \nu_n \|x_{n+1} - x_n\|^2 < +\infty, \quad \sum_n \nu_n \|y_n - x_n\|^2 < +\infty, \\
 (4.3.25c) \quad & \sum_n \nu_n \|M_\mu(z_n)\|^2 < +\infty, \quad \sum_n \nu_n^2 \|M_\mu(z_n) - M_\mu(z_{n-1})\|^2 < +\infty, \\
 (4.3.25d) \quad & \sum_n \langle M_\mu(z_{n-1}), x_n - q \rangle < +\infty \quad (\text{pour } q \in S).
 \end{aligned}$$

De plus la suite  $(\xi_n)$  définie pour  $n \geq 0$  par

$$(4.3.26) \quad \xi_n = (1 - \mu t^{-1}) z_{n-1} + \mu t^{-1} x_n,$$

converge faiblement vers  $\bar{x}$  et vérifie

$$(4.3.27) \quad \exists \xi_n^* \in (A + B)(\xi_n) \quad t.q. \quad \|\xi_n^*\| = o(\nu_n^{-1}) \quad \text{et} \quad \sum_n \nu_n \|\xi_n^*\|^2 < +\infty.$$

**Démonstration.** Nous prouvons ce théorème en considérant les étapes **(e1)**-**(e3)** suivantes :

**(e1)** Les résultats de (4.3.25a) sont donnés respectivement par (4.3.11e), (4.3.11f) et (4.3.11g). Les deux estimations de (4.3.25b) proviennent respectivement de (4.3.4e), (4.3.4g). La première et la deuxième estimation de (4.3.25c) proviennent de (4.3.11d) et de (4.3.4d) respectivement. Le dernier résultat (4.3.25d) est obtenu à partir de (4.3.4c).

**(e2)** Nous démontrons la convergence des itérés donnés par (4.1.7)-(4.1.8) à l'aide du lemme d'Opial qui garantit que la suite  $(x_n)$  converge vers un élément de  $S$ , à condition que les résultats suivants soient satisfaits :

(h1) pour tout  $q \in S$ , la suite  $(\|x_n - q\|)$  est convergente,

(h2) tout point d'accumulation faible de  $(x_n)$  appartient à  $S$ .

Démontrons (h1). Soit  $q \in S$ . Comme la suite  $(x_n)$  est bornée (d'après (4.3.4b)) et que  $\|M_\mu(z_n)\| = o(\nu_n^{-1})$  (d'après (4.3.11h)) on a

$$(4.3.28) \quad \langle M_\mu(z_{n-1}), x_n - q \rangle = o(\nu_n^{-1}).$$

De plus, nous savons que la suite  $(\mathcal{E}_n(s_0, q))_n$  est convergente (d'après le Lemme 4.3.8) et que

$$(4.3.29) \quad \begin{aligned} \mathcal{E}_n(s_0, q) = & \left(\frac{1}{2}\right) \|s_0(q - x_n) - \nu_n \dot{x}_n\|^2 + \left(\frac{1}{2}\right) s_0(e - s_0) \|x_n - q\|^2 \\ & + s_0(e + \nu_n) \langle tM_\mu(z_{n-1}), x_n - q \rangle. \end{aligned}$$

Ainsi, comme  $\nu_n \dot{x}_n \rightarrow 0$  (d'après (4.3.11e)) et que  $(e + \nu_n) \langle M_\mu(z_{n-1}), x_n - q \rangle \rightarrow 0$  lorsque  $n \rightarrow +\infty$  (d'après (4.3.28)), on en déduit que

$$(4.3.30) \quad \lim_{n \rightarrow +\infty} \mathcal{E}_n(s_0, q) = \lim_{n \rightarrow +\infty} \frac{1}{2} s_0 e \|x_n - q\|^2.$$

Ce qui nous donne (h1). Démontrons (h2). À partir de la définition de l'opérateur  $M_\mu$ , on obtient classiquement

$$(4.3.31) \quad M_\mu(z_n) \in B(z_n) + A(z_n - \mu M_\mu(z_n)).$$

Comme  $x_{n+1} - z_n = -tM_\mu(z_n)$  (d'après la Remarque 4.2.13), on a aussi

$$(4.3.32) \quad z_n - x_n = tM_\mu(z_n) + \dot{x}_{n+1}.$$

Donc, comme  $\|\dot{x}_{n+1}\| = o(\nu_n^{-1})$ ,  $\|M_\mu(z_n)\| = o(\nu_n^{-1})$  (d'après (4.3.25a)), on obtient

$$(4.3.33) \quad \|z_n - x_n\| \rightarrow 0 \quad (\text{lorsque } n \rightarrow +\infty).$$

Soit maintenant  $u$  un point d'accumulation faible de  $(x_n)$ . Autrement dit, il existe une sous-suite  $(x_{n_k})$  qui converge vers  $u$  lorsque  $k \rightarrow +\infty$ . En utilisant (4.3.25a) et (4.3.33), on obtient respectivement que  $\lim_{k \rightarrow +\infty} \|M_\mu(z_{n_k})\| = 0$  et que  $(z_{n_k})$  converge vers  $u$  lorsque  $k \rightarrow +\infty$ . En outre, la relation (4.3.31) nous donne

$$(4.3.34) \quad M_\mu(z_{n_k}) \in B(z_{n_k}) + A(z_{n_k} - \mu M_\mu(z_{n_k})).$$

Ainsi en passant à la limite lorsque  $k \rightarrow +\infty$  dans ce dernier résultat et en rappelant que le graphe d'un opérateur maximal monotone est demi-clos (voir, par exemple, [31]), on obtient  $0 \in (A + B)(u)$ , à savoir  $u \in S$ . Ceci prouve (h2).

On déduit alors du Lemme d'Opial que  $(x_n)$  converge faiblement vers un certain  $\bar{x} \in S$ . De plus, comme  $\lim_{n \rightarrow +\infty} \|z_n - x_n\| = 0$  et  $\lim_{n \rightarrow +\infty} \|y_n - x_n\| = 0$  (d'après (4.3.33) et (4.3.25a) respectivement) on obtient que les suites  $(z_n)$  et  $(y_n)$  convergent aussi vers  $\bar{x}$ .

**(e3)** Posons  $\xi_n = z_{n-1} - \mu M_\mu(z_{n-1})$ . La relation (4.3.31) nous donne alors

$$(4.3.35) \quad M_\mu(z_{n-1}) \in B(z_{n-1}) + A(\xi_n).$$

De plus, comme  $M_\mu(z_n) = t^{-1}(z_n - x_{n+1})$  (d'après la Remarque 4.2.13), nous pouvons réécrire la suite  $\xi_n$  sous la forme

$$(4.3.36) \quad \xi_n = (1 - \mu t^{-1}) z_{n-1} + \mu t^{-1} x_n.$$

On observe alors à partir de cette dernière égalité que  $\xi_n \rightarrow \bar{x}$  lorsque  $n \rightarrow +\infty$ , car  $x_n \rightarrow \bar{x}$  et  $z_n \rightarrow \bar{x}$  lorsque  $n \rightarrow +\infty$ . Par ailleurs, la relation (4.3.35) nous donne

$$(4.3.37) \quad M_\mu(z_{n-1}) + B(\xi_n) - B(z_{n-1}) \in B(\xi_n) + A(\xi_n),$$

ce qui, en posant

$$(4.3.38) \quad \xi_n^* = M_\mu(z_{n-1}) + B(\xi_n) - B(z_{n-1}),$$

s'écrit

$$(4.3.39) \quad \xi_n^* \in B(\xi_n) + A(\xi_n).$$

De plus comme  $B$  est  $(1/\lambda)$ -lipschitzien, on obtient à l'aide de la Remarque 4.2.13

$$(4.3.40) \quad \|B(\xi_n) - B(z_{n-1})\| \leq \lambda^{-1} \|\xi_n - z_{n-1}\| = \lambda^{-1} \mu \|M_\mu(z_{n-1})\|.$$

On déduit alors de (4.3.38) et (4.3.40) que

$$(4.3.41) \quad \|\xi_n^*\| \leq (1 + \lambda^{-1} \mu) \|M_\mu(z_{n-1})\|.$$

Ainsi, sachant que l'on a  $\|M_\mu(z_{n-1})\| = o(\nu_n^{-1})$  et  $\sum_n \nu_n \|M_\mu(z_{n-1})\|^2 < +\infty$  (d'après (4.3.25a) et (4.3.25b) respectivement), on en déduit que  $\|\xi_n^*\| = o(\nu_n^{-1})$  et  $\sum_n \nu_n \|\xi_n^*\|^2 < +\infty$ . Ceci achève la preuve. ■

Le résultat suivant est un cas particulier du théorème précédent lorsque  $\nu_n = n + \nu_0$ .

