



HAL
open science

Newton-Like methods for structured monotone inclusions : study of the associated dynamics and algorithms

Boushra Abbas

► **To cite this version:**

Boushra Abbas. Newton-Like methods for structured monotone inclusions : study of the associated dynamics and algorithms. Optimization and Control [math.OC]. Université Montpellier, 2015. English. NNT : 2015MONT250 . tel-02157908

HAL Id: tel-02157908

<https://theses.hal.science/tel-02157908>

Submitted on 17 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de
Docteur

Délivré par L'Université de Montpellier

Préparée au sein de l'école doctorale Informations,
Structures, et Systèmes
Et de l'unité de recherche Institut de Mathématiques et de
Modélisation de Montpellier

Spécialité : Mathématiques appliquées et Optimisation

Présentée par **Boushra ABBAS**

Méthode de Newton régularisée pour
les inclusions monotones structurées.
Etude des dynamiques et algorithmes
associés.

Soutenue le 20 Novembre 2015 devant le jury composé de

M. Samir ADLY, Prof. Dr.	Université de Limoges	Rapporteur
M. Hedy ATTOUCH, Prof. Dr.	Université de Montpellier	Directeur de thèse
M. Radu Ioan BOT, Prof. Dr.	Université de Vienne	Rapporteur
M. Michel THÉRA, Prof. Dr.	Université de Limoges	Examineur
M. Lionel THIBAUT, Prof. Dr.	Université de Montpellier	Président



A ma Syrie blessée

*Aux coeurs pleins de tendresse et d'amour
ma mère et mon père*

A ma famille

A ceux qui m'ont éclairé mon chemin

Remerciements

Tout d'abord, je voudrais remercier mon directeur de thèse, Professeur Hédya Attouch, pour sa patience, son dévouement et sa disponibilité pendant ces années. Il était le vrai exemple d'un directeur qui m'a soutenu pendant une des périodes les plus difficiles de ma vie, celle de la guerre dans mon pays, la Syrie. J'apprécie beaucoup qu'il m'a appris le sens des responsabilités et du travail bien fait. Je voudrais lui exprimer ma profonde gratitude et ma sincère reconnaissance pour l'effort et l'enthousiasme qu'il a fourni afin que ce travail soit le meilleur possible.

De plus, je suis très reconnaissante envers Messieurs Samir Adly et Radu Ioan Boț d'avoir accepté d'être les rapporteurs de ma thèse. Je voudrais aussi remercier Messieurs Lionel Thibault et Michel Théra de me faire l'honneur de faire partie du jury. Enfin, je voudrais aussi exprimer ma gratitude au Monsieur Benar F. Svaiter l'un des acteurs de mon premier article.

Je remercie l'Université Tichreen en Syrie pour la bourse de recherche qu'elle m'a accordée au cours de mes études en France. Je tiens à remercier particulièrement Monsieur Mohamad Soueykatte.

Mes meilleurs sincères remerciements aux membres du Laboratoire I3M qui m'ont accueilli et m'ont donné les outils nécessaires au développement de cette thèse. En particulier, Monsieur Patrick Redont.

Grand merci à mes amis qui ont été ma deuxième famille en France, pour leur soutien et leur présence dans ma vie. Ils ont ajouté un vrai sens à mes jours, en partageant le pur bonheur. Vraiment, c'était inoubliables ces moments qu'on est passé ensemble.

Finalement, je ne trouverai jamais les mots pour remercier ma famille. Mes parents, Rajab et Siham, qui m'ont soutenue et encouragée à poursuivre la recherche scientifique. Ainsi que mes frères et sœurs ; Bashar, Ali, Balsam et Sarab. Cette adorable famille m'a donné l'amour et la force, malgré la tristesse de la guerre et toutes les difficultés et les blessures dans mon pays Syrie.

Résumé

Méthode de Newton régularisée pour les inclusions monotones structurées. Etude des dynamiques et algorithmes associés.

Cette thèse est consacrée à la recherche des zéros d'un opérateur maximal monotone structuré, à l'aide de systèmes dynamiques dissipatifs continus et discrets. Les solutions sont obtenues comme limites des trajectoires lorsque le temps t tend vers l'infini. On s'intéressera principalement aux dynamiques obtenues par régularisation de type Levenberg-Marquardt de la méthode de Newton. On décrira aussi les approches basées sur des dynamiques voisines.

Dans un cadre Hilbertien, on s'intéresse à la recherche des zéros de l'opérateur maximal monotone structuré $M = A + B$, où A est un opérateur maximal monotone général et B est un opérateur monotone Lipschitzien. Nous introduisons des dynamiques continues et discrètes de type Newton régularisé faisant intervenir d'une façon séparée les résolvantes de l'opérateur A (implicites), et des évaluations de B (explicites). A l'aide de la représentation de Minty de l'opérateur A comme une variété Lipschitzienne, nous reformulons ces dynamiques sous une forme relevant du théorème de Cauchy-Lipschitz. Nous nous intéressons au cas particulier où A est le sous différentiel d'une fonction convexe, semi-continue inférieurement, et propre, et B est le gradient d'une fonction convexe, différentiable. Nous étudions le comportement asymptotique des trajectoires. Lorsque le terme de régularisation ne tend pas trop vite vers zéro, et en s'appuyant sur une analyse asymptotique de type Lyapunov, nous montrons la convergence des trajectoires. Par ailleurs, nous montrons la dépendance Lipschitzienne des trajectoires par rapport au terme de régularisation.

Puis nous élargissons notre étude en considérant différentes classes de systèmes dynamiques visant à résoudre les inclusions monotones gouvernées par un opérateur maximal monotone structuré $M = \partial\Phi + B$, où $\partial\Phi$ désigne le sous différentiel d'une fonction convexe, semi-continue inférieurement, et propre, et B est un opérateur monotone cocoercif. En s'appuyant sur une analyse asymptotique de type Lyapunov, nous étudions le comportement asymptotique des trajectoires de ces systèmes. La discrétisation temporelle de ces dynamiques fournit des algorithmes forward-backward (certains nouveaux).

Enfin, nous nous intéressons à l'étude du comportement asymptotique des trajectoires de systèmes dynamiques de type Newton régularisé, dans lesquels on introduit un terme supplémentaire de viscosité évanescence de type Tikhonov. On obtient ainsi la sélection asymptotique d'une solution de norme minimale.

Mots-clés : Inclusions monotones structurées, méthode de Newton, régularisation de type Levenberg-Marquardt, sous-différentiel, convergence asymptotique, analyse de Lyapunov, opérateurs cocoercifs, méthode proximal-gradient, algorithme forward-backward, régularisation de Tikhonov.

Abstract

Newton-Like methods for structured monotone inclusions. Study of the associated dynamics and algorithms.

This thesis is devoted to finding zeroes of structured maximal monotone operators, by using discrete and continuous dissipative dynamical systems. The solutions are obtained as the limits of trajectories when the time t tends towards infinity. We pay special attention to the dynamics that are obtained by Levenberg-Marquardt regularization of Newton's method. We also revisit the approaches based on some related dynamical systems.

In a Hilbert framework, we are interested in finding zeroes of a structured maximal monotone operator $M = A + B$, where A is a general maximal monotone operator, and B is monotone and locally Lipschitz continuous. We introduce discrete and continuous dynamical systems which are linked to Newton's method. They involve separately B and the resolvents of A , and are designed to splitting methods. Based on the Minty representation of A as a Lipschitz manifold, we show that these dynamics can be formulated as differential systems, which are relevant to the Cauchy–Lipschitz theorem. We focus on the particular case where A is the subdifferential of a convex lower semicontinuous proper function, and B is the gradient of a convex, continuously differentiable function. We study the asymptotic behavior of trajectories. When the regularization parameter does not tend to zero too rapidly, and by using Lyapunov asymptotic analysis, we show the convergence of trajectories. Besides, we show the Lipschitz continuous dependence of the solution with respect to the regularization term.

Then we extend our study by considering various classes of dynamical systems which aim at solving inclusions governed by structured monotone operators $M = \partial\Phi + B$, where $\partial\Phi$ is the subdifferential of a convex lower semicontinuous function, and B is a monotone cocoercive operator. By a Lyapunov analysis, we show the convergence properties of the orbits of these systems. The time discretization of these dynamics gives various forward–backward splitting methods (some new).

Finally, we focus on the study of the asymptotic behavior of trajectories of the regularized Newton dynamics, in which we introduce an additional vanishing Tikhonov-like viscosity term. We thus obtain the asymptotic selection of the solution of minimal norm.

Key words : Structured monotone inclusions, Newton method, Levenberg–Marquardt regularization, subdifferential operators, asymptotic convergence, Lyapunov analysis, cocoercive operators, proximal-gradient method, forward–backward algorithms, Tikhonov regularization.

Table des matières

1	Introduction	1
1.1	Présentation générale et objectifs	1
1.2	Organisation de la thèse	5
2	Méthode de Newton régularisée et algorithmes forward-backward pour les inclusions monotones structurées dans les espaces de Hilbert	9
2.1	Description et résultats principaux :	9
2.2	Article en anglais	13
2.2.1	Introduction.	13
2.2.2	Preliminaries. Theoretical background	14
2.2.3	Existence and uniqueness of global solutions.	16
2.2.4	The convex subdifferential case	23
2.2.5	Convergence analysis	26
2.2.6	Gradient-projection and forward-backward methods	35
2.2.7	Conclusions	38
3	Stabilité de la méthode de Newton régularisée avec deux potentiels	41
3.1	Description et résultats principaux :	41
3.2	Article en anglais	43
3.2.1	Introduction	43
3.2.2	Stability results	50
3.2.3	Bounded Variation Regularization Coefficient $\lambda(\cdot)$	53
3.2.4	Perspective	57
4	Systèmes dynamiques et algorithmes forward-backward associés à la somme d'un sous-différentiel convexe et d'un opérateur monotone cocoercif	60
4.1	Description et résultats principaux :	60
4.2	Article en anglais	64
4.2.1	Introduction	65
4.2.2	The continuous regularized Newton-like dynamic	67
4.2.3	Backward-forward algorithms associated with the regularized Newton dynamic	78
4.2.4	Semigroup generated by $-(\partial\Phi + B)$, and FB algorithms	83
4.2.5	Proximal-gradient dynamics and relaxed FB algorithms	87
4.2.6	Perspective	92
4.3	Numerical Illustrations	95
4.3.1	Introduction	95

4.3.2	Constrained least squares problems :	96
4.3.3	Perspective	99
5	Méthode de viscosité et sélection asymptotique d'équilibre pour la dynamique de Newton régularisée	100
5.1	Description et résultats principaux :	100
5.2	Article en anglais	104
5.2.1	Introduction	104
5.2.2	Existence and Uniqueness of Global Solutions	106
5.2.3	Asymptotic analysis and convergence properties	110
5.2.4	Other viscosity selection principles	120
5.2.5	Perspective	122
6	Conclusion et perspectives	125

Notations et Glossaire

Les notations suivantes seront utilisées dans toute la thèse. De plus, nous rappelons certaines définitions de base en analyse convexe.

Notations générales

- H : Espace de Hilbert réel.
- $\langle \cdot, \cdot \rangle$: Produit scalaire de l'espace H .
- 2^H : Ensemble des parties de H .
- \rightarrow : Convergence forte.
- \rightharpoonup : Convergence faible.

Soit C un sous ensemble non vide de H .

- $\delta_C : x \mapsto \begin{cases} 0, & \text{si } x \in C; \\ +\infty, & \text{si } x \notin C \end{cases}$: Fonction indicatrice de C .
- $d_C : x \mapsto \inf_{y \in C} \|x - y\|$: Fonction distance à C associée à la norme $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$.
- $\sigma_C : x \mapsto \sup_{y \in C} \langle x, y \rangle$: Fonction d'appui de C .
- P_C : Projecteur sur le sous ensemble convexe fermé non vide C de H .
- $N_C : x \mapsto \begin{cases} \{u \in H; (\forall y \in C) \langle u, y - x \rangle \leq 0\} & \text{si } x \in C \\ \emptyset & \text{sinon} \end{cases}$: Opérateur cône normal à C .

Notations et définitions relatives à un opérateur multivoque $A : H \rightarrow 2^H$

- $\text{dom}A = \{x \in H; Ax \neq \emptyset\}$: Domaine de A .
- $\text{gr}A = \{(x, u) \in H \times H; u \in Ax\}$: Graphe de A .
- $A^{-1} : H \rightrightarrows H : u \mapsto \{x \in H, u \in Ax\}$: Inverse de A .
- $\text{Fix}A = \{x \in H; x \in Ax\}$: Points fixes de A .
- $\text{zer}A = \{x \in H; 0 \in Ax\}$: Zéros de A .
- $\text{ran}A = \{u \in H; (\exists x \in H) u \in Ax\}$: Image 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 $\lambda > 0$.
- A est monotone :

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

- A est maximal monotone : A est monotone, et il n'existe pas d'opérateur monotone prolongeant strictement A , au sens de l'inclusion des graphes.
- A est γ -fortement monotone :

$$(\forall (x, u) \in \text{gr}A) (\forall (y, v) \in \text{gr}A) \quad \langle x - y, u - v \rangle \geq \gamma \|x - y\|^2.$$

Quelque définitions relatives à un opérateur univoque $T : H \longrightarrow H$

- L'ensemble des points fixes de T :

$$\text{Fix}T = \{x \in H; Tx = x\}.$$

- T est Lipschitzien de constante $L \in]0, +\infty[$ (ou T est L -Lipschitzien)

$$(\forall (x, y) \in H \times H) \|Tx - Ty\| \leq L \|x - y\|.$$

- T est une contraction :

$$(\forall (x, y) \in H \times H) \|Tx - Ty\| \leq \|x - y\|.$$

- T est β -cocoercif : $\exists \beta > 0$ tel que

$$(\forall (x, y) \in H \times H) \langle Tx - Ty, x - y \rangle \geq \beta \|Tx - Ty\|^2.$$

- T est demi fermé :

$$(\forall (x_n)_{n \geq 0} \subset H) (x_n \rightharpoonup x \text{ et } Tx_n \rightarrow y) \implies y = Tx$$

Notions relatives à une fonction convexe, semi-continue inférieurement, propre

$\varphi : H \rightarrow \mathbb{R} \cup \{+\infty\}$

- Domaine de φ

$$\text{dom}\varphi = \{x \in H; \varphi(x) < +\infty\}.$$

- Ensemble des minimiseurs de φ

$$\arg \min_H \varphi.$$

- Conjuguée de φ

$$\varphi^* : u \mapsto \sup_{x \in H} (\langle x, u \rangle - \varphi(x)).$$

- Enveloppe de Moreau d'indice $\lambda \in]0, +\infty[$ de φ

$$\varphi_\lambda : x \mapsto \inf_{y \in H} \left(\varphi(y) + \frac{1}{2\lambda} \|x - y\|^2 \right).$$

- Le sous différentiel de φ

$$\partial\varphi(x) = \{u \in H : \varphi(y) \geq \varphi(x) + \langle u, y - x \rangle \forall y \in H\}$$

- L'opérateur proximal de φ

$$\text{prox}_\varphi : H \rightarrow H : x \mapsto \arg \min_{y \in H} \left(\varphi(y) + \frac{1}{2} \|x - y\|^2 \right).$$

Chapitre 1

Introduction

1.1 Présentation générale et objectifs

Cette thèse est consacrée à la recherche des zéros d'un opérateur maximal monotone structuré, à l'aide de systèmes dynamiques dissipatifs continus et discrets. Les solutions sont obtenues comme limites des trajectoires lorsque le temps t tend vers l'infini. Les dynamiques considérées sont obtenues par régularisation de la méthode de Newton. Nous comparons les résultats avec ceux provenant d'autres dynamiques.

Le cadre mathématique est le suivant. Soit H un espace de Hilbert réel, $A : H \rightrightarrows H$ un opérateur maximal monotone général, et $B : H \rightarrow H$ un opérateur monotone Lipschitzien. On s'intéresse à la recherche, et au calcul par des méthodes algorithmiques, des zéros de l'opérateur maximal monotone structuré $M = A + B$, i.e.,

$$\text{Trouver } x \in H \text{ tel que } A(x) + B(x) \ni 0. \quad (1.1.1)$$

Dans tout le mémoire, on supposera acquis que l'ensemble des solutions de (1.1.1) est non vide, la question de l'existence de solutions relevant d'autres théories, bien établies, qui ne sont pas l'objet de notre étude. La formulation ci-dessus contient comme cas particuliers les problèmes de minimisation convexe, de point selle convexe-concave et de point fixe de contractions. En outre, elle permet de combiner entre eux ces différents types de problèmes. La résolution du problème (1.1.1) par des méthodes d'éclatement (par exemple "forward-backward") trouve des applications directes dans de nombreux domaines, en traitement du signal, théorie des jeux, transport optimal, équations aux dérivées partielles, et contrôle optimal.

L'objet principal de cette thèse est l'introduction et l'analyse de la convergence de (nouveaux) systèmes dynamiques continus, dont les versions discrètes sont liées à la résolution de l'inclusion monotone structurée (1.1.1) par la méthode de Newton.

En général, la méthode de Newton exige beaucoup de régularité, à la fois dans sa version discrète et continue. Dans le cas continu, pour contourner le caractère mal posé de la méthode de Newton, Attouch et Svaiter introduisent dans [7] une méthode de Levenberg-Marquardt régularisée, qui a les propriétés intéressantes suivantes : pour un opérateur maximal monotone général M , qui agit sur un espace de Hilbert réel H , cette dynamique est bien posée, stable, et a un bon comportement asymptotique. De plus, ils ont montré la convergence faible des trajectoires, lorsque le terme de régularisation $\lambda(\cdot)$ ne tend pas trop vite vers zéro. Le problème de Cauchy pour ce système de Newton régularisé s'écrit sous la forme du système en (x, v)

suivant

$$v(t) \in M(x(t)), \quad (1.1.2)$$

$$\lambda(t) \dot{x}(t) + \dot{v}(t) + v(t) = 0 \quad (1.1.3)$$

$$x(0) = x_0, v(0) = v_0, v_0 \in M(x_0). \quad (1.1.4)$$

En s'appuyant sur la représentation de Minty de l'opérateur maximal monotone M comme une variété Lipschitzienne, et en prenant comme nouvelle fonction inconnue $z : [0, +\infty[\rightarrow H$ définie par $z(t) = x(t) + \mu(t)v(t)$ où $(\mu(t) = \frac{1}{\lambda(t)})$, ils ont reformulé de façon équivalente le système précédent sous la forme

$$x(t) = J_{\mu(t)}^M(z(t)) \quad (1.1.5)$$

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t)) M_{\mu(t)} z(t) = 0 \quad (1.1.6)$$

$$z(0) = x_0 + \mu(0)v_0. \quad (1.1.7)$$

Cette formulation joue un rôle important dans notre approche. Insistons sur le fait que les opérateurs $J_{\mu}^M(z) := (I + \mu M)^{-1} : H \rightarrow H$, et $M_{\mu} = \frac{1}{\mu} (I - J_{\mu}^M) : H \rightarrow H$ sont partout définis et lipschitziens. Le système (1.1.5)-(1.1.7) peut alors être traité par le théorème de Cauchy Lipschitz, d'où l'existence et l'unicité d'une solution globale forte du problème de Cauchy (1.1.5)-(1.1.7).

Ce travail est prolongé dans [6] par l'étude de la régularisation en boucle fermée,

$$v(t) \in M(x(t)), \quad (1.1.8)$$

$$G(x(t), v(t)) \dot{x}(t) + \dot{v}(t) + v(t) = 0, \quad (1.1.9)$$

où le coefficient de régularisation $G : H \times H \rightarrow [0, +\infty[$ est, pour $v \neq 0$, strictement positif. Le passage de la régularisation en boucle ouverte à la régularisation en boucle fermée se fait à l'aide d'un argument de point fixe. A cet effet, dans la dynamique en boucle ouverte, on établit la dépendance lipschitzienne de la solution par rapport à λ .

Il y a un lien étroit entre le comportement asymptotique des systèmes dynamiques continus et de leurs versions discrètes (algorithmes). C'est un domaine actif de recherche [18].

Cependant, l'approche précédente, malgré son grand intérêt théorique, soulève une difficulté notable pour ses applications pratiques et numériques. En effet, les dynamiques décrites ci-dessus, et étudiées dans [7, 6], utilisent les résolvantes de l'opérateur maximal monotone M , qui sont, en général, difficiles à calculer.

Dans notre travail, nous étendons les résultats précédents, et définissons un cadre pratique d'utilisation, en considérant le cas où M est un opérateur maximal monotone structuré $M = A + B$, où A est un opérateur maximal monotone général, et B est monotone Lipschitzien. Comme nous le verrons, ce cadre est bien adapté aux applications.

Nous introduisons des dynamiques continues et discrètes de type Newton régularisé faisant intervenir d'une façon séparée les résolvantes de l'opérateur A (implicites), et des évaluations de B (explicites) :

$$v(t) \in A(x(t)), \quad (1.1.10)$$

$$\lambda(t) \dot{x}(t) + \dot{v}(t) + v(t) + B(x(t)) = 0, \quad (1.1.11)$$

La discrétisation en temps de ces dynamiques débouche sur de nouveaux algorithmes combinant les méthodes forward-backward et la méthode de Newton.

En particulier, nous examinons le cas où $A = \partial\Phi$ est le sous différentiel d'une fonction convexe, semi-continue inférieurement, et propre $\Phi : H \rightarrow \mathbb{R} \cup \{+\infty\}$, et $B = \nabla\Psi$ est le gradient d'une fonction $\Psi : H \rightarrow \mathbb{R}$ convexe et différentiable.

Le problème de minimisation convexe structuré correspondant s'écrit

$$\min \{ \Phi(x) + \Psi(x) : x \in H \}. \quad (1.1.12)$$

Lorsque le terme de régularisation $\lambda(\cdot)$ ne tend pas trop vite vers zéro (la taille critique est e^{-t}), et en s'appuyant sur une analyse asymptotique de type Lyapunov, nous étudions le comportement asymptotique des trajectoires.

Par ailleurs, nous montrons la dépendance Lipschitzienne des trajectoires par rapport au terme de régularisation. Ceci permet de prendre le coefficient λ à variation bornée. A cet effet, pour étudier l'existence et l'unicité, nous utilisons une méthode de régularisation par convolution, qui permet de se ramener à un cas simple, puis on passe à la limite dans les équations.

Puis nous portons une attention particulière au cas où A est le sous différentiel d'une fonction convexe, semi-continue inférieurement, et propre Φ , et B est opérateur monotone β -cocoercif. Le problème s'écrit

$$\text{Trouver } x \in H \text{ tel que } \partial\Phi(x) + B(x) \ni 0. \quad (1.1.13)$$

Nous nous intéressons à résoudre le problème (1.1.13) par la méthode de Newton régularisée décrite dans les lignes précédentes et développée dans [4] et dans [7]

$$v(t) \in \partial\Phi(x(t)), \quad (1.1.14)$$

$$\lambda\dot{x}(t) + \dot{v}(t) + v(t) + B(x(t)) = 0, \quad (1.1.15)$$

en supposant que le terme de régularisation λ est une constante positive.

La discrétisation en temps débouche sur un nouvel algorithme backward-forward (implicite-explicite)

$$(BF) \quad \begin{cases} x_k = \text{prox}_{\mu\Phi}(y_k) \\ y_{k+1} = (1-h)y_k + h(x_k - \mu B(x_k)) \end{cases} \quad (1.1.16)$$

qui est étroitement lié à l'algorithme forward-backward relaxé, cf. [11]. Les propriétés asymptotiques du système dynamique continu (et ses fonctions de Lyapunov) nous donnent un guide pour l'étude des propriétés asymptotiques de l'algorithme.

Nous considérons un autre système dynamique attaché à la résolution de (1.1.13), qui est le semi-groupe de contractions généré par $-(\partial\varphi + B)$

$$\dot{x}(t) + \partial\Phi(x(t)) + B(x(t)) \ni 0. \quad (1.1.17)$$

Cette dynamique relève de la théorie générale des semi-groupes de contractions dans les Hilbert, dont les générateurs sont précisément les opérateurs maximaux monotones. La théorie générale correspondante, cf. [13], nous donne l'existence et l'unicité d'une solution globale forte pour le problème de Cauchy. Le comportement asymptotique des trajectoires, étudié par Baillon-Brézis dans [10], montre que chaque trajectoire de (1.1.17) converge faiblement de façon ergodique, vers un zéro de l'opérateur $\partial\Phi + B$.

Bruck dans [15] a étudié les propriétés de convergence faible (au lieu de convergence ergodique faible). La théorie de la convergence de Bruck est applicable séparément à $\partial\Phi$ et B , qui sont des opérateurs demi-positifs. Mais on ne sait pas si la somme de ces deux opérateurs reste un

opérateur demi-positif, et donc la théorie de Bruck n'est pas directement applicable à notre cadre.

Une nouveauté de notre travail est que nous montrons que toute trajectoire du système (1.1.17) converge faiblement vers un zéro de $\partial\Phi + B$.

Nous considérons aussi un autre système dynamique qui est associé à la reformulation de (1.1.13) comme un problème de point fixe

$$\dot{x}(t) + x(t) - \text{prox}_{\mu\Phi}(x(t) - \mu B(x(t))) = 0. \quad (1.1.18)$$

Les propriétés de la convergence du système (1.1.18) ont été étudiées par Antipin [5] et Bolte [12], dans le cas particulier où Φ est la fonction indicatrice d'un ensemble convexe fermé C , et B désigne le gradient d'une fonction convexe différentiable Ψ . Dans ce cas le système (1.1.18) se ramène à la méthode de gradient-projeté suivante :

$$\dot{x}(t) + x(t) - \text{proj}_C(x(t) - \mu\nabla\Psi(x(t))) = 0. \quad (1.1.19)$$

Dans tous les systèmes dynamiques précédents, pour certaines considérations physiques, économiques et autres, une solution stationnaire particulière est plus intéressante que d'autres. Lorsqu'on a la convergence globale des trajectoires, une méthode classique pour atteindre un point d'équilibre particulier, consiste à introduire un terme évanescent dans le système, ne changeant pas l'ensemble des équilibres mais favorisant la convergence vers le type d'équilibre désiré. Cette méthode rentre dans le cadre général des solutions de viscosité, et donne lieu à un principe de minimisation hiérarchique asymptotique. Un exemple classique est l'introduction d'un terme de régularisation de type Tikhonov $\epsilon(\cdot)x(\cdot)$ dans la dynamique pour atteindre asymptotiquement un équilibre de norme minimale.

Dans l'étude de notre système, par simplicité, nous nous limitons au cas d'un seul potentiel et considérons la méthode de Newton régularisée (à la Tikhonov) suivante :

$$v(t) \in \partial\Phi(x(t)), \quad (1.1.20)$$

$$\lambda\dot{x}(t) + \dot{v}(t) + v(t) + \epsilon(t)x(t) = 0. \quad (1.1.21)$$

Sous l'hypothèse que $\epsilon(\cdot)$ tend vers zéro de façon modérée lorsque t tend vers $+\infty$, nous montrons que le terme $\epsilon(\cdot)x(\cdot)$ agit comme une régularisation de Tikhonov. Nous utilisons la représentation de Minty de l'opérateur maximal monotone $\partial\Phi$, et le fait que $\text{prox}_{\mu\Phi} = \nabla\Psi$ avec $\Psi(y) = \mu(\Phi^*)_{\frac{1}{\mu}}\left(\frac{1}{\mu}y\right)$, où Φ^* est la conjuguée de Fenchel. Le système (1.1.20)-(1.1.21) se reformule alors comme suit :

$$x(t) = \text{prox}_{\mu\Phi}(y(t)) \quad (1.1.22)$$

$$\dot{y}(t) + \nabla\Theta(y(t)) + \epsilon(t)\nabla\Psi(y(t)) = 0, \quad (1.1.23)$$

où $\Theta(\cdot) := \mu\Phi_{\mu}(\cdot)$. C'est un cas particulier du système dynamique multi-échelle, qui a été étudié en détail dans [9]

$$(\text{MAG})_{\epsilon} \quad \dot{y}(t) + \partial\Theta(y(t)) + \epsilon(t)\partial\Psi(y(t)) \ni 0.$$

Si $C = \text{argmin}\Phi \neq \emptyset$, $\int_0^{+\infty} \epsilon(t) dt = +\infty$, et sous une hypothèse de croissance modérée de $\epsilon(\cdot)$, nous montrons que chaque trajectoire converge faiblement vers l'élément de norme minimale

de l'ensemble convexe fermé C .