**Théorème 4.3.2.** *Soient  $\kappa \in ]0, 2[$  et  $e > 0$ . Soient  $\{z_n, x_n, y_n\} \subset \mathcal{H}$  les suites générées par (4.1.7)-(4.1.8) sous les conditions (4.1.2). Soient  $\{\mu, \nu_0\} \subset ]0, +\infty[$  des paramètres vérifiant*

$$(4.3.42a) \quad \mu \in ]0, 2\lambda[, \quad 0 < t < \mu(1 - \frac{\mu}{4\lambda}),$$

$$(4.3.42b) \quad \nu_n = n + \nu_0 \quad (\text{pour un certain } \nu_0 \geq 0),$$

$$(4.3.42c) \quad e > 2\kappa^{-1}, \quad s_0 \in ]2\kappa^{-1}, e[.$$

Alors les suites  $(x_n)$ ,  $(y_n)$  et  $(z_n)$  convergent faiblement vers un élément  $\bar{x} \in S$  et nous avons les estimations suivantes

$$(4.3.43a) \quad \|x_{n+1} - x_n\| = o(n^{-1}), \quad \|y_n - x_n\| = o(n^{-1}), \quad \|M_\mu(z_n)\| = o(n^{-1}),$$

$$(4.3.43b) \quad \sum_n n \|x_{n+1} - x_n\|^2 < +\infty, \quad \sum_n n \|y_n - x_n\|^2 < +\infty, \quad \sum_n n \|M_\mu(z_n)\|^2 < +\infty,$$

$$(4.3.43c) \quad \sum_n n^2 \|M_\mu(z_n) - M_\mu(z_{n-1})\|^2 < +\infty,$$

$$(4.3.43d) \quad \sum_n \langle M_\mu(z_{n-1}), x_n - q \rangle < +\infty \quad (\text{pour } q \in S).$$

De plus, la suite  $(\xi_n)$  définie pour  $n \geq 0$  par

$$(4.3.44) \quad \xi_n = (1 - \mu t^{-1}) z_{n-1} + \mu t^{-1} x_n.$$

converge faiblement vers  $\bar{x}$  et

$$(4.3.45) \quad \exists \xi_n^* \in (A + B)(\xi_n) \quad t.q. \quad \|\xi_n^*\| = o(n^{-1}) \quad \text{et} \quad \sum_n n \|\xi_n^*\|^2 < +\infty.$$

**Démonstration.** Le Théorème 4.3.2 découle immédiatement du Théorème 4.3.1 en remarquant que pour  $\nu_n = n + \nu_0$  (avec  $\nu_0 \geq 0$ ), on a

$$(4.3.46) \quad \limsup_{n \rightarrow +\infty} \dot{\nu}_{n+1} < \frac{1}{2} s_0 \kappa \iff s_0 > \frac{2}{\kappa}.$$

On vérifie alors, sous les conditions (4.3.42), que les conditions (4.3.1) sont satisfaites (pour  $n$  assez grand). ■

## 4.4 Une variante plus générale de l'algorithme proposé

Notre objectif ici est d'adapter l'algorithme [ALG1](#) au problème général d'inclusion monotone structurée suivant

$$(4.4.1) \quad \text{trouver } \bar{x} \in S := (B + \sum_{k=1}^p A_k)^{-1}(0) \neq \emptyset,$$

où  $B : \mathcal{H} \rightarrow \mathcal{H}$  est  $\lambda$ -cocoercif sur  $\mathcal{H}$ , et où  $(A_i)_{i=1}^p : \mathcal{H} \rightarrow \mathcal{H}$  est une famille de  $p$  opérateurs maximaux monotones pour lesquels la résolvante se calcule facilement.

Soulignons que l'algorithme [ALG2](#) fournit une autre façon de calculer la résolvante d'une somme d'opérateurs maximaux monotones en un point  $y \in \text{ran}(I + \sum_{i=1}^p A_i)$ . Cela s'observe en prenant  $B$  comme étant l'opérateur défini pour  $x \in \mathcal{H}$  par  $B(x) = x - y$  (donc  $B$  est 1-cocoercif).

La suite de notre méthodologie suit les mêmes lignes que celles de Maingé [71] et la preuve du théorème suivant est donnée par souci d'exhaustivité.

**Remarque 4.4.19** (voir Maingé [71]). Soient  $\{\varsigma_k\}_{k=1}^p \subset ]0, 1[$  tels que  $\sum_{k=1}^p \varsigma_k = 1$  et considérons l'espace de Hilbert  $E = \mathcal{H}^p$  muni du produit scalaire  $(\cdot|\cdot)$  défini, pour  $x = \{x_k\}_{k=1}^p$  et  $y = \{y_k\}_{k=1}^p$  appartenant à  $E$ , par  $(x|y) = \sum_{k=1}^p \varsigma_k \langle x_k, y_k \rangle$ . La norme induite de  $(\cdot|\cdot)$  est notée  $\|\cdot\|_{\mathcal{H}^p}$ .

Considérons également le problème auxiliaire

$$(4.4.2) \quad \text{trouver } \bar{z} \in S_p := \{z \in \mathcal{H}^p \mid \sum_{i=1}^p \varsigma_i z_i \in S\},$$

pour lequel il a été démontré qu'il possède un ensemble de solutions  $S_p$  non vides (dès lors que  $S \neq \emptyset$ ). Le problème (4.4.2) peut également être reformulé comme une inclusion monotone correspondant au problème (4.1.1)-(4.1.2) sur  $E$ . Introduisons, en effet, les applications  $\bar{A}_G$  et  $\bar{B}$  de  $E$  dans  $E$  définies pour  $\{x_i\}_{i=1}^p \in E$  par

$$(4.4.3) \quad \bar{A}_G(\{x_i\}_{i=1}^p) = \left\{ \frac{\mu}{\varsigma_i} A_i(x_i) \right\}_{i=1}^p, \quad \bar{B}(\{x_i\}_{i=1}^p) = \{B(x_i)\}_{i=1}^p,$$

et soit  $N_\Gamma : E \rightarrow 2^E$  le cône normal de l'ensemble convexe fermé (non vide)

$$(4.4.4) \quad \Gamma = \{\{x_i\}_{i=1}^p \in E \mid x_1 = x_2 = \dots = x_p\}.$$

On peut alors vérifier que  $\bar{A}_G$  et  $\bar{B}$  sont des opérateurs maximaux monotones sur  $E$ . Ainsi les opérateurs de réflexion  $R_{\bar{A}_G} = 2J_{\bar{A}_G} - I_{\mathcal{H}^p}$  et  $R_{N_\Gamma} = 2J_{N_\Gamma} - I_{\mathcal{H}^p}$  sont bien définis, ce qui nous permet de considérer les applications  $T$ ,  $T_1$  et  $T_2$  de  $E$  dans  $E$  tels que

$$(4.4.5) \quad T = T_1 \circ T_2, \quad T_1 = \frac{1}{2}(R_{\bar{A}_G} \circ R_{N_\Gamma} + I_{\mathcal{H}^p}) \quad \text{et} \quad T_2 = I_{\mathcal{H}^p} - \mu \bar{B} \circ J_{N_\Gamma}.$$

Il a été établi (voir [94, Propositions 4.11, 4.12 et 4.15]), que pour un certain opérateur maximal monotone  $A'_G : E \rightarrow 2^E$ , l'ensemble des solutions  $S_p$  de (4.4.2) peut s'écrire

$$(4.4.6) \quad S_p = (A'_G + \bar{B} \circ J_{N_\Gamma})^{-1}(0),$$

et que l'opérateur  $\bar{B} \circ J_{N_\Gamma}$  est  $\lambda$ -cocoercif (voir la section 4.2 de [94, Proposition 4.1]). De plus,  $S_p$  n'est rien d'autre que l'ensemble des points fixes de l'opérateur  $T$  où  $T$  s'écrit

$$(4.4.7) \quad T = J_{A'_G} \circ (I_{\mathcal{H}^p} - \mu \bar{B} \circ J_{N_\Gamma}).$$

Compte tenu de la remarque précédente, une stratégie pour résoudre (4.4.1) consiste d'abord à approcher un élément de  $S_p$  (à savoir un point fixe de  $T$ ) au moyen d'une suite  $(\zeta_n) = (\{\zeta_{n,k}\}_{k=1}^p) \subset E := \mathcal{H}^p$  générée par **ALG1**, dans le cadre de (4.4.7) et (4.4.6), comme suit :

$$(4.4.8a) \quad z_n = \zeta_n - \theta_n(y_n - \zeta_n) - \gamma_n(\zeta_n - z_{n-1}),$$

$$(4.4.8b) \quad \zeta_{n+1} = (1 - \frac{t}{\mu})z_n + \frac{t}{\mu}T(z_n),$$

$$(4.4.8c) \quad y_{n+1} = (1 - \kappa)y_n + \kappa\zeta_n.$$

Ensuite, nous obtenons un élément de  $S$  comme étant la limite de  $(x_n) \subset \mathcal{H}$  donnée par  $x_n = \sum_{k=1}^p \varsigma_k \zeta_{n,k}$ . Cela nous conduit (voir la preuve du théorème 4.4.1) à l'algorithme **ALG2** suivant

### ALG2

► **Initialisation** : Choisir  $\{z_{-1}, y_0, \zeta_0\} \subset \mathcal{H}^p$ ,  $\kappa \in ]0, 2[$ ,  $\{e, s_0, \nu_0, \mu, t\} \subset ]0, +\infty[$ ,  $(\varsigma_k)_{k=1}^p \subset ]0, 1[$ .

► **Étape principale** : Pour  $\{z_{n-1}, y_n, \zeta_n\} \subset \mathcal{H}^p$  donnés (avec  $n \geq 0$ ), on effectue

$$(4.4.9a) \quad z_n = \zeta_n - \theta_n(y_n - \zeta_n) - \gamma_n(\zeta_n - z_{n-1}),$$

$$(4.4.9b) \quad \omega_n = \sum_{k=1}^p \varsigma_k z_{n,k},$$

$$(4.4.9c) \quad \zeta_{n+1} = \left\{ z_{n,k} + \frac{t}{\mu} \left( J_{\frac{\mu}{\varsigma_k}}^{A_k} (2\omega_n - \mu B(\omega_n) - z_{n,k}) - \omega_n \right) \right\}_{k=1}^p,$$

$$(4.4.9d) \quad y_{n+1} = (1 - \kappa)y_n + \kappa\zeta_n,$$

$$(4.4.9e) \quad x_{n+1} = \sum_{k=1}^p \varsigma_k \zeta_{n,k}.$$

où les paramètres sont

$$(4.4.10) \quad \nu_n = n + \nu_0, \quad \theta_n = \frac{\nu_n - (1 - \kappa)\nu_{n+1}}{e + \nu_{n+1}} \quad \text{et} \quad \gamma_n := 1 - \frac{s_0}{e + \nu_{n+1}} \left( 1 - (1 - \kappa) \frac{\nu_{n+1}}{\nu_n} \right)$$

Le théorème suivant présente les taux de convergence  $\|\dot{\zeta}_n\|_{\mathcal{H}^p} = o(n^{-1})$  (pour la vitesse discrète) et  $\|T(z_n) - z_n\|_{\mathcal{H}^p} = o(n^{-1})$  (concernant les résidus de points fixes) pour des itérés  $(z_n)$  générés par **ALG2**.