Puis nous considérons les propriétés de sélection asymptotiques liées à d'autres types de perturbation :

$$v(t) \in \partial\Phi(x(t)) \tag{1.1.24}$$

$$\lambda\dot{x}(t) + \dot{v}(t) + v(t) + \varepsilon(t)\partial g(x(t)) \ni 0, \tag{1.1.25}$$

En utilisant la représentation de Minty de $\partial\Phi$, et en prenant $g(\cdot) = \frac{1}{2\mu}\|\cdot\|^2 + \Phi(\cdot)$, nous reformulons le système (1.1.24)-(1.1.25) sous la forme suivante :

$$\dot{y}(t) + \nabla\Theta(y(t)) + \varepsilon(t)y(t) = 0 \tag{1.1.26}$$

Le premier résultat de convergence vers l'élément de norme minimale pour ce type de système (1.1.26) a été démontré dans [17] (basé sur [14]) : en supposant que $\varepsilon(\cdot)$ est décroissante, et ne tend pas trop vite vers zéro, alors toutes les trajectoires convergent fortement vers l'élément de norme minimale dans C . Puis Cominetti-Peypouquet-Sorin ont montré dans [16] que le résultat de convergence reste vrai, sans supposer que $\varepsilon(\cdot)$ est décroissante. Le système (1.1.24)-(1.1.25) avec le choix particulier de $g(\cdot) = \frac{1}{2\mu}\|\cdot\|^2 + \Phi(\cdot)$ devient :

$$v(t) \in \partial\Phi(x(t)) \tag{1.1.27}$$

$$\lambda\dot{x}(t) + \dot{v}(t) + (1 + \mu\varepsilon(t))v(t) + \varepsilon(t)x(t) = 0. \tag{1.1.28}$$

Nous montrons que toutes les trajectoires convergent fortement vers l'élément de norme minimale dans C .

1.2 Organisation de la thèse

Dans le chapitre 2, qui s'appuie sur l'article [4], nous introduisons des dynamiques continues et discrètes de type Newton régularisé qui visent à résoudre le problème (1.1.1). A l'aide de la représentation de Minty, nous montrons l'existence et l'unicité d'une solution globale forte de notre problème. Puis nous nous intéressons au cas particulier où A est le sous différentiel d'une fonction convexe, semi-continue inférieurement, et propre, et B est le gradient d'une fonction convexe, différentiable, B étant supposé seulement localement Lipschitzien (hypothèse un peu plus faible que dans le cas général, où B est supposé monotone globalement Lipschitzien). Dans ce cadre, bien adapté aux problèmes de minimisation structurés, nous montrons l'existence et l'unicité d'une solution globale forte de notre problème d'évolution. Puis, en s'appuyant sur une analyse asymptotique de type Lyapunov, et en utilisant le lemme d'Opial, nous étudions le comportement asymptotique des trajectoires, lorsque le terme de régularisation ne tend pas trop vite vers zéro. Finalement, nous introduisons de nouveaux algorithmes forward-backward (explicite-implicite) et gradient-projeté, qui sont obtenus par la discrétisation en temps des systèmes dynamiques continus étudiés.

Le chapitre 3, qui s'appuie sur l'article [2], porte sur l'étude de questions de stabilité pour le problème précédent, dans le cas où A est le sous différentiel d'une fonction convexe, semi-continue inférieurement, et propre, et B est le gradient d'une fonction convexe, différentiable, B étant supposé seulement localement Lipschitzien. Dans ce cas, l'équation d'équilibre est le problème de minimisation structuré (1.1.12).

Nous établissons tout d'abord des estimations du type énergie pour les trajectoires. Dans le cas où le coefficient de régularisation $\lambda(\cdot)$ est absolument continu, nous prouvons la dépendance Lipschitzienne des solutions par rapport au coefficient de régularisation. Par un argument de prolongement par continuité, nous pouvons ainsi étendre notre étude au cas où le coefficient $\lambda(\cdot)$ est à variation bornée.

Dans le chapitre 4, qui correspond à l'article [3], nous nous intéressons au cas où A est le sous différentiel d'une fonction convexe, semi-continue inférieurement, et propre et B est un opérateur monotone cocoercif. Nous étudions diverses dynamiques visant à la résolution du problème d'équilibre correspondant (1.1.13). Nous commençons par l'étude des dynamiques continues et discrètes de type Newton régularisé. En supposant que le terme de régularisation est une constante positive, nous montrons l'existence et l'unicité d'une solution globale forte, et le fait que toute trajectoire de ce système converge faiblement vers un équilibre. La discrétisation en temps du système continu fournit un nouvel algorithme backward-forward (implicite-explicite) qui génère des suites qui convergent faiblement vers des équilibres. Puis, nous considérons un autre système dynamique attaché à la résolution de (1.1.13), qui est le semi-groupe de contractions généré par $-(\partial\varphi + B)$, où la discrétisation en temps donne l'algorithme forward-backward classique. Finalement, nous considérons le système dynamique proximal-gradient associé à la reformulation de (1.1.13) comme un problème de point fixe, et nous étudions ses propriétés de convergence. Dans la section 4.3, on présente une illustration numérique dans un cas modèle simple, permettant de comparer le comportement asymptotique des algorithmes forward-backward respectivement associés à la dynamique Newton régularisé, au semi-groupe de contractions généré par $-(\partial\varphi + B)$, et au système dynamique proximal-gradient.

Le chapitre 5 s'appuie sur l'article [1], où nous nous intéressons à l'étude du comportement asymptotique des trajectoires du système avec viscosité évanescence de type Tikhonov (1.1.20)-(1.1.21). Nous montrons que toute trajectoire de ce système converge faiblement vers l'équilibre de norme minimale. Puis nous montrons que dans le cas où Φ est une fonction convexe différentiable, dont le gradient est Lipschitzien, toutes les trajectoire du système (1.1.20)-(1.1.21) convergent fortement vers cet équilibre. Finalement, nous considérons les propriétés de sélection asymptotiques liées à d'autres types de méthode de viscosité.

Bibliographie

- [1] B. Abbas, An asymptotic viscosity selection result for the regularized Newton dynamic, arXiv :1504.07793v1 [math.OC] 29 Apr 2015.
- [2] B. Abbas, Stability of a regularized Newton method with two potentials, arXiv :1505.05418v1 [math.OC] 20 May 2015
- [3] B. Abbas, H. Attouch, Dynamical systems and forward-backward algorithms associated with the sum of a convex subdifferential and a monotone cocoercive operator, Optimization, (2014) <http://dx.doi.org/10.1080/02331934.2014.971412>.
- [4] B. Abbas, H. Attouch, B. F. Svaiter, Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces, J. Optim. Theory Appl., 161 (2014), No. 2, pp. 331-360.
- [5] A. Antipin, Minimization of convex functions on convex sets by means of differential equations, Differ. Equ. 2010 ;30 :3246–3270.
- [6] H. Attouch, P. Redont, B.F Svaiter, Global convergence of a closed-loop regularized Newton method for solving monotone inclusions in Hilbert spaces, J. Optim. Theory Appl. 157(3), 624–650 (2013)
- [7] H. Attouch, B.F. Svaiter, A continuous dynamical Newton-like approach to solving monotone inclusions, SIAM J. Control Optim. 49, 574–598 (2011).
- [8] H. Attouch, J. Peypouquet, P. Redont, Backward-forward algorithms for structured monotone inclusions in Hilbert spaces, to appear.
- [9] H. Attouch, M.-O. Czarnecki, Asymptotic behavior of coupled dynamical systems with multiscale aspects, Journal of Differential Equations, 248 (2010), pp. 1315-1344.
- [10] J.-B. Baillon, H. Brézis, Une remarque sur le comportement asymptotique des semi-groupes non linéaires [A remark on the asymptotic behavior of nonlinear semigroups], Houston J. Math. 1976 ;2 :5–7.
- [11] H.H. Bauschke, P.-L. Combettes, Convex analysis and monotone operator theory in Hilbert spaces. New York (NY) : Springer ; 2011.
- [12] J. Bolte, Continuous gradient projection method in Hilbert spaces, J. Optim. Theory Appl. 2003 ;119 :235–259.
- [13] H. Brézis, Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert [Maximal monotone operators and semi-groups of contractions in Hilbert spaces]. New York (NY) : North-Holland/Elsevier ; 1973.
- [14] F.E. Browder, Nonlinear operators and nonlinear equations of evolution in Banach spaces, Proc. Sympos. Pure Math., vol. 18 (part 2), Amer. Math. Soc., Providence, RI, 1976.

- [15] R. Bruck, Asymptotic convergence of nonlinear contraction semigroups in Hilbert spaces. *J. Funct. Anal.* 1975;18 :15–26.
- [16] R. Cominetti, J. Peypouquet, S. Sorin, Strong asymptotic convergence of evolution equations governed by maximal monotone operators with Tikhonov regularization, *J. Differential Equations*, 245 (2008), pp. 3753–3763.
- [17] S. Reich, Nonlinear evolution equations and nonlinear ergodic theorems, *Nonlinear Anal.* 1 (1976) 319–330.
- [18] J. Peypouquet, S. Sorin, Evolution equations for maximal monotone operators : asymptotic analysis in continuous and discrete time, *J. Convex Anal.* 17, 1113–1163 (2010).

Chapitre 2

Méthode de Newton régularisée et algorithmes forward-backward pour les inclusions monotones structurées dans les espaces de Hilbert

2.1 Description et résultats principaux :

Dans ce chapitre, nous étudions des systèmes dynamiques visant à résoudre les inclusions monotones gouvernées par un opérateur maximal monotone structuré $M = A + B$. On se place dans un espace de Hilbert H , A est un opérateur maximal monotone général, et B est un opérateur monotone univoque, qui est supposé Lipschitzien sur les bornés.

Problème 2.1.1 *Soit H un espace de Hilbert réel.*

$$\text{Trouver } x \in H \text{ tel que } A(x) + B(x) \ni 0.$$

Les solutions de ce problème seront obtenues comme limites des trajectoires de systèmes dynamiques dissipatifs continus et discrets (algorithmes) de type Newton régularisé.

La dynamique continue, associée à la méthode de Newton, étant mal posée, nous nous intéressons à sa régularisation par une méthode de type Levenberg-Marquardt, le terme de régularisation étant vu comme un contrôle.

En utilisant la représentation de Minty de l'opérateur maximal monotone A (variété lipschitzienne), nous montrons que la dynamique peut se formuler comme une ODE relevant du théorème de Cauchy-Lipschitz. En s'appuyant sur une analyse asymptotique de type Lyapunov, nous montrons que les trajectoires convergent faiblement vers des zéros de l'opérateur $M = A + B$.

La discrétisation temporelle de ces dynamiques fournit de nouveaux algorithmes combinant la méthode de Newton et les méthodes forward-backward.

Comme guide de notre travail, nous utilisons la méthode de Newton régularisée introduite par Attouch-Svaiter (2011).

Pour résoudre le problème (2.1.1), nous considérons le problème de Cauchy suivant :

$$v(t) \in A(x(t)), \quad (2.1.1)$$

$$\lambda(t) \dot{x}(t) + \dot{v}(t) + v(t) + B(x(t)) = 0, \quad (2.1.2)$$

$$x(0) = x_0, v(0) = v_0, v_0 \in A(x_0). \quad (2.1.3)$$

Dans ce système, le choix du terme de régularisation ($\lambda(\cdot) > 0$) joue un rôle essentiel dans l'analyse de la convergence des trajectoires vers des équilibres, qui sont les zéros de l'opérateur $A+B$. Notons que, dans (2.1.2), les opérateurs A et B jouent des rôles dissymétriques. Lorsque $B = 0$, on obtient une méthode de Newton régularisée par rapport à A , alors que, lorsque $A = 0$, on obtient le semi-groupe de contractions généré par B .

En s'appuyant sur la représentation de Minty de l'opérateur maximal monotone A comme une variété lipschitzienne, et en prenant comme nouvelle fonction inconnue $z : [0, +\infty[\rightarrow H$ définie par $z(t) = x(t) + \mu(t)v(t)$ où ($\mu(t) = \frac{1}{\lambda(t)}$), nous pouvons reformuler de façon équivalente le système (2.1.1-2.1.3) relativement aux nouvelles inconnues (x, z) sous la forme :

$$x(t) = J_{\mu(t)}^A(z(t)) \quad (2.1.4)$$

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t)) A_{\mu(t)} z(t) + \mu(t) B(J_{\mu(t)}^A z(t)) = 0 \quad (2.1.5)$$

$$z(0) = x_0 + \mu(0) v_0. \quad (2.1.6)$$

Cette formulation joue un rôle important dans notre approche. Insistons sur le fait que les opérateurs $J_{\mu}^A(z) := (I + \mu A)^{-1} : H \rightarrow H$, et $A_{\mu} = \frac{1}{\mu} (I - J_{\mu}^A) : H \rightarrow H$ sont partout définis et lipschitziens, ce qui permet d'appliquer le théorème de Cauchy-Lipschitz.

Notre résultat principal concernant l'existence et l'unicité d'une solution globale forte du système (2.1.1-2.1.3) se formule comme suit :

Théorème 2.1.2 *Soit $\lambda : [0, +\infty[\rightarrow]0, +\infty[$ absolument continue sur tout intervalle borné $[0, b]$, $b > 0$. On pose $\mu(t) = \frac{1}{\lambda(t)}$. Soit $A : H \rightrightarrows H$ un opérateur maximal monotone général, et soit $B : H \rightarrow H$ un opérateur maximal monotone Lipschitzien continu sur H .*

Soit $(x_0, v_0) \in H \times H$ et $v_0 \in A(x_0)$. Alors,

(i) il existe une solution unique globale forte $(x(\cdot), v(\cdot)) : [0, +\infty[\rightarrow H \times H$ du problème de Cauchy (2.1.1)-(2.1.3).

(ii) la solution $(x(\cdot), v(\cdot))$ du (2.1.1)-(2.1.3) peut être représentée par, pour tout $t \in [0, +\infty[$

$$\begin{aligned} x(t) &= J_{\mu(t)}^A(z(t)), \\ v(t) &= A_{\mu(t)} z(t), \end{aligned}$$

où $z(\cdot) : [0, +\infty[\rightarrow H$ est la solution unique globale forte du problème de Cauchy (2.1.4)-(2.1.6).

Puis, nous examinons le cas où A est le sous-différentiel d'une fonction convexe, semi-continue inférieurement, et étudions le problème de minimisation convexe structuré suivant :

Problème 2.1.3

Trouver $x \in H$ tel que $\partial\varphi(x) + B(x) \ni 0$.

Dans ce problème :

$A = \partial\varphi$ est le sous différentiel d'une fonction convexe, semi-continue inférieurement, et propre $\varphi : H \longrightarrow \mathbb{R} \cup \{+\infty\}$.

$B = \nabla\psi$ est le gradient d'une fonction $\psi : H \longrightarrow \mathbb{R}$ convexe, continue et différentiable, B étant supposé seulement localement Lipschitzien (hypothèse un peu plus faible que dans le cas général, où B est supposé monotone globalement Lipschitzien).

Par un résultat classique $\partial\varphi$ et $\nabla\psi$ sont deux opérateurs maximaux monotones, et le système (2.1.1)-(2.1.3) se reformule alors comme suit :

$$v(t) \in \partial\varphi(x(t)), \quad (2.1.7)$$

$$\lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) + \nabla\psi(x(t)) = 0, \quad (2.1.8)$$

$$x(0) = x_0, v(0) = v_0, v_0 \in \partial\varphi(x_0). \quad (2.1.9)$$

Nous montrons d'abord que le système (2.1.7)-(2.1.9) génère des trajectoires qui satisfont la propriété de la descente suivante. Cette propriété, ainsi que l'estimation de type énergie des trajectoires vont jouer un rôle clé dans l'établissement du résultat d'existence global.

Proposition 2.1.4 *Soit (x, v) une solution forte du système (2.1.7)-(2.1.9) sur $[0, b]$, $b > 0$. Alors*

(a) $t \longrightarrow \varphi(x(t)) + \psi(x(t))$ est une fonction décroissante.

(b) $\int_0^b \lambda(t) \|\dot{x}(t)\|^2 dt < +\infty$, et par conséquent $\dot{x} \in L^2(0, b; H)$.

Nous montrons le résultat suivant d'existence et d'unicité dans le cas où $B = \nabla\psi$ est localement Lipschitzien.

Théorème 2.1.5 *Soit $\lambda : [0, +\infty[\longrightarrow]0, +\infty[$ absolument continue sur tout intervalle borné $[0, b]$, $b > 0$. Soit $A = \partial\varphi$ le sous différentiel d'une fonction convexe, semi-continue inférieurement, et propre $\varphi : H \longrightarrow \mathbb{R} \cup \{+\infty\}$, et $B = \nabla\psi$ le gradient d'une fonction convexe, continue et différentiable $\psi : H \longrightarrow \mathbb{R}$. Nous supposons que B est localement Lipschitzien. Supposons que $\varphi + \psi$ est minorée sur H . Soit $(x_0, v_0) \in H \times H$ et $v_0 \in \partial\varphi(x_0)$. Alors,*

(i) *il existe une solution unique globale forte $(x(\cdot), v(\cdot)) : [0, +\infty[\longrightarrow H \times H$ du problème de Cauchy (2.1.7)-(2.1.9).*

(ii) *on pose $\mu(t) = \frac{1}{\lambda(t)}$, la solution $(x(\cdot), v(\cdot))$ du système (2.1.7)-(2.1.9) peut être représentée par, pour tout $t \in [0, +\infty[$*

$$x(t) = J_{\mu(t)}^{\partial\varphi}(z(t)),$$

$$v(t) = \nabla\varphi_{\mu(t)}z(t),$$

où $z(\cdot) : [0, +\infty[\longrightarrow H$ est la solution unique globale forte du problème de Cauchy suivant :

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t)) \nabla\varphi_{\mu(t)}z(t) + \mu(t) \nabla\psi\left(J_{\mu(t)}^{\partial\varphi}z(t)\right) = 0 \quad (2.1.10)$$

$$z(0) = x_0 + \mu(0)v_0. \quad (2.1.11)$$

Puis, nous étudions la convergence asymptotique des trajectoires lorsque $t \longrightarrow +\infty$. En s'appuyant sur une analyse asymptotique de type Lyapunov et le lemme d'Opial, nous montrons que, si le terme de régularisation $\lambda(\cdot)$ ne tend pas trop vite vers zéro, les trajectoires convergent faiblement vers un équilibre qui est une solution du problème de minimisation (2.1.3).

Le théorème suivant est notre résultat principal concernant la convergence des trajectoires. Rappelons que notre stratégie est de prendre le coefficient de régularisation $\lambda(\cdot)$ aussi petit que possible pour que, asymptotiquement on soit le plus proche possible de la méthode de Newton. Notre résultat est que $\lambda(t) = e^{-t}$ est la taille critique.

Théorème 2.1.6 *Supposons que λ est décroissante. Alors*

$$\lim_{t \rightarrow +\infty} (\varphi + \psi)(x(t)) = \inf_H (\varphi + \psi).$$

Supposons que $\arg \min (\varphi + \psi) \neq \emptyset$, et

$$\exists \epsilon > 0 : 1 + \frac{\dot{\lambda}(t)}{\lambda(t)} \geq \epsilon, \quad \text{presque pour tout } t > 0$$

Alors,

(i) chaque trajectoire $x(\cdot)$ converge faiblement vers un minimiseur de $\varphi + \psi$.

(ii) $v(\cdot) + \nabla \psi(x(\cdot))$ converge fortement vers zéro.

Finalement, nous nous intéressons à la discrétisation en temps du système précédent, où $A = \partial \varphi$ est le sous différentiel d'une fonction convexe, semi-continue inférieurement, et propre $\varphi : H \rightarrow \mathbb{R} \cup \{+\infty\}$. Dans ce cas, les résolvantes se reformulent de façon équivalente sous la forme proximale due à Moreau. Précisément, pour tout $x \in H$ et $\mu > 0$, la résolvante et l'approximation de Yosida de l'opérateur A d'indice μ sont donnés par :

$$J_\mu^A x = \text{prox}_{\mu\varphi} x = \operatorname{argmin} \left\{ \varphi(y) + \frac{1}{2\mu} \|x - y\|^2 : y \in H \right\}$$

$$A_\mu = \nabla \varphi_\mu, \quad \varphi_\mu(x) = \inf \left\{ \varphi(y) + \frac{1}{2\mu} \|x - y\|^2 : y \in H \right\}$$

Le système dynamique (2.1.10)-(2.1.11) s'écrit de façon équivalente sous la forme du système en (x, z) , qui est étudié dans le théorème 2.1.5, par

$$\begin{cases} x(t) = \text{prox}_{\mu(t)\varphi}(z(t)), \\ \dot{z}(t) + \left(1 - \frac{\dot{\mu}(t)}{\mu(t)}\right) (z(t) - \text{prox}_{\mu(t)\varphi}(z(t))) + \mu(t) \nabla \psi(\text{prox}_{\mu(t)\varphi}(z(t))) = 0. \end{cases} \quad (2.1.12)$$

Dans le cas où μ est une constante positive, on a $\dot{\mu} = 0$, et le système (2.1.12) devient

$$\begin{cases} x(t) = \text{prox}_{\mu(t)\varphi}(z(t)), \\ \dot{z}(t) + z(t) - \text{prox}_{\mu(t)\varphi}(z(t)) + \mu(t) \nabla \psi(\text{prox}_{\mu(t)\varphi}(z(t))) = 0. \end{cases} \quad (2.1.13)$$

La discrétisation en temps du système (2.1.13) débouche sur un nouvel algorithme forward-backward

$$\begin{cases} x_k = \text{prox}_{\mu\varphi}(z_k), \\ z_{k+1} = (1 - h_k)z_k + h_k(x_k - \mu \nabla \psi(x_k)) \end{cases} \quad (2.1.14)$$

L'algorithme (2.1.14) partage les bonnes propriétés de convergence globale de la dynamique continue associée. Une étude détaillée est faite dans le chapitre 4.

2.2 Article en anglais

Newton-like Dynamics and Forward-Backward Methods for Structured Monotone Inclusions in Hilbert Spaces¹

abstract

In a Hilbert space setting we introduce dynamical systems, which are linked to Newton and Levenberg-Marquardt methods. They are intended to solve, by splitting methods, inclusions governed by structured monotone operators $M = A + B$, where A is a general maximal monotone operator, and B is monotone and locally Lipschitz continuous. Based on the Minty representation of A as a Lipschitz manifold, we show that these dynamics can be formulated as differential systems, which are relevant to the Cauchy-Lipschitz theorem, and involve separately B and the resolvents of A . In the convex subdifferential case, by using Lyapunov asymptotic analysis, we prove a descent minimizing property, and weak convergence to equilibria of the trajectories. Time discretization of these dynamics gives algorithms combining Newton's method and forward-backward methods for solving structured monotone inclusions.

2.2.1 Introduction.

Discrete and continuous dynamical systems governed by maximal monotone operators play an important role in optimization, game theory, equilibrium, fixed-point theory, partial differential equations, mechanics, among others. The study of the relationship between continuous time and discrete time models is an active area of research, see [1] for a recent review on the subject. Viewing an iterative algorithm as a discrete version of a continuous dynamical system sheds new light on the properties of this algorithm, offers Liapunov functions that are useful for the study of its asymptotic behavior, and suggests new neighboring classes of algorithms. As a classical situation, Cauchy's steepest descent algorithm can be viewed as an explicit (forward) discretization of the minus gradient flow, while the implicit (backward) discretization gives the proximal algorithm. When considering gradient flows associated to nonsmooth convex lower semicontinuous potentials, one obtains differential inclusions governed by subdifferential operators, which are maximal monotone.

Newton's method, however, requires a lot of regularity both in its discrete and continuous version. In the continuous case, to circumvent the ill-posedness of Newton's method, Attouch and Svaiter in [2] introduced a Levenberg-Marquardt regularized dynamic enjoying interesting properties : for a general maximal monotone operator M , acting on a Hilbert space H , this dynamic is proved well-posed, stable, and having a good asymptotic behavior. A central question studied in [2] is the asymptotic behavior of trajectories, when the Levenberg-Marquardt regularization parameter tends rapidly to zero. In that case, one may expect to obtain rates of convergence close to Newton's method. This analysis is completed in [3] by the study of a closed-loop regularization method, i.e., at time t the regularization parameter is taken as a function of the current state of the system.

Yet, the dynamics in [2], [3], makes use of the resolvent of M , which may be difficult to compute in practical situations. In this paper, we will adapt (and partially extend) the dynamical

1. Abbas B, Attouch H, Svaiter BF.. Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces. *J. Optim. Theory Appl.* 2014;161 :331-360.

approach in [2], so as to deal with structured maximal monotone operators. Precisely, we suppose that $M = A + B$ can be written as the sum of a maximal monotone operator A (whose resolvents can be computed, at least approximatively), and a Lipschitz continuous monotone operator B . Our objective is to design continuous and discrete Newton-like dynamics attached to finding zeroes of M (the equilibria), and which exploit the particular structure of the operator M . Precisely, we want to use only the resolvents of A (implicit, backward step), and evaluation of B (explicit, forward step). So doing, we will obtain new insight on forward-backward, and gradient-projection methods for solving structured monotone inclusions.

2.2.2 Preliminaries. Theoretical background

Throughout this paper, H is a real Hilbert space with scalar product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. As a guideline to our study, we use the Newton-like dynamic introduced by Attouch and Svaiter in [2]. Let us first recall the main lines of this approach. Then, we shall explain how to adapt it, in order to fit our splitting program. When M is a differentiable mapping, the classical Newton-Raphson method generates sequences $(x_k)_{k \in \mathbb{N}}$ in H verifying

$$M(x_k) + M'(x_k)(x_{k+1} - x_k) = 0.$$

When the current iterate is far from the solution, it is convenient to introduce a positive step-size Δt_k , and consider

$$M(x_k) + M'(x_k) \left(\frac{x_{k+1} - x_k}{\Delta t_k} \right) = 0.$$

Unless restrictive assumptions on M are made, this is not a well-posed equation. To overcome this difficulty, we consider the following regularized version of the Newton-Raphson method

$$M(x_k) + (\lambda_k I + M'(x_k)) \left(\frac{x_{k+1} - x_k}{\Delta t_k} \right) = 0,$$

where I is the identity operator on H , and $(\lambda_k)_{k \in \mathbb{N}}$ is a sequence of positive real numbers (in the particular case of the Gauss-Newton method, this is the Levenberg-Marquardt regularization method). The algorithm has a natural interpretation as a time discretized version of the continuous dynamic

$$\lambda(t) \dot{x}(t) + M'(x(t)) \dot{x}(t) + M(x(t)) = 0,$$

where $\dot{x}(t) = \frac{dx}{dt}(t)$ is the derivative at time t of the mapping $t \mapsto x(t)$, and $t \mapsto \lambda(t)$ is a positive real-valued function. By using the derivation rule for the composition of smooth mappings, we can rewrite the previous equation as follows : find (x, v) solution of the differential-algebraic system

$$\begin{cases} v(t) = M(x(t)), \\ \lambda(t) \dot{x}(t) + \dot{v}(t) + v(t) = 0. \end{cases}$$

When $M : H \rightrightarrows H$ is a general maximal operator, the corresponding differential-algebraic inclusion system writes

$$\begin{cases} v(t) \in M(x(t)), \\ \lambda(t) \dot{x}(t) + \dot{v}(t) + v(t) = 0. \end{cases} \quad (2.2.1)$$

It involves an inclusion instead of an equality in the first equation. In order to study (2.2.1) with the help of the classical differential equations theory, we make use of Minty's representation of maximal monotone operators as Lipschitz manifolds.