**Théorème 4.4.1.** Soient  $(x_n, \zeta_n) \in (\mathcal{H}^p)^2$  des suites générées par *ALG2*. Soient  $\{s_0, t, e, \nu_0, \mu\} \subset ]0, +\infty[$ ,  $\kappa \in ]0, 2[$ , et  $(\varsigma_k)_{k=1}^p \subset ]0, 1[$  vérifiant  $\sum_{k=1}^p \varsigma_k = 1$ , ainsi que

$$(4.4.11) \quad 0 < \mu < 2\lambda, \quad 0 < t < \mu \left(1 - \frac{\mu}{4\lambda}\right).$$

$$(4.4.12) \quad \nu_n = n + \nu_0 \quad (\text{pour un certain } \nu_0 \geq 0),$$

$$(4.4.13) \quad e > 2\kappa^{-1} \quad \text{et} \quad s_0 \in ]2\kappa^{-1}, e[.$$

Alors nous avons les résultats suivants

$$(4.4.14a) \quad \|\dot{\zeta}_n\|_{\mathcal{H}^p} = o(n^{-1}), \quad \sum_n n \|\dot{\zeta}_n\|_{\mathcal{H}^p}^2 < +\infty,$$

$$(4.4.14b) \quad \|y_n - \zeta_n\|_{\mathcal{H}^p} = o(n^{-1}), \quad \sum_n n \|y_n - \zeta_n\|_{\mathcal{H}^p}^2 < +\infty,$$

$$(4.4.14c) \quad \|T(z_n) - z_n\|_{\mathcal{H}^p} = o(n^{-1}), \quad \sum_n n \|T(z_n) - z_n\|_{\mathcal{H}^p}^2 < +\infty,$$

$$(4.4.14d) \quad \sum_n n^2 \|(T(z_n) - z_n) - (T(z_{n-1}) - z_{n-1})\|_{\mathcal{H}^p}^2 < +\infty,$$

$$(4.4.14e) \quad \exists \bar{\zeta} \in S_p, \text{ t.q., (pour } k = 1, \dots, p), \bar{\zeta}_{n,k} \rightharpoonup \bar{\zeta}_k \text{ faiblement dans } \mathcal{H},$$

$$(4.4.14f) \quad x_n \rightharpoonup \bar{x} = \sum_{k=1}^p \varsigma_k \bar{\zeta}_k \in S, \text{ faiblement dans } \mathcal{H},$$

où  $T : \mathcal{H}^p \rightarrow \mathcal{H}^p$  est donné par (4.4.7) et où  $\|\cdot\|_{\mathcal{H}^p}^2 = \sum_{k=1}^p \varsigma_k \|\cdot\|^2$  sur  $\mathcal{H}^p$ .

**Démonstration.** On peut vérifier (voir [94, Lemma 4.10]), que pour  $\{x_k\}_{k=1}^p \in E$  l'on a

$$(4.4.15) \quad J_{N_\Gamma}(\{x_i\}_{i=1}^p) = \left\{ \sum_{j=1}^p \varsigma_j x_j \right\}_{i=1}^p,$$

$$(4.4.16) \quad R_{N_\Gamma}(\{x_k\}_{k=1}^p) = \left\{ 2 \sum_{i=1}^p \varsigma_i x_i - x_k \right\}_{k=1}^p,$$

$$(4.4.17) \quad R_{\bar{A}_G}(\{x_k\}_{k=1}^p) = \left\{ 2 J_{\frac{\mu}{\varsigma_k} A_k}(x_k) - x_k \right\}_{k=1}^p.$$

Alors, pour  $\{z_k\}_{k=1}^p \in E$ , on obtient d'après  $T_2 = I_{\mathcal{H}^p} - \mu \bar{B} \circ J_{N_\Gamma}$ , que

$$\begin{aligned} T_2(\{z_k\}_{k=1}^p) &= \{z_k\}_{k=1}^p - \mu(\bar{B} \circ J_{N_\Gamma})(\{z_k\}_{k=1}^p) \\ &= \{z_k\}_{k=1}^p - \mu \bar{B} \left( \left\{ \sum_{i=1}^p \varsigma_i z_i \right\}_{k=1}^p \right) \\ &= \{z_k - \mu B(\sum_{i=1}^p \varsigma_i z_i)\}_{k=1}^p. \end{aligned}$$

De plus, pour  $\{y_k\}_{k=1}^p \in E$  et en posant  $\bar{y} = \sum_{i=1}^p \varsigma_i y_i$ , on obtient d'après

$T_1 = \frac{1}{2}(R_{\bar{A}_G} \circ R_{N_\Gamma} + I_{\mathcal{H}^p})$  que

$$\begin{aligned} T_1(\{y_k\}_{k=1}^p) &= (1/2) \left( R_{\bar{A}_G}(\{2\bar{y} - y_k\}_{k=1}^p) + \{y_k\}_{k=1}^p \right) \\ &= (1/2) \left( \left\{ 2 J_{\frac{\mu}{\varsigma_k} A_k}(2\bar{y} - y_k) - 2\bar{y} + y_k \right\}_{k=1}^p + \{y_k\}_{k=1}^p \right) \\ &= \left\{ J_{\frac{\mu}{\varsigma_k} A_k}(2\bar{y} - y_k) - \bar{y} + y_k \right\}_{k=1}^p. \end{aligned}$$

Ainsi, en prenant  $\{y_k\}_{k=1}^p = T_2(\{z_k\}_{k=1}^p)$  et en posant  $\bar{z} = \sum_{i=1}^p \varsigma_i z_i$ , on a  $T(\{z_k\}_{k=1}^p) = T_1(\{y_k\}_{k=1}^p)$ , ou de manière équivalente

$$T(\{z_k\}_{k=1}^p) = T_1(\{z_k - \mu B(\bar{z})\}_{k=1}^p) = \left\{ J_{\frac{\mu}{\varsigma_k} A_k} (2\bar{z} - \mu B(\bar{z}) - z_k) - \bar{z} + z_k \right\}_{k=1}^p.$$

Il en résulte que

$$\left(1 - \frac{t}{\mu}\right) \{z_k\}_{k=1}^p + \frac{t}{\mu} T(\{z_k\}_{k=1}^p) = \left\{ z_k + \frac{t}{\mu} \left( J_{\frac{\mu}{\varsigma_k} A_k} (2\bar{z} - \mu B(\bar{z}) - z_k) - \bar{z} \right) \right\}_{k=1}^p.$$

On en déduit alors que [ALG2](#) peut se réécrire sous la forme [\(4.4.8\)](#). Par conséquent, les résultats [\(4.4.14a\)](#)-[\(4.4.14d\)](#) proviennent directement du [Théorème 4.3.2](#) tandis que [\(4.4.14e\)](#) est déduit de [\(4.4.14d\)](#). ■

## 4.5 Simulations numériques

### 4.5.1 Résolutions d'inclusion monotones structurées non potentielles

On se place dans  $\mathcal{H} = \mathbb{R}^{100}$  et on considère le problème d'inclusion monotone structuré suivant

$$(4.5.1) \quad \text{trouver } \bar{x} \in \mathbb{R}^{100} \text{ tel que } 0 \in A\bar{x} + B\bar{x},$$

où les opérateurs  $A, B \in \mathcal{M}_{100}(\mathbb{R})$  sont définis par

$$(4.5.2) \quad A = \begin{pmatrix} 0 & -1 & \dots & \dots & -1 \\ 1 & \ddots & -1 & \dots & \vdots \\ \vdots & 1 & \ddots & -1 & \vdots \\ \vdots & \vdots & 1 & \ddots & \vdots \\ 1 & \dots & \dots & \dots & 0 \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & \vdots \\ 0 & 0 & 1 & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 1 \end{pmatrix}.$$

On vérifie aisément que  $A$  et  $B$  sont des opérateurs (non potentiels) maximaux monotones tels que  $B$  est 1/2-cocoercif. On peut donc appliquer l'algorithme [ALG1](#) pour résoudre le problème [\(4.5.1\)](#). Pour chacune des Figures [4.5.1](#), [4.5.2](#) et [4.5.3](#), les conditions initiales sont  $x_0 = y_0 = z_0 = (10, 10, \dots, 10) \in \mathbb{R}^{100}$ .

La Figure 4.5.1 compare les performances de l'algorithme **ALG1** pour le calcul des zéros de  $A + B$  à l'aide de la suite  $\xi_n^* := A\xi_n + B\xi_n$  (où  $\xi_n = (1 - \mu t^{-1})z_{n-1} + \mu t^{-1}x_n$ ) avec différentes valeurs de  $\kappa$  et de  $e$ . Comme pour l'algorithme **RIGPA**, on observe que les performances de **ALG1** sont meilleures pour des valeurs croissantes de  $\kappa$  dans  $]0, 2[$ . La Figure 4.5.2, montre les performances de **ALG1** pour différentes valeurs de  $t$  dans  $]0, \sigma[$ . On observe que la convergence est meilleure pour des valeurs croissantes de  $t$  dans  $]0, \sigma[$ . La Figure 4.5.3, quant à elle, montre le comportement asymptotique des itérés  $\xi_n^*$  et de la vitesse discrète  $x_{n+1} - x_n$ .