Let $J_\mu^M = (I + \mu M)^{-1}$ and $M_\mu = \frac{1}{\mu}(I - J_\mu^M)$ be respectively the resolvent and the Yosida approximation of index $\mu > 0$ of M . For any $t \in [0, +\infty[$, set $\mu(t) = \frac{1}{\lambda(t)}$, and introduce the new unknown function $z : [0, +\infty[\rightarrow H$, which is defined by

$$z(t) = x(t) + \mu(t)v(t). \quad (2.2.2)$$

Minty's representation gives the equivalent form of the inclusion $v(t) \in M(x(t))$

$$\begin{aligned} x(t) &= J_{\mu(t)}^M(z(t)), \\ v(t) &= M_{\mu(t)}(z(t)). \end{aligned}$$

It is a straightforward computation to verify that (2.2.1) is equivalent to

$$\begin{cases} x(t) = J_{\mu(t)}^M(z(t)), \\ \dot{z}(t) + (\mu(t) - \dot{\mu}(t))M_{\mu(t)}(z(t)) = 0. \end{cases} \quad (2.2.3)$$

As an important point note that, because of the Lipschitz continuous property of M_μ , the above differential system in z can be treated by the Cauchy-Lipschitz theorem.

This approach, which has been developed in [2], makes use of the resolvent of M . In typical situations (like convex constrained minimization), $M = N_C + \nabla\Phi$, where N_C is the normal cone mapping to the constraint set $C \subset H$, and $\nabla\Phi$ is the gradient of the differentiable criterion $\Phi : H \rightarrow \mathbb{R}$. In such situations, computing the resolvent of M turns out to be a difficult problem. By contrast, computing the resolvent of $A = N_C$ is equivalent to finding the projection onto C , a simpler problem. Other particular important situations come from inverse problems in signal/imaging, see [4], [5]. In order to recover sparse solutions of under-determined linear systems, one usually considers the l^1 regularization of the least squares minimization problem

$$\min\{\gamma\|x\|_1 + \frac{1}{2}\|Ax - b\|^2 : x \in \mathbb{R}^n\}. \quad (2.2.4)$$

This is a structured minimization problem with $\Phi(x) = \gamma\|x\|_1$ a convex nondifferentiable function, and $\Psi(x) = \frac{1}{2}\|Ax - b\|^2$ a convex C^1 function. The crucial point is that the proximity mapping of the l^1 norm (equivalently, the resolvent of the subdifferential of the l^1 norm) can be computed explicitly, via elementary operations. It is just the soft thresholding operator. The classical forward-backward method provides the so-called iterative shrinkage/thresholding algorithm (ISTA).

Thus, in the spirit of the splitting forward-backward methods, we are going to exploit the structure of the operator $M = A + B$, and first develop continuous dynamics, then algorithms which only use the resolvents of the nonsmooth operator A . Namely, instead of (2.2.1), we are going to study the dynamic

$$\begin{cases} v(t) \in A(x(t)), \\ \lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) + B(x(t)) = 0. \end{cases} \quad (2.2.5)$$

Note that in (2.2.5), by contrast with (2.2.1), the operators A and B play asymmetric roles : if $A = 0$, the operator B governs a gradient type dynamic, while if $B = 0$ the operator A governs a Newton-like dynamic. By applying the Minty transformation to A , we will obtain the new dynamic

$$\begin{cases} x(t) = J_{\mu(t)}^A z(t) \\ \dot{z}(t) + (\mu(t) - \dot{\mu}(t)) A_{\mu(t)} z(t) + \mu(t) B \left(J_{\mu(t)}^A z(t) \right) = 0. \end{cases} \quad (2.2.6)$$

System (2.2.6) is relevant to the Cauchy-Lipschitz theorem, and makes use only of the resolvents of A . Equivalent systems (2.2.5) and (2.2.6) provide a dynamical approach to forward-backward algorithms. Indeed, time discretization of (2.2.6) naturally yields a large class of such algorithms, including the classical ones.

The choice of the regularization parameter $\lambda(t) > 0$ is a crucial point of our analysis. Following [2], to obtain a dynamic that is close to the Newton method, one should take $\lambda(t) \rightarrow 0$ as $t \rightarrow +\infty$, with $\lambda(t)$ taken as small as possible. In doing so, we can expect a rapid convergence of the trajectories towards equilibria, which are zeros of the operator $A + B$.

The paper is organized as follows : In Section 2.2.3, Theorem 2.2.4, by relying on the equivalence with (2.2.6), we show the existence and uniqueness of a strong global solution to the Cauchy problem (2.2.5), with B monotone and globally Lipschitz continuous. From Section 2.2.4, we will focus on the convex subdifferential case. We first show a descent property and a classical energy estimate. Then, in Theorem 2.2.7, we prove existence and uniqueness of a strong global solution to the Cauchy problem (2.2.5), just assuming that $B = \nabla\Psi$ is locally Lipschitz continuous. In Section 2.2.5, we study the asymptotic convergence as $t \rightarrow +\infty$ of the trajectories of our dynamical system. In Theorems 2.2.15 and 2.2.16, by relying on a careful Lyapunov analysis (in the spirit of energy arguments developed in [6]), and Opial's lemma, we show that if the regularization parameter $\lambda(t)$ does not tend to zero too rapidly (roughly speaking no faster than e^{-t}), then each trajectory converges weakly to an equilibrium which is a solution of the corresponding minimization problem. Indeed, the asymptotic convergence analysis is based on energy estimates that require the quantity $1 + \frac{\dot{\lambda}(t)}{\lambda(t)}$ to be nonnegative (see Proposition 2.2.10), which gives the critical size $\lambda(t) = e^{-t}$. Finally, in Section 2.2.6, we introduce some new forward-backward, and gradient-projection algorithms which are obtained by time discretization of the continuous dynamics.

2.2.3 Existence and uniqueness of global solutions.

We consider the Cauchy problem for the differential inclusion system

$$v(t) \in A(x(t)) \quad (2.2.7a)$$

$$\lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) + B(x(t)) = 0 \quad (2.2.7b)$$

$$x(0) = x_0, \quad v(0) = v_0, \quad v_0 \in A(x_0). \quad (2.2.7c)$$

First, we are going to define a notion of strong solution to the above system. Then, we shall reformulate this system with the help of the Minty representation of A . Finally, we shall prove the existence and uniqueness of a strong solution to system (2.2.7a)-(2.2.7c), by applying the Cauchy-Lipschitz theorem to this equivalent formulation.

Definition of Strong Solutions.

Let us first recall some notions concerning vector-valued functions of real variables (see [6, Appendix] for further details).

Definition 2.2.1 *Given $b \in \mathbb{R}^+$, a function $f : [0, b] \rightarrow H$ is said to be absolutely continuous iff one of the following equivalent properties holds :*

(i) *there exists an integrable function $g : [0, b] \rightarrow H$ such that*

$$f(t) = f(0) + \int_0^t g(s) ds, \quad \forall t \in [0, b];$$

(ii) *f is continuous, and its distributional derivative is Lebesgue integrable on $[0, b]$.*

(iii) *for every $\epsilon > 0$, there exists $\eta > 0$ such that, for any finite family of intervals $I_k = (a_k, b_k)$,*

$$(I_k \cap I_j = \emptyset \text{ for } i \neq j, \text{ and } \sum_k |b_k - a_k| \leq \eta) \implies (\sum_k \|f(b_k) - f(a_k)\| \leq \epsilon.)$$

Moreover, an absolutely continuous function is differentiable almost everywhere, its derivative coincide with its distributional derivative almost everywhere, and one can recover the function from its derivative $f' = g$ by integration formula (i).

Definition 2.2.2 *We say that the pair $(x(\cdot), v(\cdot))$ is a strong global solution of (2.2.7a)-(2.2.7c) iff the following properties are satisfied :*

(i) *$x(\cdot), v(\cdot) : [0, +\infty[\rightarrow H$ are absolutely continuous on each interval $[0, b]$, $0 < b < +\infty$;*

(ii) *$v(t) \in A(x(t))$ for all $t \in [0, +\infty[$;*

(iii) *$\lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) + B(x(t)) = 0$ for almost all $t \in [0, +\infty[$;*

(iv) *$x(0) = x_0, v(0) = v_0$.*

The last condition makes sense because of the continuity property of $x(\cdot)$ and $v(\cdot)$. Let us now make our standing assumption on function $\lambda(\cdot)$:

$$\lambda : [0, +\infty[\rightarrow]0, +\infty[\text{ is absolutely continuous on each interval } [0, b], \quad 0 < b < +\infty. \quad (2.2.8)$$

Hence $\dot{\lambda}(t)$ exists for almost all $t > 0$, and $\dot{\lambda}(\cdot)$ is Lebesgue integrable on each bounded interval $[0, b]$. We stress the fact that we assume $\lambda(t) > 0$ for any $t \geq 0$. By continuity of $\lambda(\cdot)$, this implies that, for any $b > 0$, there exists some positive finite lower and upper bounds for $\lambda(\cdot)$ on $[0, b]$, i.e., for any $t \in [0, b]$

$$0 < \lambda_{b, \min} < \lambda(t) < \lambda_{b, \max} < +\infty. \quad (2.2.9)$$

This fact will be of importance for proving existence of strong solutions.

Equivalent Formulation as a Classical Differential Equation

In order to solve system (2.2.7a)-(2.2.7c) we use Minty's device. Set $\mu(t) = \frac{1}{\lambda(t)}$. Let us rewrite inclusion (2.2.7a) by using the following equivalences : for any $t \in [0, +\infty)$

$$v(t) \in A(x(t)) \Leftrightarrow \quad (2.2.10)$$

$$x(t) + \mu(t)v(t) \in x(t) + \mu(t)A(x(t)) \Leftrightarrow \quad (2.2.11)$$

$$x(t) = (I + \mu(t)A)^{-1}(x(t) + \mu(t)v(t)). \quad (2.2.12)$$

Let us introduce the new unknown function $z : [0, +\infty[\rightarrow H$ which is defined for $t \in [0, +\infty[$ by

$$z(t) = x(t) + \mu(t)v(t), \quad (2.2.13)$$

and rewrite system (2.2.7a)-(2.2.7c) with the help of (x, z) . From (2.2.12) and (2.2.13)

$$\begin{aligned} x(t) &= (I + \mu(t)A)^{-1}(z(t)) \\ v(t) &= \frac{1}{\mu(t)} (z(t) - (I + \mu(t)A)^{-1}(z(t))). \end{aligned}$$

Equivalently,

$$x(t) = J_{\mu(t)}^A(z(t)) \quad (2.2.14)$$

$$v(t) = A_{\mu(t)}(z(t)), \quad (2.2.15)$$

where $J_{\mu}^A = (I + \mu A)^{-1}$ and $A_{\mu} = \frac{1}{\mu} (I - J_{\mu}^A)$ are respectively the resolvent and the Yosida approximation of index $\mu > 0$ of the maximal monotone operator A (one can consult [6], [7], [8] for basic facts concerning maximal monotone operators and Yosida approximation).

Indeed, this is Minty's representation of maximal monotone operators, see [9]. In a finite dimensional setting, this technique has been developed by Rockafellar in [10] : he shows that a maximal monotone operator can be represented as a Lipschitz manifold, which allows him to define second-order derivatives of convex lower semicontinuous functions (a program close to ours, since, formally, Newton's method for a subdifferential operator ∂f involves a second-order generalized derivative of f).

Let us show how (2.2.7b) can be reformulated as a classical differential equation with respect to $z(\cdot)$. First, let us rewrite (2.2.7b) as

$$\dot{x}(t) + \mu(t)\dot{v}(t) + \mu(t)v(t) + \mu(t)B(x(t)) = 0. \quad (2.2.16)$$

Differentiating (2.2.13) and using (2.2.16) we obtain

$$\dot{z}(t) = \dot{x}(t) + \mu(t)\dot{v}(t) + \dot{\mu}(t)v(t) \quad (2.2.17)$$

$$= -\mu(t)v(t) + \dot{\mu}(t)v(t) - \mu(t)B(x(t)). \quad (2.2.18)$$

From (2.2.14), (2.2.15) and (2.2.18) we deduce that

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t))A_{\mu(t)}(z(t)) + \mu(t)B(J_{\mu(t)}^A z(t)) = 0.$$

Finally, the (x, z) system can be written as

$$x(t) = J_{\mu(t)}^A(z(t)) \quad (2.2.19a)$$

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t))A_{\mu(t)}(z(t)) + \mu(t)B(J_{\mu(t)}^A z(t)) = 0. \quad (2.2.19b)$$

Conversely, if z is solution of (2.2.19b), then (x, v) with $x(t) = J_{\mu(t)}^A(z(t))$, $v(t) = A_{\mu(t)}(z(t))$ is solution of (2.2.7a)-(2.2.7c). This is detailed in the proof of Theorem 2.4.

As a nice feature of system (2.2.19a)-(2.2.19b), let us stress the fact that the operators $J_{\mu}^A : H \rightarrow H$ and $A_{\mu} : H \rightarrow H$ are everywhere defined and Lipschitz continuous, which makes this system relevant to the Cauchy-Lipschitz theorem.

Global Existence and Uniqueness Results.

The formulation (2.2.19a)-(2.2.19b) of our dynamic as a (x, z) system involves the time-dependent operators $J_{\mu(t)}^A$ and $A_{\mu(t)}$. In order to establish existence results for the evolution equation (2.2.19b), we will need the following regularity properties of the mappings $\mu \mapsto J_{\mu}^A x$ and $\mu \mapsto A_{\mu} x$.

Proposition 2.2.3 *For any $\lambda > 0$, $\mu > 0$, $x \in H$, the following properties hold :*

$$(i) \quad J_{\lambda}^A x = J_{\mu}^A \left(\frac{\mu}{\lambda} x + \left(1 - \frac{\mu}{\lambda}\right) J_{\lambda}^A x \right);$$

$$(ii) \quad \|J_{\lambda}^A x - J_{\mu}^A x\| \leq |\lambda - \mu| \|A_{\lambda} x\|.$$

As a consequence, for any $x \in H$ and any $0 < \delta < +\infty$, the mapping $\tau \mapsto J_{\tau}^A x$ is Lipschitz continuous on $[\delta, +\infty[$. More precisely, for any λ, μ belonging to $[\delta, +\infty[$

$$\|J_{\lambda}^A x - J_{\mu}^A x\| \leq |\lambda - \mu| \|A_{\delta} x\|. \quad (2.2.20)$$

Proof. Equality (i) is known as the resolvent equation. It is a classical result ; see, [6], [7, Proposition 23.28].