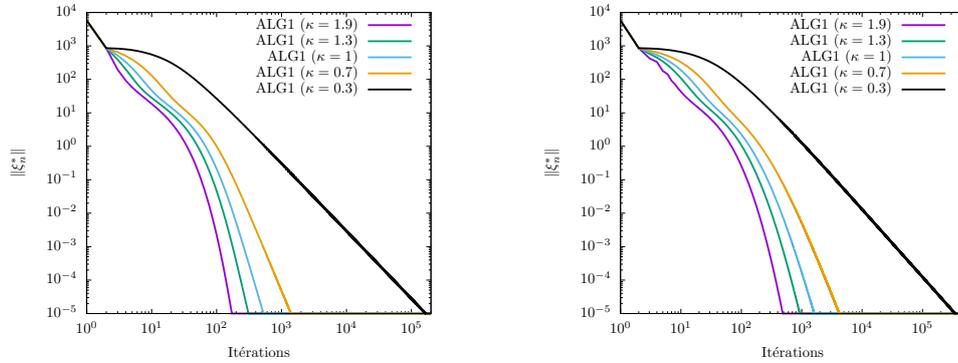


FIGURE 4.5.1 – Évolution des itérés  $\xi_n^*$  pour  $e = 7$  (gauche) et  $e = 15$  (droite) avec différentes valeurs de  $\kappa$ . Les autres paramètres sont  $\mu = 0.1$ ,  $s_0 = 6.75$ ,  $t = 0.5\sigma$ ,  $\nu_n = n + 2$

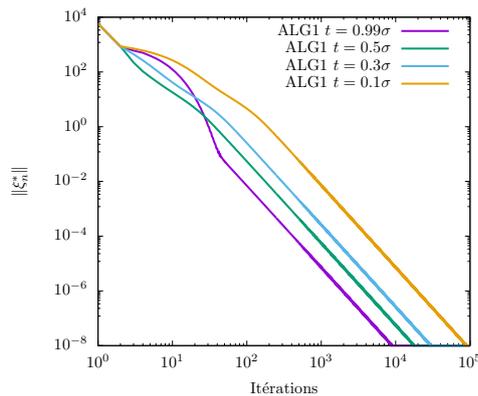


FIGURE 4.5.2 – Évolution des itérés  $\xi_n^*$  pour  $\kappa = 1.9$  avec différentes valeurs de  $t$ . Les autres paramètres sont  $\mu = 0.1$ ,  $e = 1.5 \times 2/\kappa$ ,  $\nu_n = n + 2$ ,  $s_0 = c \times 2/\kappa + (1 - c) \times e$  avec  $c = 0.001$ .

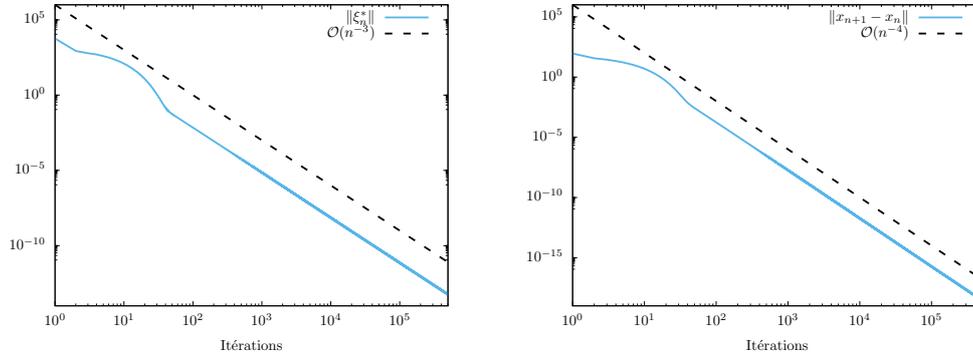


FIGURE 4.5.3 – Comparaison asymptotique itérés  $\xi_n^*$  (gauche) et de la vitesse discrète (droite). Les paramètres sont  $\kappa = 1.9$ ,  $e = 1.5 \times 2/\kappa$ ,  $\mu = 0.1$ ,  $t = 0.99\sigma$ ,  $\nu_n = n + 2$ ,  $s_0 = c \times 2/\kappa + (1 - c) \times e$  avec  $c = 0.001$ .

## Utilisation du « restarting » en présence de forte monotonie

Pour des opérateurs cocoercifs et fortement monotones, la méthode forward-backward possède un taux de convergence linéaire (voir chapitre 1, Section 1.4.2), tandis rien ne garantit que la méthode [ALG1](#) proposée dans ce chapitre possède un tel taux de convergence. En procédant de façon similaire à Kim [65], nous proposons d’avoir recours à du « restarting ». Supposons, pour cela, que l’on redémarre l’algorithme [ALG1](#) toutes les  $R$  itérations en initialisant la boucle extérieur ( $j + 1$ ) suivante par :

$$(4.5.3) \quad x_{j+1,0} = y_{j+1,0} = z_{j+1,-1} \leftarrow x_{j,R}.$$

Cela donne alors lieu à l’algorithme décrit ci-dessous que nous appellerons ALG1-r.

**ALG1-r**

**Données:**  $x_{0,0} \in \mathcal{H}$ ,  $\{\mu, e, t, \nu_k\} \subset ]0, +\infty[$ ,  $\kappa \in ]0, 2[$ ,  $s_0 \in ]0, e[$ ,  
 $\theta_k = \frac{\nu_k - (1-\kappa)\nu_{k+1}}{e + \nu_{k+1}}$  et  $\gamma_k = 1 - \frac{s_0}{e + \nu_{k+1}} \left(1 - (1-\kappa)\frac{\nu_{k+1}}{\nu_k}\right)$ ;

**Initialisation :**

$y_{0,0} \leftarrow x_{0,0}$  ;

$z_{0,-1} \leftarrow x_{0,0}$  ;

**pour**  $j$  allant de 0 à  $S - 1$  **faire**

**pour**  $k$  allant de 0 à  $R - 1$  **faire**

$z_{j,k} \leftarrow x_{j,k} - \theta_k(y_{j,k} - x_{j,k}) - \gamma_k(x_{j,k} - z_{j,k-1})$ ;

$x_{j,k+1} \leftarrow \left(1 - \frac{t}{\mu}\right) z_{j,k} + \left(\frac{t}{\mu}\right) J_{\mu}^A(z_{j,k} - \mu B(z_{j,k}))$ ;

$y_{j,k+1} \leftarrow (1 - \kappa)y_{j,k} + \kappa x_{j,k}$ ;

**fin**

$x_{j+1,0} \leftarrow x_{j,R}$  ;

$y_{j+1,0} \leftarrow x_{j,R}$  ;

$z_{j+1,-1} \leftarrow x_{j,R}$  ;

**fin**

La Figure 4.5.4 montre l'évolution de la norme de  $x_n$  en fonction du nombre d'itérations pour les algorithmes FB, ALG1, ALG1-r-5000, ALG1-r-2000 et ALG1-r-200 où ces trois derniers algorithmes correspondent à l'algorithme ALG1 lequel a été redémarré toutes les 5000, 2000 et 200 itérations respectivement avec l'initialisation (4.5.3). On observe la convergence linéaire de la méthode FB ainsi que la convergence sous-linéaire de l'algorithme ALG1 proposé. Cependant, en ayant recours au « restarting » mentionné ci-dessus, il semble possible d'obtenir une convergence linéaire et meilleure que celle de la méthode FB.

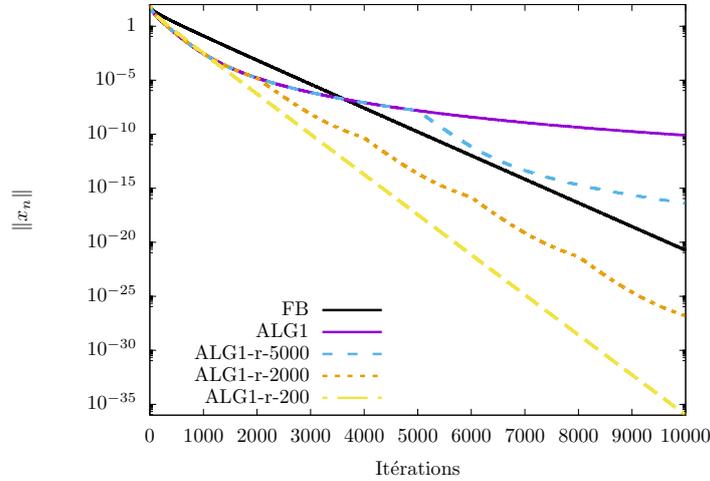


FIGURE 4.5.4 – Évolution de la norme  $x_n$  en fonction du nombre d'itérations pour les algorithmes FB, ALG1, et ALG1-r avec divers valeurs de « restarting ». Les paramètres sont les mêmes que ceux de la Figure 4.5.3

## 4.6 Annexe

### 4.6.1 Une formulation plus classique de ALG1.

Il est intéressant de noter que ALG1 entre dans le cadre d'un algorithme inertiel forward-backward relaxé plus familier.

**Proposition 4.6.1.** *Toutes suites  $\{z_n, x_n, y_n\}_{n \geq 1} \subset \mathcal{H}$  générées par ALG1 vérifient l'algorithme (4.1.10) pour  $n \geq 2$ .*

**Démonstration.** Soient  $\{z_n, x_n, y_n\}_{n \geq 1} \subset \mathcal{H}$  les suites générées par ALG1 et posons

$$(4.6.1) \quad u_n = y_n - x_n \quad \text{et} \quad v_n = x_n - z_{n-1}.$$

L'égalité (4.6.1) nous donne  $x_n = y_n - u_n$ . Ainsi d'après (4.1.7a)-(4.1.7c) on obtient

$$(4.6.2a) \quad v_n = y_n - u_n - z_{n-1},$$

$$(4.6.2b) \quad x_n = \left(1 - \frac{t}{\mu}\right)z_{n-1} + \frac{t}{\mu}J_{\mu}^A(z_{n-1} - \mu Bz_{n-1}),$$

$$(4.6.2c) \quad u_n = -\kappa^{-1}\dot{y}_{n+1}.$$

De plus, (4.1.7a) peut se réécrire

$$\begin{aligned}
 (4.6.3) \quad z_n &= x_n - \theta_n u_n - \gamma_n v_n \\
 &= y_n - u_n - \theta_n u_n - \gamma_n v_n \\
 &= y_{n+1} - \dot{y}_{n+1} - (1 + \theta_n) u_n - \gamma_n v_n.
 \end{aligned}$$

En combinant alors l'égalité précédente avec (4.6.2a) et (4.6.2c), on obtient

$$(4.6.4) \quad z_n = y_{n+1} - \dot{y}_{n+1} + \kappa^{-1} (1 + \theta_n) \dot{y}_{n+1} - \gamma_n (y_n + \kappa^{-1} \dot{y}_{n+1} - z_{n-1}),$$

à savoir

$$(4.6.5) \quad z_n = y_{n+1} + \kappa^{-1} (1 - \kappa + \theta_n - \gamma_n) \dot{y}_{n+1} - \gamma_n (y_n - z_{n-1}).$$

Par ailleurs, d'après (4.1.7c) on a aussi  $y_{n+1} = (1 - \kappa)y_n + \kappa x_n$ , ce qui d'après (4.6.2b) nous donne

$$(4.6.6) \quad y_{n+1} = (1 - \kappa)y_n + \kappa \left( \left(1 - \frac{t}{\mu}\right) z_{n-1} + \frac{t}{\mu} J_{\mu}^A(z_{n-1} - \mu B z_{n-1}) \right).$$

Par conséquent, on déduit de (4.6.2b), (4.6.5) et de (4.6.6) que les itérés  $\{z_n, x_n, y_n\}$  vérifient pour  $n \geq 2$ ,

$$(4.6.7a) \quad z_{n-1} = y_n + \kappa^{-1} (\xi_n - \gamma_n) (y_n - y_{n-1}) - \gamma_{n-1} (y_{n-1} - z_{n-2}),$$

$$(4.6.7b) \quad x_n = \left(1 - \frac{t}{\mu}\right) z_{n-1} + \frac{t}{\mu} J_{\mu}^A(z_{n-1} - \mu B z_{n-1}),$$

$$(4.6.7c) \quad y_{n+1} = (1 - \kappa)y_n + \kappa x_n,$$

où  $\xi_n = 1 - \kappa + \theta_{n-1}$ . En utilisant la définition de  $\theta_n$  (donnée par (4.1.8a)), on obtient de plus  $\xi_n = \frac{(1-\kappa)e + \nu_{n-1}}{(e + \nu_n)}$ . ■

## 4.6.2 Démonstration du Lemme 4.2.7

On observe d'après la Remarque 4.2.16 que les itérés  $\{z_n, x_n, y_n\}$  générés par (4.1.7a)-(4.1.7c) entrent dans le cas particulier de l'algorithme (4.2.14) pour  $n \geq 1$  en prenant

$$(4.6.8) \quad d_n = t M_{\mu}(z_n) - t \gamma_n M_{\mu}(z_{n-1}).$$

Par conséquent, en posant  $W_n = \|\dot{x}_{n+1} + \theta_n(y_n - x_n)\|^2$ , la Proposition 4.2.1 nous donne immédiatement

$$(4.6.9) \quad \begin{aligned} \dot{G}_{n+1}(s, q) + \frac{1}{2}\rho_n^{-1}\tau_n^2 W_n \\ + (s\tau_n)\langle d_n, x_{n+1} - q \rangle + \vartheta_n\rho_n^{-1}\tau_n^2\langle d_n, \dot{x}_{n+1} \rangle = -T_n(s), \end{aligned}$$

où  $\tau_n = e + \nu_{n+1}$  et  $\vartheta_n = 1 - s\frac{\rho_n}{\tau_n}$ . De plus, en posant

$$U_n := \langle tM_\mu(z_{n-1}), x_n - q \rangle \text{ et } Q_n := \langle tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle,$$

on obtient

$$\begin{aligned} \langle d_n, x_{n+1} - q \rangle \\ &= \langle tM_\mu(z_n) - t\gamma_n M_\mu(z_{n-1}), x_{n+1} - q \rangle \\ &= \langle tM_\mu(z_n), x_{n+1} - q \rangle - \gamma_n \langle tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle - \gamma_n \langle tM_\mu(z_{n-1}), x_n - q \rangle \\ &= U_{n+1} - \gamma_n Q_n - \gamma_n U_n, \end{aligned}$$

ainsi que

$$(4.6.10) \quad \langle d_n, \dot{x}_{n+1} \rangle = \langle tM_\mu(z_n) - t\gamma_n M_\mu(z_{n-1}), \dot{x}_{n+1} \rangle$$

$$(4.6.11) \quad = \langle tM_\mu(z_n), \dot{x}_{n+1} \rangle - \gamma_n \langle tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle$$

$$(4.6.12) \quad = \langle tM_\mu(z_n), \dot{x}_{n+1} \rangle - \gamma_n Q_n.$$

Il résulte de ces deux derniers résultats que

$$\begin{aligned} (s\tau_n)\langle d_n, x_{n+1} - q \rangle + \vartheta_n\rho_n^{-1}\tau_n^2\langle d_n, \dot{x}_{n+1} \rangle \\ &= (s\tau_n)\left(U_{n+1} - \gamma_n Q_n - \gamma_n U_n\right) + \vartheta_n\rho_n^{-1}\tau_n^2\left(\langle tM_\mu(z_n), \dot{x}_{n+1} \rangle - \gamma_n Q_n\right) \\ &= (s\tau_n)\left(U_{n+1} - \gamma_n U_n\right) - \left((s\tau_n) + \vartheta_n\rho_n^{-1}\tau_n^2\right)\gamma_n Q_n + \vartheta_n\rho_n^{-1}\tau_n^2\langle tM_\mu(z_n), \dot{x}_{n+1} \rangle. \end{aligned}$$

De plus, en remarquant que  $s\tau_n + \vartheta_n\rho_n^{-1}\tau_n^2 = \rho_n^{-1}\tau_n^2$ , on obtient

$$\begin{aligned} (s\tau_n)\langle d_n, x_{n+1} - q \rangle + \vartheta_n\rho_n^{-1}\tau_n^2\langle d_n, \dot{x}_{n+1} \rangle \\ &= (s\tau_n)\left(U_{n+1} - \gamma_n U_n\right) - \rho_n^{-1}\tau_n^2\gamma_n Q_n + \vartheta_n\rho_n^{-1}\tau_n^2\langle tM_\mu(z_n), \dot{x}_{n+1} \rangle. \end{aligned}$$

Introduisons à présent la suite  $Z_n = \langle tM_\mu(z_n) - tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle$ . On déduit alors de l'égalité précédente et de (4.6.9) que

$$\begin{aligned} \dot{G}_{n+1}(s, q) &+ \frac{1}{2}\rho_n^{-1}\tau_n^2 W_n + (s\tau_n)(U_{n+1} - \gamma_n U_n) + T_n(s) \\ &= -\rho_n^{-1}\tau_n^2 (-\gamma_n Q_n + \vartheta_n \langle tM_\mu(z_n), \dot{x}_{n+1} \rangle) \\ &= -\rho_n^{-1}\tau_n^2 (-\gamma_n Q_n + \vartheta_n \langle tM_\mu(z_n) - tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle + \vartheta_n \langle tM_\mu(z_{n-1}), \dot{x}_{n+1} \rangle) \\ &= -\rho_n^{-1}\tau_n^2 (-\gamma_n Q_n + \vartheta_n Z_n + \vartheta_n Q_n) \\ &= -\rho_n^{-1}\tau_n^2 ((\vartheta_n - \gamma_n)Q_n + \vartheta_n Z_n). \end{aligned}$$

Par ailleurs, on vérifie aisément d'après  $\vartheta_n = 1 - s\rho_n\tau_n^{-1}$  et  $\gamma_n = 1 - s_0\rho_n\tau_n^{-1}$  que

$$(4.6.13) \quad \vartheta_n - \gamma_n = s_0\rho_n\tau_n^{-1} - s\rho_n\tau_n^{-1} = \rho_n\tau_n^{-1}(s_0 - s).$$

Par conséquent, on déduit de ces deux derniers résultats que

$$(4.6.14) \quad \begin{aligned} \dot{G}_{n+1}(s, q) &+ \frac{1}{2}\rho_n^{-1}\tau_n^2 W_n + (s\tau_n)(U_{n+1} - \gamma_n U_n) + T_n(s) \\ &+ \tau_n(s_0 - s)Q_n + \rho_n^{-1}\tau_n^2 \vartheta_n Z_n = 0. \end{aligned}$$

De plus nous avons clairement

$$\tau_n(U_{n+1} - \gamma_n U_n) = \tau_n U_{n+1} - \tau_{n-1} U_n + (\tau_{n-1} - \tau_n \gamma_n) U_n,$$

tandis que les définitions de  $\gamma_n$  et  $\tau_n$  (à savoir  $\gamma_n = 1 - s_0 \frac{\rho_n}{\tau_n}$  et  $\tau_n = e + \nu_{n+1}$ ), nous donne

$$\tau_{n-1} - \tau_n \gamma_n = \tau_{n-1} - \tau_n + s_0 \rho_n = -\dot{\nu}_{n+1} + s_0 \rho_n.$$

On obtient alors

$$(4.6.15) \quad \tau_n(U_{n+1} - \gamma_n U_n) = \tau_n U_{n+1} - \tau_{n-1} U_n + (s_0 \rho_n - \dot{\nu}_{n+1}) U_n.$$

Ainsi, comme  $\mathcal{E}_n(s, q) = G_n(s, q) + (s\tau_{n-1})U_n$ , il résulte de (4.6.14) et (4.6.15), que l'on a

$$(4.6.16) \quad \begin{aligned} \dot{\mathcal{E}}_{n+1}(s, q) &+ s(e\rho_n - \dot{\nu}_{n+1})U_n + \frac{1}{2}\rho_n^{-1}\tau_n^2 W_n \\ &+ \rho_n^{-1}\tau_n^2 \vartheta_n Z_n + \tau_n(s_0 - s)Q_n = -T_n(s). \end{aligned}$$

Ceci nous donne (4.2.19) . ■



# Chapitre 5

## Conclusions et perspectives

Un cadre continu et des schémas numériques sont proposés dans cette thèse tant pour la résolution de problèmes de minimisation convexe structurés que pour la résolution d'inclusions monotones structurées.