(ii) is a consequence of the resolvent equation, and the nonexpansive property of the resolvents, for a detailed proof see [7, Proposition 23.28]. According to the non-increasing property of $\lambda \mapsto \|A_{\lambda} x\|$, see [6, Proposition 2.6], we obtain (2.2.20). \square

We can now state our main existence and uniqueness result for system (2.2.7a)-(2.2.7c).

Theorem 2.2.4 *Let $\lambda : [0, +\infty[\rightarrow]0, +\infty[$ be a continuous function which is absolutely continuous on each bounded interval $[0, b]$, $b > 0$. Set $\mu(t) = \frac{1}{\lambda(t)}$. Let $A : H \rightrightarrows H$ be a maximal monotone operator, and let $B : H \rightarrow H$ be a maximal monotone operator which is Lipschitz continuous on H . Let $(x_0, v_0) \in H \times H$ be such that $v_0 \in A(x_0)$. Then the following properties hold :*

i) there exists a unique strong global solution $(x(\cdot), v(\cdot)) : [0, +\infty[\rightarrow H \times H$ of the Cauchy problem (2.2.7a)-(2.2.7c) ;

ii) the solution pair $(x(\cdot), v(\cdot))$ of (2.2.7a)-(2.2.7c) can be represented as follows : for any $t \in [0, +\infty[$,

$$x(t) = J_{\mu(t)}^A (z(t)), \quad (2.2.21)$$

$$v(t) = A_{\mu(t)} (z(t)), \quad (2.2.22)$$

where $z(\cdot) : [0, +\infty[\rightarrow H$ is the unique strong global solution of the Cauchy problem

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t)) A_{\mu(t)}(z(t)) + \mu(t) B (J_{\mu(t)}^A z(t)) = 0 \quad (2.2.23a)$$

$$z(0) = x_0 + \mu(0) v_0. \quad (2.2.23b)$$

Proof. The proof is an adaptation to our situation (with two operators A and B) of the proof of [2, Theorem 2.4] (with a single operator A).

1) Let us first prove existence and uniqueness of a strong global solution of the Cauchy problem (2.2.23a)-(2.2.23b). By the definition of $\mu(t) = \frac{1}{\lambda(t)}$, we have $\dot{\mu}(t) = -\frac{\dot{\lambda}(t)}{\lambda^2(t)}$. Thus, (2.2.23a) can be written as

$$\dot{z}(t) + \left(1 + \frac{\dot{\lambda}(t)}{\lambda(t)}\right) \frac{1}{\lambda(t)} A_{\frac{1}{\lambda(t)}}(z(t)) + \frac{1}{\lambda(t)} B \left(J_{\frac{1}{\lambda(t)}}^A z(t) \right) = 0. \quad (2.2.24)$$

The Cauchy problem (2.2.23a)-(2.2.23b) is equivalent to

$$\dot{z}(t) = F(t, z(t)) \quad (2.2.25a)$$

$$z(0) = x_0 + \mu(0) v_0, \quad (2.2.25b)$$

with

$$F(t, z) = \theta(t)G(t, z) + K(t, z), \quad (2.2.26a)$$

$$\theta(t) = - \left(1 + \frac{\dot{\lambda}(t)}{\lambda(t)}\right), \quad (2.2.26b)$$

$$G(t, z) = \frac{1}{\lambda(t)} A_{\frac{1}{\lambda(t)}}(z), \quad (2.2.26c)$$

$$K(t, z) = -\frac{1}{\lambda(t)} B \left(J_{\frac{1}{\lambda(t)}}^A(z) \right). \quad (2.2.26d)$$

In order to apply the Cauchy-Lipschitz theorem to (2.2.25a)-(2.2.25b), let us first examine the Lipschitz continuity properties of $F(t, \cdot)$.

a) Take arbitrary $z_i \in H$, $i = 1, 2$. For any $t \geq 0$, $G(t, \cdot) : H \rightarrow H$ is nonexpansive, i.e.,

$$\|G(t, z_2) - G(t, z_1)\| \leq \|z_2 - z_1\|. \quad (2.2.27)$$

Let L be the Lipschitz constant of B . By the nonexpansive property of the resolvent operators $J_{\frac{1}{\lambda(t)}}^A$ we have

$$\|B \left(J_{\frac{1}{\lambda(t)}}^A(z_2) \right) - B \left(J_{\frac{1}{\lambda(t)}}^A(z_1) \right)\| \leq L \|z_2 - z_1\|. \quad (2.2.28)$$

By definition (2.2.26a)-(2.2.26d) of F , (2.2.27), and (2.2.28), we obtain

$$\|F(t, z_2) - F(t, z_1)\| \leq L_F(t) \|z_2 - z_1\|, \quad (2.2.29)$$

with

$$L_F(t) = 1 + \frac{L + |\dot{\lambda}(t)|}{\lambda(t)}. \quad (2.2.30)$$

Since $\dot{\lambda}(\cdot)$ is locally integrable, and $\lambda(\cdot)$ is bounded away from zero on any bounded interval, (2.2.30) shows that the Lipschitz constant $L_F(t)$ of $F(t, \cdot)$ satisfies

$$L_F(\cdot) \in L^1([0, b]) \text{ for any } 0 < b < +\infty. \quad (2.2.31)$$

b) Let us show that

$$\forall z \in H, \forall b > 0, F(\cdot, z) \in L^1([0, b]; H). \quad (2.2.32)$$

By (2.2.9), for any $t \in [0, b]$, we have $0 < \lambda_{b, \min} \leq \lambda(t) \leq \lambda_{b, \max} < +\infty$. Returning to the definition (2.2.26a)-(2.2.26d) of F , we deduce that

$$\|F(t, z)\| \leq \left(1 + \frac{|\dot{\lambda}(t)|}{\lambda_{b, \min}}\right) \frac{1}{\lambda_{b, \min}} \|A_{\frac{1}{\lambda_{b, \max}}} z\| + \|K(t, z)\|. \quad (2.2.33)$$

On the other hand,

$$\|K(t, z)\| \leq \frac{1}{\lambda_{b, \min}} \left(\|B(t, z_0)\| + L \|J_{\frac{1}{\lambda(t)}}^A(z) - z_0\| \right), \quad (2.2.34)$$

where z_0 is arbitrarily chosen in H . Using again the local integrability of $\dot{\lambda}(\cdot)$ and Proposition 2.2.3 (ii) we obtain (2.2.32).

From properties (2.2.29), (2.2.31), and (2.2.32), we can deduce the existence and uniqueness of a strong global solution of differential equation (2.2.25a)-(2.2.25b), with given Cauchy data. To that end, we use the version of the Cauchy-Lipschitz theorem relying on the integrability of $t \mapsto F(t, x)$, and involving absolutely continuous trajectories ; see for example [11, Proposition 6.2.1], [12, Theorem 54].

2) Let us now return to the initial problem(2.2.7a)-(2.2.7c). Given $z(\cdot) : [0, +\infty) \rightarrow H$, which is the unique strong solution of Cauchy problem (2.2.23), let us define $x(\cdot), v(\cdot) : [0, +\infty) \rightarrow H$ by

$$x(t) = J_{\mu(t)}^A(z(t)), \quad v(t) = A_{\mu(t)}(z(t)). \quad (2.2.35)$$

a) Let us show that $x(\cdot), v(\cdot)$ are absolutely continuous on each bounded interval, and satisfy system (2.2.7a)-(2.2.7c). Let us give arbitrary $z_1 \in H, z_2 \in H$ and $\mu_1 > 0, \mu_2 > 0$. Combining Proposition 2.2.3 (ii) and the contraction property of the resolvents, we obtain

$$\|J_{\mu_2}^A(z_2) - J_{\mu_1}^A(z_1)\| \leq \|J_{\mu_2}^A(z_2) - J_{\mu_2}^A(z_1)\| + \|J_{\mu_2}^A(z_1) - J_{\mu_1}^A(z_1)\| \quad (2.2.36)$$

$$\leq \|z_2 - z_1\| + |\mu_2 - \mu_1| \|A_{\mu_1} z_1\|. \quad (2.2.37)$$

Assuming that $s, t \in [0, b]$, by taking $z_1 = z(s), z_2 = z(t)$ and $\mu_1 = \mu(s), \mu_2 = \mu(t)$ in (2.2.37), and with the same notations as before (for any $t \in [0, b]$, $0 < \lambda_{b, \min} \leq \lambda(t) \leq \lambda_{b, \max} < +\infty$), setting more briefly $\Lambda = \lambda_{b, \max}$, we obtain

$$\|J_{\mu(t)}^A(z(t)) - J_{\mu(s)}^A(z(s))\| \leq \|z(t) - z(s)\| + |\mu(t) - \mu(s)| \|A_{\mu(t)}(z(t))\| \quad (2.2.38)$$

$$\leq \|z(t) - z(s)\| + |\mu(t) - \mu(s)| \|A_{\frac{1}{\Lambda}}(z(t))\|. \quad (2.2.39)$$

Noting that $\|A_{\frac{1}{\Lambda}}(z(t))\| \leq \|A_{\frac{1}{\Lambda}}(0)\| + \Lambda \|z(t)\|$ remains bounded on $[0, b]$, and owing to the absolute continuity property of $z(\cdot)$ and $\mu(\cdot)$, we deduce that $x(t) = J_{\mu(t)}^A(z(t))$ is absolutely continuous on $[0, b]$ for any $b > 0$. The same property holds true for $v(t) = \lambda(t) (z(t) - x(t))$, because $\lambda(\cdot)$ is absolutely continuous on $[0, b]$ for any $b > 0$, and the product of two absolutely continuous functions is still absolutely continuous ; see [13, Corollaire VIII.9]. Indeed this last property is a straight consequence of Definition 2.2.1(iii) of absolute continuity.

Moreover, for any $t \in [0, +\infty[$

$$v(t) \in A(x(t)), \quad z(t) = x(t) + \mu(t) v(t).$$

Differentiation of the above equation shows that for almost every $t > 0$

$$\dot{x}(t) + \mu(t)\dot{v}(t) + \dot{\mu}(t)v(t) = \dot{z}(t).$$

On other hand, owing to $v(t) = A_{\mu(t)}(z(t))$, $x(t) = J_{\mu(t)}^A(z(t))$, (2.2.23a) can be equivalently written as

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t))v(t) + \mu(t)B(x(t)) = 0.$$

Combining the two above equations, we obtain

$$\dot{x}(t) + \mu(t)\dot{v}(t) + \mu(t)v(t) + \mu(t)B(x(t)) = 0.$$

From $\mu(t) = \lambda(t)^{-1}$ we conclude that $(x(\cdot), v(\cdot))$ is a solution of system (2.2.7a)-(2.2.7c). Regarding the initial condition, we observe that

$$z(0) = x_0 + \mu(0)v_0, \tag{2.2.40}$$

$$= x(0) + \mu(0)v(0), \tag{2.2.41}$$

with $v_0 \in A(x_0)$ and $v(0) \in A(x(0))$. Hence

$$x(0) = x_0 = (I + \mu(0)A)^{-1}(x_0 + \mu(0)v_0).$$

Returning to (2.2.40)-(2.2.41), after simplification, we obtain $v(0) = v_0$.

b) Let us now prove uniqueness. Suppose that

$$x(\cdot), v(\cdot) : [0, +\infty[\longrightarrow H \times H$$

is a solution pair of (2.2.7a)-(2.2.7c). Defining $\mu(t) = \lambda(t)^{-1}$ and

$$z(t) = x(t) + \mu(t)v(t), \tag{2.2.42}$$

we conclude that $z(\cdot)$ is absolutely continuous (we use again that the product of two absolutely continuous functions is still absolutely continuous), $z_0 = x_0 + \mu v_0$, and for any $t \in [0, +\infty[$,

$$x(t) = (I + \mu(t)A)^{-1}(z(t)), \quad v(t) = A_{\mu(t)}(z(t)). \tag{2.2.43}$$

The functions involved in the definition (2.2.42) of z , (namely $x(\cdot), v(\cdot), \mu(\cdot)$) are differentiable for almost all $t > 0$. Hence, by using the the usual derivation rule for the sum and product of functions, and using (2.2.7b), we conclude that for almost all $t \in [0, +\infty[$

$$\begin{aligned} \dot{z}(t) &= \dot{x}(t) + \mu(t)\dot{v}(t) + \dot{\mu}(t)v(t). \\ &= -\mu(t)(\dot{v}(t) + v(t) + B(x(t))) + \mu(t)\dot{v}(t) + \dot{\mu}(t)v(t) \\ &= (-\mu(t) + \dot{\mu}(t))v(t) - \mu(t)B(x(t)). \end{aligned}$$

Since $v(t) = A_{\mu(t)}(z(t))$, we finally obtain

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t))A_{\mu(t)}(z(t)) + \mu(t)B(J_{\mu(t)}^A z(t)) = 0.$$

Moreover

$$z_0 = x_0 + \mu(0)v_0.$$

Arguing as before, by the Cauchy-Lipschitz theorem, $z(\cdot)$ is uniquely determined, and locally absolutely continuous. Thus, by (2.2.43), $x(\cdot)$ and $v(\cdot)$ are uniquely determined. \square

Remark 2.2.5 There are some other possibilities for the choice for the auxiliary variable z . For example, by taking $z(t) = \mu(t)y(t)$, i.e., $y(t) = \lambda(t)x(t) + v(t)$, one obtains the equivalent (x, v, y) system

$$x(t) = J_{\mu(t)}^A(\mu(t)y(t)), \quad (2.2.44a)$$

$$v(t) = A_{\mu(t)}(\mu(t)y(t)), \quad (2.2.44b)$$

$$\dot{y}(t) + y(t) - \frac{1}{\mu^2(t)}(\mu(t) - \dot{\mu}(t))J_{\mu(t)}^A(y(t)) + B(J_{\mu(t)}^A(\mu(t)y(t))) = 0. \quad (2.2.44c)$$

2.2.4 The convex subdifferential case

Let us now suppose that $A = \partial\varphi$ is the subdifferential of a convex lower semicontinuous proper function $\varphi : H \rightarrow \mathbb{R} \cup \{+\infty\}$, and that $B = \nabla\psi$ is the gradient of a convex, continuously differentiable function $\psi : H \rightarrow \mathbb{R}$. By a classical result, $\partial\varphi$ and $\nabla\psi$ are maximal monotone operators. System (2.2.7a)-(2.2.7c) reads as follows

$$v(t) \in \partial\varphi(x(t)) \quad (2.2.45a)$$

$$\lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) + \nabla\psi(x(t)) = 0. \quad (2.2.45b)$$

Descent Property

Let us show that the trajectories generated by system (2.2.45a)-(2.2.45b) satisfy a descent property. This property, and the accompanying energy estimate, will play a key role in establishing global existence results, just assuming that $B = \nabla\psi$ is locally Lipschitz continuous. That's why the following result is first stated on a bounded interval $[0, b]$.