S'agissant de la minimisation convexe structurée, nous avons étudié un algorithme (issu d'une variante discrète du cadre continu proposé). Nous avons démontré, entre autres, la propriété de convergence rapide vers zéro des sous-gradients de la fonction objectif même dans le cas non lisse, en sus des taux de convergence établis pour les algorithmes récemment proposés dans la littérature. Il serait intéressant de poursuivre ce travail dans le cadre des fonctions fortement convexes.

Concernant la résolution d'inclusions monotones structurées plus générales, nous avons développé un algorithme de type Forward-Backward qui nous permet d'obtenir des taux de convergence intéressants en termes de vitesse discrète et de résidus de points fixes, à l'instar de ceux obtenus pour la minimisation convexe structurée.

Il serait intéressant de transposer au cadre de la minimisation stochastique, les résultats obtenus dans cette thèse concernant la minimisation structurée non lisse. Il s'agira, par exemple, de remplacer le gradient de la fonction régulière par un estimateur de celui-ci afin d'éviter les calculs trop coûteux du gradient entier. Par ailleurs, afin de réduire la variance induite par cet estimateur, il serait intéressant d'avoir recours à une technique de réduction de la variance qui serait adaptée à des problèmes de machine learning régularisés de (très) grandes échelles.

On peut envisager l'étude d'une variante continue du schéma proposé dans le cadre de la résolution d'inclusions monotones, afin de mieux comprendre les processus sous-jacents.

Il serait, par ailleurs, intéressant d'étudier la convergence de ce schéma numérique en présence d'opérateurs fortement monotones.

Enfin, le cadre continu considéré nous a permis de construire un modèle intéressant pour des dynamiques discrètes, mais peut également servir de base pour le développement de nombreux autres systèmes dynamiques.

## Bibliographie

- [1] Z. ALLEN-ZHU. « Katyusha : The First Direct Acceleration of Stochastic Gradient Methods ». (2018). arXiv : [1603.05953](https://arxiv.org/abs/1603.05953).
- [2] F. ALVAREZ. « On the minimizing property of a second order dissipative system in Hilbert spaces ». *SIAM Journal Control and Optimization* **38**.(4) (2000), page(s): 1102-1119. DOI : [10.1137/S0363012998335802](https://doi.org/10.1137/S0363012998335802).
- [3] F. ALVAREZ et H. ATTOUCH. « An Inertial Proximal Method for Maximal Monotone Operators via Discretization of a Nonlinear Oscillator with Damping ». *Set-Valued Analysis* **9** (2001), page(s): 3-11. DOI : [10.1023/A:1011253113155](https://doi.org/10.1023/A:1011253113155).
- [4] F. ALVAREZ et al. « A second-order gradient-like dissipative dynamical system with Hessian driven damping. Application to Optimization and Mechanics ». *J. Math. Pures appl* **81**.(8) (2002), page(s): 747-779. DOI : [10.1016/S0021-7824\(01\)01253-3](https://doi.org/10.1016/S0021-7824(01)01253-3).
- [5] V. APIDOPOULOS, J.F. AUJOL et C. DOSSAL. « Convergence rate of inertial forward-backward algorithm beyond Nesterov's rule ». *Mathematical Programming* **180** (2020), page(s): 137-156. DOI : [10.1007/s10107-018-1350-9](https://doi.org/10.1007/s10107-018-1350-9).
- [6] H. ATTOUCH et F. ALVAREZ. « The heavy ball with friction dynamical system for convex constrained minimization problems. » *Nguyen V.H., Strodiot J.J., Tossings P. (eds) Optimization. Lecture Notes in Economics and Mathematical Systems* **481** (2000). DOI : [10.1007/978-3-642-57014-8\\_2](https://doi.org/10.1007/978-3-642-57014-8_2).
- [7] H. ATTOUCH, J. BOLTE et P. REDONT. « Optimizing properties of an inertial dynamical system with geometric damping. Link with proximal methods. » *Control and Cybernetics* **31** (2002), page(s): 643-657.

- [8] H. ATTOUCH et A. CABOT. « Convergence of a relaxed inertial forward-backward algorithm for structured monotone inclusions ». *Applied Math. Optimization* **80** (2019), page(s): 547-598. DOI : [10.1007/s00245-019-09584-z](https://doi.org/10.1007/s00245-019-09584-z).
- [9] H. ATTOUCH et A. CABOT. « Convergence rates of inertial Forward-Backward algorithms ». *SIAM Journal on Optimization* **28** (2018), page(s): 849-874.
- [10] H. ATTOUCH, X. GOUDOU et P. REDONT. « The heavy ball with friction method, I. The continuous dynamical system : global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system ». *Communications in Contemporary Mathematics, World Scientific Publishing Company* **2** (2000), page(s): 1-34. DOI : [10.1142/S0219199700000025](https://doi.org/10.1142/S0219199700000025).
- [11] H. ATTOUCH et S.C. LÁSZLÓ. « Continuous Newton-like inertial dynamics for monotone inclusions ». *Set-valued and variational Analysis* (2020). DOI : [10.1007/s11228-020-00564-y](https://doi.org/10.1007/s11228-020-00564-y).
- [12] H. ATTOUCH et S.C. LÁSZLÓ. « Newton-like inertial dynamics and proximal algorithms governed by maximally monotone operators ». *SIAM J. Control Optimization* **30.4** (2020), page(s): 3252-3283. DOI : [10.1137/20M1333316](https://doi.org/10.1137/20M1333316).
- [13] H. ATTOUCH et J. PEYPOUQUET. « Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators ». *Mathematical Programming* **174** (2019), page(s): 391-432. DOI : [10.1007/s10107-018-1252-x](https://doi.org/10.1007/s10107-018-1252-x).
- [14] H. ATTOUCH et J. PEYPOUQUET. « The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than  $1/k^2$  ». *SIAM J. Optimization* (2013), page(s): 1824-1834. DOI : [10.1137/15M1046095](https://doi.org/10.1137/15M1046095).
- [15] H. ATTOUCH, J. PEYPOUQUET et P. REDONT. « Fast convex optimization via inertial dynamics with hessian driven damping ». *Journal of Differential Equations* **261** (2016), page(s): 5734-5783. DOI : [10.1016/j.jde.2016.08.020](https://doi.org/10.1016/j.jde.2016.08.020).
- [16] H. ATTOUCH et al. « Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity ». *Math. Programming* **168** (2018), page(s): 123-175. DOI : [10.1007/s10107-016-0992-8](https://doi.org/10.1007/s10107-016-0992-8).

- [17] H. ATTOUCH et al. « Fast convex optimization via inertial dynamics combining viscous and Hessian-driven damping with time rescaling ». *Evolution Equations and Control Theory* (2021). DOI : [10.3934/eect.2021010](https://doi.org/10.3934/eect.2021010).
- [18] H. ATTOUCH et al. « First-order optimization algorithms via inertial systems with Hessian driven damping ». *Math. Programming* (2019). DOI : [10.1007/s10107-020-01591-1](https://doi.org/10.1007/s10107-020-01591-1).
- [19] J.F. AUJOL et C. DOSSAL. « Optimal rate of convergence of an ODE associated to the Fast Gradient Descent schemes for  $b > 0$ . » (2017). HAL : [hal-01547251](https://hal.archives-ouvertes.fr/hal-01547251).
- [20] J.F. AUJOL, C. DOSSAL et A. RONDEPIERRE. « Optimal convergence rates for Nesterov Acceleration ». *SIAM J. Optimization* **29**.(4) (2019), page(s): 3131-3153. DOI : [10.1137/18M1186757](https://doi.org/10.1137/18M1186757).
- [21] J.B. BAILLON. « Un exemple concernant le comportement asymptotique de la solution du problème  $du + \partial\phi(u) \ni 0$ . » *Journal of Functional Analysis* **28** (1978), page(s): 369-376.
- [22] H.H. BAUSCHKE et P.L. COMBETTES. « Convex analysis and monotone operator theory in Hilbert spaces ». *2nd edition, Springer, New York* (2017). DOI : [10.1007/978-3-319-48311-5](https://doi.org/10.1007/978-3-319-48311-5).
- [23] A. BECK. *First-order methods in optimization*. Society for Industrial & Applied Mathematics, 2017. DOI : [10.1137/1.9781611974997](https://doi.org/10.1137/1.9781611974997).
- [24] A. BECK et M. TEBoulLE. « A fast iterative shrinkage-thresholding algorithm for linear inverse problems ». *SIAM J. Imaging Sci.* **2**.(1) (2009), page(s): 183-202. DOI : [10.1137/080716542](https://doi.org/10.1137/080716542).
- [25] G. BELGIOSO et S. GRAMMATICO. « Semi-decentralized Nash equilibrium seeking in aggregative games with separable coupling constraints and non-differentiable cost functions ». *IEEE Control Systems Letters* **1**.(2) (2017), page(s): 400-405. DOI : [10.1109/LCSYS.2017.2718842](https://doi.org/10.1109/LCSYS.2017.2718842).
- [26] R.I. BOŦ et E.R. CSETNEK. « Second order forward-backward dynamical systems for monotone inclusion problems ». *SIAM Journal on Control and Optimization* **54**.(3) (2016), page(s): 1423-1443. DOI : [10.1137/15M1012657](https://doi.org/10.1137/15M1012657).
- [27] L. BOTTOU et O. BOUSQUET. « The tradeoffs of large scale learning ». *Advances in Neural Information Processing Systems (NIPS)* **20** (2008), page(s): 161-168.

- [28] L. BOTTOU et Y. LE CUN. « On-line learning for very large data sets. » *Applied Stochastic Models in Business and Industry* **21** (2005), page(s): 137-151. DOI : [10.1002/asmb.538](https://doi.org/10.1002/asmb.538).
- [29] S. BOYD. « Subgradient Methods ». *Notes for EE364b, Stanford University*, (2013).
- [30] H. BREZIS. *Function Analysis, Sobolev Spaces and Partial Differential Equations*. 2010. DOI : [10.1007/978-0-387-70914-7](https://doi.org/10.1007/978-0-387-70914-7).
- [31] H. BREZIS. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. T. 5. Burlington, MA : Elsevier Science, 1973.
- [32] H. BREZIS et P. L. LIONS. « Produits infinis de résolvantes. » *Israel Journal of Mathematics* **29**.(4) (1978), page(s): 329-345.
- [33] L.M. BRICEÑO-ARIAS et P.L. COMBETTES. « Monotone operator methods for Nash equilibria in non-potential games ». *Springer Proceedings in Mathematics & Statistics Computational and Analytical Mathematics* **50** (2013), page(s): 143-159. DOI : [10.1007/978-1-4614-7621-4\\_9](https://doi.org/10.1007/978-1-4614-7621-4_9).
- [34] R. E. BRUCK. « An iterative solution of a variational inequality for certain monotone operators in Hilbert space ». *Bulletin of the American Mathematical Society* **81** (1975), page(s): 890-892. DOI : [10.1090/S0002-9904-1975-13874-2](https://doi.org/10.1090/S0002-9904-1975-13874-2).
- [35] R. E. BRUCK. « Asymptotic convergence of nonlinear contraction semigroups in Hilbert space. » *Journal of Functional Analysis* **18**.(1) (1975), page(s): 15-26. DOI : [10.1016/0022-1236\(75\)90027-0](https://doi.org/10.1016/0022-1236(75)90027-0).
- [36] A. CABOT, H. ENGLER et S. GADAT. « Second order differential equations with asymptotically small dissipation and piecewise flat potentials ». *Elect. J. of Differential Equations* **17** (2009), page(s): 33-38.
- [37] A. CABOT et P. FRANKEL. « Asymptotics for some semilinear hyperbolic equations with non-autonomous damping ». *Journal of Differential Equations* **252** (2012), page(s): 294-322. DOI : [10.1016/j.jde.2011.09.012](https://doi.org/10.1016/j.jde.2011.09.012).
- [38] V. CEVHER et B.C. VU. « A Reflected Forward-Backward Splitting Method for Mono- tone Inclusions Involving Lipschitzian Operators ». *Set Valued and Variational Analysis* (2020). DOI : [10.1007/s11228-020-00542-4](https://doi.org/10.1007/s11228-020-00542-4).

- [39] A. CHAMBOLLE. « An algorithm for total variation minimization and applications ». *J. Math. Imaging Vis* (**20**) (2004), page(s): 89-97. DOI : [10.1023/B:JMIV.0000011325.36760.1e](https://doi.org/10.1023/B:JMIV.0000011325.36760.1e).
- [40] A. CHAMBOLLE et C. DOSSAL. « On the convergence of the iterates of “Fast Iterative Shrinkage/Thresholding Algorithm” ». *Journal of Optimization Theory and Applications* **166** (2015), page(s): 968-982. DOI : [10.1007/s10957-015-0746-4](https://doi.org/10.1007/s10957-015-0746-4).
- [41] A. CHAMBOLLE et P.L. LIONS. « Image recovery via total variation minimization and related problems ». *Numer. Math* **76** (1997), page(s): 167-188. DOI : [10.1007/s002110050258](https://doi.org/10.1007/s002110050258).
- [42] A. CHAMBOLLE et T. POCK. « An introduction to continuous optimization for imaging ». *Acta Numerica, Cambridge University Press (CUP)* **25** (2016), page(s): 161-319. DOI : [10.1017/S096249291600009X](https://doi.org/10.1017/S096249291600009X).
- [43] P.L. COMBETTES et J.C. PESQUET. « Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators ». *Set Valued and Variational Analysis* **20.2** (2012), page(s): 307-330.
- [44] P.L. COMBETTES et J.C. PESQUET. « Proximal Splitting Methods in Signal Processing ». *Bauschke H., Burachik R., Combettes P., Elser V., Luke D., Wolkowicz H. (eds) Fixed-Point Algorithms for Inverse Problems in Science and Engineering. Springer Optimization and Its Applications* **49** (2011). DOI : [10.1007/978-1-4419-9569-8\\_10](https://doi.org/10.1007/978-1-4419-9569-8_10).
- [45] P.L. COMBETTES et V.R. WAJS. « Signal recovery by proximal forward-backward splitting ». *Multiscale Model Simul* **4** (2005), page(s): 1168-1200. DOI : [10.1137/050626090](https://doi.org/10.1137/050626090).
- [46] E. CORMAN et X. YUAN. « A generalized proximal point algorithm and its convergence rate ». *SIAM J. Optim* **24.4** (2014), page(s): 1614-1638. DOI : [10.1137/130940402](https://doi.org/10.1137/130940402).
- [47] J.B. CRUZ et T. NGHIA. « On the convergence of the proximal forward-backward splitting method with linesearches ». *Optim. Methods and Software* **31.(6)** (2016), page(s): 1209-1238. DOI : [10.1080/10556788.2016.1214959](https://doi.org/10.1080/10556788.2016.1214959).

- [48] E.R. CSETNEK. « Continuous dynamics related to monotone inclusions and non-smooth optimization problems, » *Set Valued and Variational Analysis* **28** (2020), page(s): 611-642. DOI : [10.1007/s11228-020-00548-y](https://doi.org/10.1007/s11228-020-00548-y).
- [49] I. DAUBECHIES, M. DEFRISE et C. DE MOL. « An iterative thresholding algorithm for linear inverse problems with a sparsity constraint ». *Comm. Pure Appl. Math.* **57** (2004), page(s): 1413-1457. DOI : [10.1002/cpa.20042](https://doi.org/10.1002/cpa.20042).
- [50] K. DEIMLING et S. GRAMMATICO. « Zeros of accretive operators ». *Manuscripta Mathematica* (2017), page(s): 365-374. DOI : [10.1007/bf01171148](https://doi.org/10.1007/bf01171148).
- [51] J. DOUGLAS et H. H. RACHFORD. « On the numerical solution of heat conduction problems in two and three space variables ». *Trans. Amer. Math. Society* **82** (1956), page(s): 421-439.
- [52] Y. DRORI et M. TEBoulLE. « Performance of first-order methods for smooth convex minimization : A novel approach ». *Math. Programming* **145.2** (2014), page(s): 451-482.
- [53] I. EKELAND et R. TEMAM. *Convex Analysis and Variational Problems*. 1976.
- [54] P. GETREUER. « Rudin-Osher-Fatemi Total Variation Denoising using Split Bregman ». *Image Processing On Line* **2** (2012), page(s): 74-95. DOI : [10.5201/ipol.2012.g-tvd](https://doi.org/10.5201/ipol.2012.g-tvd).
- [55] A. A. GOLDSTEIN. « Convex programming in Hilbert spaces ». *Bull. Amer. Math. Soc.* (70) (1964), page(s): 709-710.
- [56] X. GOUDOU et J. MUNIER. « The gradient and heavy ball with friction dynamical systems : the quasiconvex case. » *Math. Programming* **116** (2007), page(s): 173-191. DOI : [10.1007/s10107-007-0109-5](https://doi.org/10.1007/s10107-007-0109-5).
- [57] O. GÜLER. « New proximal point algorithms for convex minimization ». *SIAM J. Optimization* **2**.(4) (1992), page(s): 649-664. DOI : [10.1137/0802032](https://doi.org/10.1137/0802032).
- [58] O. GÜLER. « On the convergence of the proximal point algorithm for convex minimization ». *SIAM J. Control Optimization* **29** (1991), page(s): 403-419. DOI : [10.1137/0329022](https://doi.org/10.1137/0329022).
- [59] I GUYON et al. « Structural risk minimization for character recognition ». *NIPS'91 : Proceedings of the 4th International Conference on Neural Information Processing Systems* (1991), page(s): 471-479.

- [60] A. HARAUX. « Systèmes dynamiques dissipatifs et applications ». *Masson RMA17* (1991).
- [61] A. HARAUX et M.A. JENDOUBI. « Convergence of solutions of second order gradient like systems with analytic nonlinearities ». *Journal of Differential Equations* **144**.(2) (1998), page(s): 312-320. DOI : [10.1006/jdeq.1997.3393](https://doi.org/10.1006/jdeq.1997.3393).
- [62] F. IUTZELER et J M. HENDRICKX. « A Generic online acceleration scheme for Optimization algorithms via Relaxation and Inertia ». (2017). arXiv : [1603.05398v3](https://arxiv.org/abs/1603.05398v3).
- [63] R. JOHNSON et T. ZHANG. « Accelerating stochastic gradient descent using predictive variance reduction ». *Advances in Neural Information Processing Systems* **26** (2013).
- [64] J. KIEFER et J. WOLFOWITZ. « Stochastic estimation of the maximum of a regression function ». *Ann. Math. Stat.* **23**.(3) (1952), page(s): 462-466. DOI : [10.1214/aoms/1177729392](https://doi.org/10.1214/aoms/1177729392).
- [65] D. KIM. « Accelerated proximal point method for maximally monotone operators ». *Math. Program* (2021). DOI : [10.1007/s10107-021-01643-0](https://doi.org/10.1007/s10107-021-01643-0).
- [66] F. LABARRE et P.E. MAINGÉ. « First-Order frameworks for continuous Newton-like dynamics governed by maximally monotone operators ». *Set-Valued and Variational Analysis* (2021). DOI : [10.1007/s11228-021-00593-1](https://doi.org/10.1007/s11228-021-00593-1).
- [67] N. LE ROUX, M. SCHMIDT et F. BACH. « A Stochastic Gradient Method with an Exponential Convergence Rate for Finite Training Sets ». *NIPS'12 - 26 th Annual Conference on Neural Information Processing Systems* (2012). HAL : [hal-00674995v4](https://hal.archives-ouvertes.fr/hal-00674995v4).
- [68] M. LI et al. « Shrinking gradient descent algorithms for total variation regularized image denoising ». *Comput Optim Appl* **68** (2017), page(s): 643-660. DOI : [10.1007/s10589-017-9931-8](https://doi.org/10.1007/s10589-017-9931-8).
- [69] P.L LIONS et B. MERCIER. « Splitting algorithms for the sum of two nonlinear operators ». *SIAM J. Numer. Anal.* **16**.(6) (1979), page(s): 964-979. DOI : [10.1137/0716071](https://doi.org/10.1137/0716071).