Proposition 2.2.6 *Let (x, v) be a strong solution of system (2.2.45a)-(2.2.45b) on $[0, b]$, $b > 0$. Then*

a) $t \mapsto \varphi(x(t)) + \psi(x(t))$ is a non-increasing function.

b) $\int_0^b \lambda(t)\|\dot{x}(t)\|^2 dt < +\infty$, and hence $\dot{x} \in L^2(0, b; H)$.

Proof. After scalar multiplication of (2.2.45b) by $\dot{x}(t)$, we obtain

$$\lambda(t)\|\dot{x}(t)\|^2 + \langle \dot{x}(t), \dot{v}(t) \rangle + \langle v(t), \dot{x}(t) \rangle + \langle \nabla\psi(x(t)), \dot{x}(t) \rangle = 0. \quad (2.2.46)$$

For almost all $t \in [0, b]$, $\dot{x}(t)$ and $\dot{v}(t)$ are well defined; thus

$$\langle \dot{x}(t), \dot{v}(t) \rangle = \lim_{\substack{h \rightarrow 0 \\ h \neq 0}} \frac{1}{h^2} \langle x(t+h) - x(t), v(t+h) - v(t) \rangle.$$

By (2.2.45a), we have $v(t) \in \partial\varphi(x(t))$. Since $\partial\varphi : H \rightrightarrows H$ is monotone,

$$\langle x(t+h) - x(t), v(t+h) - v(t) \rangle \geq 0.$$

Dividing by h^2 and passing to the limit as $h \rightarrow 0$ preserves the inequality, which yields, for almost all $t \in [0, b]$

$$\langle \dot{x}(t), \dot{v}(t) \rangle \geq 0. \quad (2.2.47)$$

Combining (2.2.46) and (2.2.47) we obtain

$$\lambda(t) \|\dot{x}(t)\|^2 + \langle v(t), \dot{x}(t) \rangle + \langle \nabla \psi(x(t)), \dot{x}(t) \rangle \leq 0. \quad (2.2.48)$$

Noticing that x and v are continuous on $[0, b]$, hence bounded, and that λ is bounded from below on $[0, b]$ by a positive number, one easily gets from (2.2.48) that

$$\dot{x} \in L^2(0, b; H). \quad (2.2.49)$$

On the other hand, by the classical derivation chain rule

$$\frac{d}{dt} \psi(x(t)) = \langle \nabla \psi(x(t)), \dot{x}(t) \rangle. \quad (2.2.50)$$

We appeal to a similar formula which is still valid for a convex lower semicontinuous proper function $\varphi : H \rightarrow \mathbb{R} \cup \{+\infty\}$. Notice that

i) $v(t) \in \partial \varphi(x(t))$ for almost every $t \in [0, b]$;

ii) v is continuous on $[0, b]$ and hence belongs to $L^2(0, b; H)$;

iii) $\dot{x} \in L^2(0, b; H)$ by (2.2.49).

By *i)*, *ii)*, *iii)*, conditions of [6, Lemma 3.3] are satisfied, which allows to deduce that $t \mapsto \varphi(x(t))$ is absolutely continuous on $[0, b]$, and, for almost every $t \in [0, b]$

$$\frac{d}{dt} \varphi(x(t)) = \langle v(t), \dot{x}(t) \rangle. \quad (2.2.51)$$

Combining (2.2.48) with (2.2.50) and (2.2.51), we obtain

$$\lambda(t) \|\dot{x}(t)\|^2 + \frac{d}{dt} (\varphi(x(t)) + \psi(x(t))) \leq 0. \quad (2.2.52)$$

Hence

$$\frac{d}{dt} (\varphi(x(t)) + \psi(x(t))) \leq 0,$$

which implies that the absolutely continuous function $t \mapsto \varphi(x(t)) + \psi(x(t))$ is non-increasing. Integration of (2.2.52) immediately yields $\int_0^b \lambda(t) \|\dot{x}(t)\|^2 dt < +\infty$. \square

Global Existence with $B = \nabla \psi$ Locally Lipschitz Continuous

We consider the Cauchy problem

$$v(t) \in \partial \varphi(x(t)), \quad (2.2.53a)$$

$$\lambda(t) \dot{x}(t) + \dot{v}(t) + v(t) + \nabla \psi(x(t)) = 0, \quad (2.2.53b)$$

$$x(0) = x_0, \quad v(0) = v_0, \quad v_0 \in \partial \varphi(x_0). \quad (2.2.53c)$$

We suppose that the operator $B = \nabla \psi$ is Lipschitz continuous on bounded sets, i.e., for any $R > 0$, there exists $L_R > 0$ such that for any $z_i \in H$, $i = 1, 2$

$$\|z_2\| \leq R, \quad \|z_1\| \leq R \implies \|\nabla \psi(z_2) - \nabla \psi(z_1)\| \leq L_R \|z_1 - z_2\|.$$

For short, we say that $B = \nabla \psi$ is locally Lipschitz continuous.

Theorem 2.2.7 *Let $\lambda : [0, +\infty[\rightarrow]0, +\infty[$ be a continuous function which is absolutely continuous on each bounded interval $[0, b]$, $b > 0$. Let $A = \partial\varphi$ be the subdifferential of a convex lower semicontinuous proper function $\varphi : H \rightarrow \mathbb{R} \cup \{+\infty\}$. Let $B = \nabla\psi$ be the gradient of a convex, continuously differentiable function $\psi : H \rightarrow \mathbb{R}$, and suppose that B is locally Lipschitz continuous. Suppose that $\varphi + \psi$ is bounded from below on H . Let $(x_0, v_0) \in H \times H$ be such that $v_0 \in \partial\varphi(x_0)$.*

Then the following properties hold

i) There exists a unique strong global solution $(x(\cdot), v(\cdot)) : [0, +\infty[\rightarrow H \times H$ of the Cauchy problem (2.2.53a)-(2.2.53c).

ii) Set $\mu(t) = \frac{1}{\lambda(t)}$. The solution pair $(x(\cdot), v(\cdot))$ of the problem (2.2.53a)-(2.2.53c) can be represented as follows : for any $t \in [0, +\infty[$,

$$x(t) = J_{\mu(t)}^{\partial\varphi}(z(t)), \quad (2.2.54)$$

$$v(t) = \nabla\varphi_{\mu(t)}(z(t)), \quad (2.2.55)$$

where $z(\cdot) : [0, +\infty[\rightarrow H$ is the unique strong solution of the Cauchy problem

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t)) \nabla\varphi_{\mu(t)}(z(t)) + \mu(t) \nabla\psi\left(J_{\mu(t)}^{\partial\varphi}(z(t))\right) = 0 \quad (2.2.56)$$

$$z(0) = x_0 + \mu(0)v_0. \quad (2.2.57)$$

Proof. Relying on the local Lipschitz continuity of $B = \nabla\psi$, and the Cauchy-Lipschitz theorem in the locally Lipschitz case, a demonstration similar to that used in Theorem 2.2.4 gives the local existence and uniqueness of a strong solution of the Cauchy problem (2.2.53a)-(2.2.53c).

In order to pass from a local solution to a global solution we follow a standard argument. Let $[0, T_{max}[$ be the maximal interval of existence of a solution. We want to show that $T_{max} = +\infty$. We follow a proof by contradiction, and assume that $T_{max} < +\infty$. By Proposition 2.2.6, $t \mapsto \varphi(x(t)) + \psi(x(t))$ is a decreasing function. By (2.2.52) and the assumption $\inf_H(\varphi + \psi) > -\infty$, we obtain the energy estimate

$$\int_0^{T_{max}} \lambda(t) \|\dot{x}(t)\|^2 dt \leq (\varphi + \psi)(x_0) - \inf_H(\varphi + \psi). \quad (2.2.58)$$

According to (2.2.9) and the fact that $T_{max} < +\infty$, we have that, for any $t \in [0, T_{max}]$

$$0 < \lambda_{T_{max}, min} < \lambda(t). \quad (2.2.59)$$

For short, set

$$T_m = T_{max} \quad \text{and} \quad \lambda_{min} = \lambda_{T_m, min}.$$

From (2.2.58), (2.2.59), and the minorization assumption concerning $\varphi + \psi$, we infer

$$\lambda_{min} \int_0^{T_m} \|\dot{x}(t)\|^2 dt \leq \int_0^{T_m} \lambda(t) \|\dot{x}(t)\|^2 dt < +\infty.$$

Hence

$$\dot{x}(\cdot) \in L^2(0, T_m; H). \quad (2.2.60)$$

By a standard argument, this implies that $\lim_{t \rightarrow T_m} x(t)$ exists. Indeed, for any sequence $t_n \rightarrow T_m$, by using successively that $x(\cdot)$ is absolutely continuous on bounded sets, and Cauchy-Schwarz inequality we have

$$\begin{aligned} \|x(t_n) - x(t_m)\| &\leq \int_{t_n}^{t_m} \|\dot{x}(s)\| ds \\ &\leq \sqrt{|t_n - t_m|} \left(\int_0^{T_m} \|\dot{x}(s)\|^2 ds \right)^{\frac{1}{2}}. \end{aligned}$$

Since $T_m < +\infty$ and $t_n \rightarrow T_m$, we deduce from the above inequality that $(x(t_n))$ is a Cauchy sequence in the Hilbert space H , and hence $(x(t_n))$ converges. The same argument shows that the limit does not depend on the sequence $t_n \rightarrow T_m$. Thus

$$\lim_{t \rightarrow T_m} x(t) \text{ exists.} \quad (2.2.61)$$

Let us show that $v(t)$ and $z(t)$ also have a limit as $t \rightarrow T_m$. By integration of equation (2.2.53b) with respect to t we successively obtain

$$\begin{aligned} \lambda(t) \dot{x}(t) + \dot{v}(t) + v(t) + \nabla \psi(x(t)) &= 0, \\ \frac{d}{dt} e^t v(t) &= e^t (-\lambda(t) \dot{x}(t) - \nabla \psi(x(t))), \\ v(t) &= v(0) e^{-t} + e^{-t} \int_0^t e^{-s} (-\lambda(s) \dot{x}(s) - \nabla \psi(x(s))) ds. \end{aligned}$$

From (2.2.60), (2.2.61), and the above relation it follows that

$$\lim_{t \rightarrow T_m} v(t) \text{ exists.} \quad (2.2.62)$$

On the other hand, the relation $z(t) := x(t) + \mu(t)v(t)$ and (2.2.61-2.2.62) shows that

$$\lim_{t \rightarrow T_m} z(t) = z_{T_m} \text{ exists.} \quad (2.2.63)$$

We can reapply the Cauchy-Lipschitz theorem at the point (T_m, z_{T_m}) . So doing we obtain a strong solution which is defined on an interval of length strictly greater than T_m , a clear contradiction with the maximality of T_m . Thus $T_m = +\infty$, and problem (2.2.53a)-(2.2.53c) has a unique strong global solution. \square

2.2.5 Convergence analysis

We work in the framework which has been delineated in the previous section, namely

i) $A = \partial\varphi$ is the subdifferential of a convex lower semicontinuous proper function $\varphi : H \rightarrow \mathbb{R} \cup \{+\infty\}$.

ii) $B = \nabla\psi$ is the gradient of a convex, continuously differentiable function $\psi : H \rightarrow \mathbb{R}$, and $\nabla\psi$ is assumed to be locally Lipschitz continuous.

iii) $\varphi + \psi$ is bounded from below on H .

We are going to study the minimizing and the convergence properties of the trajectories of system (2.2.45a)-(2.2.45b). The existence of strong global solutions of this system is given by Theorem (2.2.53a)-(2.2.53c). Here and subsequently $(x(\cdot), v(\cdot)) : [0, +\infty) \rightarrow H \times H$ is the solution of the Cauchy problem (2.2.53). We will often omit the time variable t and write $x, v \dots$ for $x(t), v(t) \dots$ when no ambiguity arises.

We make use of the following Lyapunov functions : given $z \in H$,

$$g_z(t) := \varphi(z) - [\varphi(x(t)) + \langle z - x(t), v(t) \rangle], \quad (2.2.64)$$

$$h_z(t) := \psi(z) - [\psi(x(t)) + \langle z - x(t), \nabla\psi(x(t)) \rangle]. \quad (2.2.65)$$

Proposition 2.2.8 *For any $t \geq 0$ and z in the effective domain of $\varphi + \psi$*

$$g_z(t) \geq 0, \quad h_z(t) \geq 0,$$

and for almost all $t \geq 0$

$$\frac{d}{dt}g_z(t) = \langle x(t) - z, \dot{v}(t) \rangle.$$

Proof. The first inequality, $g_z(t) \geq 0$, follows from the subdifferential inequality for φ at $x(t)$, and $v(t) \in \partial\varphi(x(t))$, see (2.2.45a), while the second follows from the convexity of ψ .

We now use the derivation chain rule in the nonsmooth convex subdifferential case, see (2.2.51), and again $v \in \partial\varphi(x)$ to obtain

$$\begin{aligned} \frac{d}{dt}g_z(t) &= -\frac{d}{dt}\varphi(x(t)) + \langle \dot{x}(t), v(t) \rangle + \langle x(t) - z, \dot{v}(t) \rangle \\ &= -\langle \dot{x}(t), v(t) \rangle + \langle \dot{x}(t), v(t) \rangle + \langle x(t) - z, \dot{v}(t) \rangle \\ &= \langle x(t) - z, \dot{v}(t) \rangle. \end{aligned}$$

□

Minimizing Property

Proposition 2.2.9 *Suppose that λ is nonincreasing. Then*

$$\lim_{t \rightarrow +\infty} (\varphi + \psi)(x(t)) = \inf_H (\varphi + \psi).$$

Moreover, if $\arg \min(\varphi + \psi) \neq \emptyset$, then, for any $t > 0$,

$$(\varphi + \psi)(x(t)) - \min(\varphi + \psi) \leq \frac{1}{t} \left(\frac{\lambda_0}{2} \|x_0 - \bar{z}\|^2 + g_{\bar{z}}(0) \right)$$

where $\lambda_0 = \lambda(0)$ and \bar{z} is the orthogonal projection of x_0 onto $\arg \min(\varphi + \psi)$.