- [70] D.A LORENZ et T. POCK. « An inertial forward-backward algorithm for monotone inclusions ». *J. Math. Imaging Vision* **51** (2015), page(s): 311-325. DOI : [10.1007/s10851-014-0523-2](https://doi.org/10.1007/s10851-014-0523-2).
- [71] P.E. MAINGÉ. « Fast convergence of generalized forward-backward algorithm for structured monotone inclusions ». (2021). arXiv : [2107.10107](https://arxiv.org/abs/2107.10107).
- [72] P.E. MAINGÉ et F. LABARRE. « Accelerated methods with fastly vanishing subgradients for structured non-smooth minimization ». *Numerical Algorithms* (2021). DOI : [10.1007/s11075-021-01181-y](https://doi.org/10.1007/s11075-021-01181-y).
- [73] P.E. MAINGÉ et S. MARUSTER. « Convergence in norm of modified Krasnoselski-Mann iterations for fixed points of demicontractive mappings ». *Applied Mathematics and Computation* **217**.(24) (2011), page(s): 9864-9874. DOI : [10.1016/j.amc.2011.04.068](https://doi.org/10.1016/j.amc.2011.04.068).
- [74] Y. MALITSKY. « Proximal extrapolated gradient methods for variational inequalities ». *Optimization Methods and Software* **33**.(1) (2018), page(s): 140-164. DOI : [10.1080/10556788.2017.1300899](https://doi.org/10.1080/10556788.2017.1300899).
- [75] B. MARTINET. « Détermination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox ». *C.R. Acad. Sc. Paris* (1972), page(s): 163-165.
- [76] B. MARTINET. « Régularisation d'inéquations variationnelles par approximations successives ». *Rev. Francaise Informat. Recherche Opérationnelle* **4** (1970), page(s): 154-158. DOI : [10.1051/m2an/197004R301541](https://doi.org/10.1051/m2an/197004R301541).
- [77] R. MAY. « Asymptotic for a second order evolution equation with convex potential and vanishing damping term ». *Turkish Journal of Mathematics* **41**.(3) (2015).
- [78] A. MOUDAFI et M. OLINY. « Convergence of a splitting inertial proximal method for monotone operators ». *J. Comput. Appl. Math.* **155**.(2) (2003), page(s): 447-454. DOI : [10.1016/S0377-0427\(02\)00906-8](https://doi.org/10.1016/S0377-0427(02)00906-8).
- [79] A. NEDIC et D. BERTSEKAS. « Convergence rate of incremental subgradient algorithms ». *Stochastic Optimization : Algorithms and Applications* **54** (2001), page(s): 263-304. DOI : [10.1007/978-1-4757-6594-6\\_11](https://doi.org/10.1007/978-1-4757-6594-6_11).

- [80] A. NEMIROVSKI et al. « Robust stochastic approximation approach to stochastic programming ». *SIAM Journal on Optimization* **19.4** (2009), page(s): 1574-1609. DOI : [10.1137/070704277](https://doi.org/10.1137/070704277).
- [81] Y. NESTEROV. « A method of solving a convex programming problem with convergence rate  $O(1/k^2)$  ». *Soviet Mathematics Doklady* **27** (1983), page(s): 372-376.
- [82] Y. NESTEROV. « Gradient methods for minimizing composite objective function ». *Math. Programming* (2013), page(s): 125-161. DOI : [10.1007/s10107-012-0629-5](https://doi.org/10.1007/s10107-012-0629-5).
- [83] Y. NESTEROV. « Introductory lectures on convex optimization : A basic course ». *Kluwer Academic Publishers* **87** (2004). DOI : [10.1007/978-1-4419-8853-9](https://doi.org/10.1007/978-1-4419-8853-9).
- [84] Y. NESTEROV. « Smooth minimization of non-smooth functions ». *Mathematical Programming* **103** (2005), page(s): 127-152. DOI : [10.1007/s10107-004-0552-5](https://doi.org/10.1007/s10107-004-0552-5).
- [85] A. NITANDA. « Stochastic proximal gradient descent with acceleration techniques ». *Advances in Neural Information Processing Systems* **27** (2014), page(s): 1574-1582.
- [86] P. OCHS, T. BROX et T. POCK. « iPiasco : Inertial proximal algorithm for strongly convex optimization ». *J Math Imaging Vision* **53** (2015), page(s): 171-181. DOI : [10.1007/s10851-015-0565-0](https://doi.org/10.1007/s10851-015-0565-0).
- [87] Z. OPIAL. « Weak convergence of the sequence of successive approximations for nonexpansive mappings ». *Bull. Amer. Math. Soc.* **73** (1967), page(s): 591-597.
- [88] N. PARIKH et S. BOYD. « Proximal algorithms ». *Foundations and Trends<sup>®</sup> in Optimization* **1.3** (2014), page(s): 127-239. DOI : [10.1561/2400000003](https://doi.org/10.1561/2400000003).
- [89] G.B. PASSTY. « Ergodic convergence to a zero of the sum of monotone operators in Hilbert space ». *J. Math. Anal. Appl.* **72** (1979), page(s): 383-390. DOI : [10.1016/0022-247X\(79\)90234-8](https://doi.org/10.1016/0022-247X(79)90234-8).
- [90] D.W. PEACEMAN et H. H. RACHFORD. « The numerical solution of parabolic and elliptic differential equations ». *J. Soc. Indust. Appl. Math.* **3.1** (1955), page(s): 28-41.

- [91] B.T POLYAK. « Some methods of speeding up the convergence of iteration methods ». *USSR Computational Mathematics and Mathematical Physics* **4**.(5) (1964), page(s): 1-17. DOI : [10.1016/0041-5553\(64\)90137-5](https://doi.org/10.1016/0041-5553(64)90137-5).
- [92] B.T. POLYAK. « A new method of stochastic approximation type ». *Avtomatika i Telemekhanika*, **51** (1990), page(s): 98-107.
- [93] B.T. POLYAK et A.B. JUDITSKY. « Acceleration of stochastic approximation by averaging ». *SIAM Journal on Control and Optimization* **30**.4 (1992), page(s): 838-855. DOI : [10.1137/0330046](https://doi.org/10.1137/0330046).
- [94] H. RAGUET, J. FADILI et G. PEYRÉ. « A generalized forward-backward splitting ». *SIAM J. Imaging Sci* (2013), page(s): 1199-1226. DOI : [10.1137/120872802](https://doi.org/10.1137/120872802).
- [95] S. REICH. « An iterative procedure for constructing zeros of accretive sets in Banach spaces ». *Nonlinear Analysis : Theory, Methods & Applications* (1978), page(s): 85-92. DOI : [10.1016/0362-546x\(78\)90044-5](https://doi.org/10.1016/0362-546x(78)90044-5).
- [96] H. ROBBINS et S. MONRO. « A stochastic approximation model ». *Ann. Math. Stat.* **22**.(3) (1951), page(s): 400-407. DOI : [10.1214/aoms/1177729586](https://doi.org/10.1214/aoms/1177729586).
- [97] R. T. ROCKAFELLAR. « Monotone operators and the proximal point algorithm ». *SIAM journal on control and optimization* **14**.(5) (1976), page(s): 877-898. DOI : [10.1137/0314056](https://doi.org/10.1137/0314056).
- [98] R.T ROCKAFELLAR et R.J.B. WETS. « Variational analysis fundamental principles of mathematical sciences ». *Springer* **317** (1998).
- [99] L.I. RUDIN, E. OSHER et E. FATEMI. « Nonlinear total variation based noise removal algorithms ». *Physica D : Nonlinear Phenomena* **60** (1992), page(s): 259-268. DOI : [10.1016/0167-2789\(92\)90242-F](https://doi.org/10.1016/0167-2789(92)90242-F).
- [100] D. RUPPERT. « Efficient estimations from a slowly convergent robbins-monro process ». *Technical report, Cornell University Operations Research and Industrial Engineering* (1988).
- [101] K. SCHEINBERG, D. GOLDFARB et X. BAI. « Fast first-order methods for composite convex optimization with backtracking ». *Foundations of Computational Mathematics* **14**.(3) (2014), page(s): 389-417. DOI : [10.1007/s10208-014-9189-9](https://doi.org/10.1007/s10208-014-9189-9).

- [102] S. SHALEV-SHWARTZ et al. « Stochastic convex optimization ». *Conference on Learning Theory (COLT)*, (2009).
- [103] B. SHI et al. « Understanding the acceleration phenomenon via high-resolution differential equations ». (2018). arXiv : [1810.08907v3](https://arxiv.org/abs/1810.08907v3).
- [104] M. SODOLOV. « Incremental gradient algorithms with stepsizes bounded away from zero ». *Computational Optimization and Applications* **11** (1998), page(s): 23-25. DOI : [10.1023/A:1018366000512](https://doi.org/10.1023/A:1018366000512).
- [105] W. SU, S. BOYD et E. J. CANDÈS. « A differential equation for modeling Nesterov's accelerated gradient method : theory and insights ». *Neural Information Processing Systems* **27** (2014), page(s): 2510-2518.
- [106] P. TSENG. « A modified Forward-Backward splitting method for maximal monotone mappings ». *SIAM Journal Control and Optimization* **38**.(2) (2000), page(s): 431-446. DOI : [10.1137/S0363012998338806](https://doi.org/10.1137/S0363012998338806).
- [107] Vapnik V. « Principles of risk minimization for learning theory Principles of risk minimization for learning theory ». *NIPS'91 : Proceedings of the 4th International Conference on Neural Information Processing Systems* (1991), page(s): 831-838.
- [108] S. WINKLER. « Digital Video Quality : Vision Models and Metrics. » *John Wiley & Sons Ltd., West Sussex* (2005).
- [109] L. XIAO et T. ZHANG. « A proximal stochastic gradient method with progressive variance reduction ». (2014). arXiv : [1403.4699](https://arxiv.org/abs/1403.4699).