Proof. Take z in the effective domain of φ . Multiplying equation (2.2.45b)

$$\lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) + \nabla\psi(x(t)) = 0$$

by $x(t) - z$ we obtain

$$\begin{aligned} \lambda(t) \langle \dot{x}(t), x(t) - z \rangle + \langle \dot{v}(t), x(t) - z \rangle &= \langle v(t) + \nabla\psi(x(t)), z - x(t) \rangle \\ &\leq (\varphi + \psi)(z) - (\varphi + \psi)(x(t)), \end{aligned}$$

where the inequality follows from the convexity of φ , ψ and the inclusion $v(t) \in \partial\varphi(x(t))$. Combining the above inequality with Proposition 2.2.8 we conclude that

$$\frac{d}{dt} \left[\frac{\lambda(t)}{2} \|x(t) - z\|^2 + g_z(t) \right] \leq \frac{\dot{\lambda}(t)}{2} \|x(t) - z\|^2 + (\varphi + \psi)(z) - (\varphi + \psi)(x(t)).$$

Since $\dot{\lambda}(t) \leq 0$ for almost all $t \geq 0$, we have

$$\frac{d}{dt} \left[\frac{\lambda(t)}{2} \|x(t) - z\|^2 + g_z(t) \right] \leq (\varphi + \psi)(z) - (\varphi + \psi)(x(t)).$$

After integration on $[0, T]$, we obtain that, for any $T > 0$

$$\begin{aligned} - \left[\frac{\lambda(0)}{2} \|x_0 - z\|^2 + g_z(0) \right] &\leq \left[\frac{\lambda(T)}{2} \|x(T) - z\|^2 + g_z(T) \right] \\ &\quad - \left[\frac{\lambda(0)}{2} \|x_0 - z\|^2 + g_z(0) \right] \\ &\leq \int_0^T (\varphi + \psi)(z) - (\varphi + \psi)(x(t)) dt \\ &\leq T [(\varphi + \psi)(z) - (\varphi + \psi)(x(T))], \end{aligned}$$

where the first inequality comes from $g_z \geq 0$ (Proposition 2.2.8), and the last inequality follows from $(\varphi + \psi)(x)$ being non-increasing (Proposition 2.2.6). Altogether we have

$$(\varphi + \psi)(x(T)) - (\varphi + \psi)(z) \leq \frac{(\lambda_0/2) \|x_0 - z\|^2 + g_z(0)}{T}$$

which trivially implies the desired results. \square

Note that if the trajectory x is minimizing, then a necessary condition for it being bounded is the existence of solutions of

$$\min \varphi + \psi. \tag{2.2.66}$$

So this assumption is quite natural for proving boundedness of x , which will be accomplished with the use of the Lyapunov function

$$\Gamma_z(t) = \frac{1}{2} \|x(t) - z\|^2 + \lambda^{-1} g_z(t), \quad t \geq 0, \tag{2.2.67}$$

where z is an arbitrary point in the effective domain of $\varphi + \psi$. Using the second part of Proposition 2.2.8 and (2.2.45b) we have

$$\begin{aligned} \frac{d}{dt}\Gamma_z(t) &= \langle \dot{x}(t), x(t) - z \rangle + \lambda(t)^{-1} \langle \dot{v}(t), x(t) - z \rangle - \dot{\lambda}(t)\lambda(t)^{-2}g_z(t) \\ &= \frac{1}{\lambda(t)} \left[\langle \lambda(t)\dot{x}(t) + \dot{v}(t), x(t) - z \rangle - \frac{\dot{\lambda}(t)}{\lambda(t)}g_z(t) \right] \\ &= -\frac{1}{\lambda(t)} \left[\langle v(t) + \nabla\psi, (x(t)), x(t) - z \rangle + \frac{\dot{\lambda}(t)}{\lambda(t)}g_z(t) \right]. \end{aligned}$$

Direct algebraic manipulations on the expressions for g_z, h_z in (2.2.64)-(2.2.65) yield

$$g_z(t) + h_z(t) + (\varphi + \psi)(x(t)) - (\varphi + \psi)(z) = \langle x(t) - z, v(t) + \nabla\psi(x(t)) \rangle.$$

Combining the two above equations we conclude that

$$\frac{d}{dt}\Gamma_z(t) = -\frac{1}{\lambda(t)} \left[(\varphi + \psi)(x(t)) - (\varphi + \psi)(z) + \left(1 + \dot{\lambda}(t)/\lambda(t)\right) g_z(t) + h_z(t) \right]. \quad (2.2.68)$$

Proposition 2.2.10 *If $\arg \min(\varphi + \psi) \neq \emptyset$, and for almost all $t > 0$*

$$1 + \frac{\dot{\lambda}(t)}{\lambda(t)} \geq 0 \quad (2.2.69)$$

then

1. for any $z \in \arg \min(\varphi + \psi)$, Γ_z is decreasing, bounded and

$$\|h_z/\lambda\|_1 + \|((\varphi + \psi)(x) - (\varphi + \psi)(z))/\lambda\|_1 \leq \Gamma_z(0)$$

$$\lambda^{-1}h_z, \lambda^{-1}[(\varphi + \psi)(x) - (\varphi + \psi)(z)] \in L^1[0, \infty);$$

2. x is bounded;

3. $\|\dot{x}\|^2 \in L^1([0, \infty[)$

$$\|\|\dot{x}\|^2\|_1 \leq \left\| \frac{(\varphi + \psi)(x) - (\varphi + \psi)(z)}{\lambda} \right\|_1 + \frac{(\varphi + \psi)(x_0) - (\varphi + \psi)(z)}{\lambda(0)}.$$

Proof. Take $z \in \arg \min(\varphi + \psi)$. Using (2.2.68), (2.2.69) and the first part of Proposition 2.2.8 we conclude that, for almost all $t > 0$

$$\frac{d}{dt}\Gamma_z(t) \leq -\left\{ \frac{1}{\lambda(t)} [(\varphi + \psi)(x(t)) - (\varphi + \psi)(z)] + \frac{1}{\lambda}h_z(t) \right\}.$$

Since Γ_z and both functions inside the braces are non-negative, these two functions must be integrable and Γ_z is bounded, which proves item 1. Boundedness of Γ_z trivially implies

boundedness of x .

Still with $z \in \arg \min(\varphi + \psi)$ define

$$f(t) = (\varphi + \psi)(x(t)) - (\varphi + \psi)(z), \quad t \geq 0. \quad (2.2.70)$$

Performing the inner product of $\lambda^{-1}\dot{x}$ with both sides of equality (2.2.45b) yields

$$\|\dot{x}\|^2 + \frac{1}{\lambda} \langle \dot{x}, \dot{v} \rangle + \frac{1}{\lambda} \langle \dot{x}, v + \nabla \psi \rangle = 0.$$

We have $df/dt = \langle \dot{x}, v + \nabla \psi(x) \rangle$. Adding to both sides of the above equality $-(\dot{\lambda}/\lambda^2)f(t)$, and taking into account that $\langle \dot{x}, \dot{v} \rangle$ is non-negative, we conclude that

$$\|\dot{x}\|^2 + \frac{d}{dt} \left(\frac{1}{\lambda} f(t) \right) \leq -\frac{\dot{\lambda}}{\lambda^2} f(t).$$

By assumption (2.2.69) we have $-\frac{\dot{\lambda}}{\lambda^2} \leq \frac{1}{\lambda}$. Since f is nonnegative we deduce that

$$\|\dot{x}\|^2 + \frac{d}{dt} \left(\frac{1}{\lambda} f(t) \right) \leq \frac{1}{\lambda} f(t).$$

After integration on $[0, T]$, we obtain that, for any $T > 0$

$$\int_0^T \|\dot{x}\|^2 dt + \frac{1}{\lambda(T)} f(T) \leq \frac{1}{\lambda(0)} f(0) + \int_0^T \frac{1}{\lambda(t)} f(t) dt.$$

By nonnegativity of f and item 1 ($\frac{1}{\lambda} f \in L^1([0, +\infty[))$) we obtain item 3. \square

In short, from now on, we use the following terminology : we say that \bar{x} is a weak sequential cluster point of the trajectory x iff $\bar{x} = w - \lim x(t_n)$ for some sequence $t_n \rightarrow +\infty$.

Theorem 2.2.11 *Suppose that*

$$\int_0^{+\infty} \frac{1}{\lambda(t)} dt = +\infty \quad (2.2.71)$$

and for almost all $t > 0$

$$1 + \frac{\dot{\lambda}(t)}{\lambda(t)} \geq 0. \quad (2.2.72)$$

Then

$$\lim_{t \rightarrow +\infty} (\varphi + \psi)(x(t)) = \inf (\varphi + \psi). \quad (2.2.73)$$

If, additionally, $\arg \min (\varphi + \psi) \neq \emptyset$ then the trajectory x is bounded, all weak sequential cluster points of x are minimizers of $\varphi + \psi$, and for any $z \in \arg \min (\varphi + \psi)$

$$(\varphi + \psi)(x(t)) - (\varphi + \psi)(z) \leq \frac{\Gamma_z(0)}{\int_0^t 1/\lambda(s) ds}.$$

Proof. Take an arbitrary $z \in H$. Using (2.2.68), (2.2.72) and the first part of Proposition 2.2.8 we conclude that, for almost all $t > 0$

$$\frac{d}{dt} \Gamma_z(t) \leq -\frac{1}{\lambda} [(\varphi + \psi)(x(t)) - (\varphi + \psi)(z)],$$

which combined with the non-negativity of Γ_z (see definition (2.2.67), Proposition 2.2.8), and Proposition 2.2.6 yields the inequalities

$$\begin{aligned} \Gamma_z(0) \geq \Gamma_z(0) - \Gamma_z(t) &\geq \int_0^t \frac{1}{\lambda(s)} [(\varphi + \psi)(x(s)) - (\varphi + \psi)(z)] ds \\ &\geq [(\varphi + \psi)(x(t)) - (\varphi + \psi)(z)] \int_0^t \frac{1}{\lambda(s)} ds \end{aligned}$$

valid for any $t \geq 0$. Dividing the above inequalities by $\int_0^t 1/\lambda(s) ds$, using (2.2.71), and taking the limit as $t \rightarrow \infty$ we conclude that

$$\lim_{t \rightarrow \infty} (\varphi + \psi)(x(t)) - (\varphi + \psi)(z) \leq 0,$$

which means that the trajectory x is minimizing.

The second part of the theorem follows from the first one, Proposition 2.2.10 and the weak lower semicontinuity of $\varphi + \psi$. \square

Convergence of Trajectories

The following well-known results will be used in the proof of the weak convergence of x . For the reader's convenience, we give a proof of them.

Lemma 2.2.12 *Suppose that $F : [0, \infty[\rightarrow \mathbb{R}$ is locally absolutely continuous, bounded below, and there exists $G \in L^1([0, \infty[)$ such that for almost all t*

$$\frac{d}{dt} F(t) \leq G(t). \tag{2.2.74}$$

Then there exists $\lim_{t \rightarrow \infty} F(t) \in \mathbb{R}$.

Proof. By integration of (2.2.74) it follows that, for any $0 < s < t < \infty$

$$F(t) - F(s) \leq \int_s^t G(\tau) d\tau. \tag{2.2.75}$$

Equivalently

$$F(t) - \int_0^t G(\tau) d\tau \leq F(s) - \int_0^s G(\tau) d\tau. \tag{2.2.76}$$

Hence, the function $t \mapsto F(t) - \int_0^t G(\tau) d\tau$ is nonincreasing. It is bounded from below, because F has been assumed to be bounded below, and $G \in L^1([0, \infty[)$. As a consequence $F(t) - \int_0^t G(\tau) d\tau$ has a limit as $t \rightarrow \infty$. Using again that $G \in L^1([0, \infty[)$, we finally obtain that there exists $\lim_{t \rightarrow \infty} F(t) \in \mathbb{R}$. \square

Lemma 2.2.13 *If $1 \leq p < \infty$, $1 \leq r \leq \infty$, $F \in L^p([0, \infty[)$ is a locally absolutely continuous nonnegative function, $G \in L^r([0, \infty[)$ and for almost all t*

$$\frac{d}{dt}F(t) \leq G(t) \tag{2.2.77}$$

then $\lim_{t \rightarrow \infty} F(t) = 0$.

Proof. When $r = 1$ we are in the setting of Lemma 2.2.12. Hence, $\lim_{t \rightarrow \infty} F(t)$ exists and since $F \in L^p([0, \infty[)$ with $1 \leq p < \infty$, we deduce that $\lim_{t \rightarrow \infty} F(t) = 0$.

Let us now examine the case $1 < r \leq +\infty$. Let us introduce a positive real parameter $1 < q < +\infty$ (its value will be chosen appropriately later) and consider the function $t \mapsto F(t)^q$ (recall that $F(t)$ is nonnegative). By (2.2.77) we have

$$\frac{d}{dt}F(t)^q = qF(t)^{q-1} \frac{d}{dt}F(t) \tag{2.2.78}$$

$$\leq qF(t)^{q-1}G(t). \tag{2.2.79}$$

Let us choose the parameter $1 < q < +\infty$ so that the function $t \mapsto F(t)^{q-1}G(t)$ belongs to $L^1([0, \infty[)$. Since $G \in L^r([0, \infty[)$, this will be achieved if $F^{q-1} \in L^{r'}([0, \infty[)$ where r' is the Holder conjugate exponent of r , i.e., $\frac{1}{r} + \frac{1}{r'} = 1$. This is equivalent to $F^{(q-1)r'} \in L^1([0, \infty[)$. We know that $F \in L^p[0, \infty[$. Thus one can take $(q-1)r' = p$, that is

$$q = 1 + p\left(1 - \frac{1}{r}\right), \tag{2.2.80}$$

which verifies $1 < q < +\infty$. With this choice of q we have $F^{q-1}G$ belongs to $L^1([0, \infty[)$. By (2.2.78) and Lemma 2.2.12 we deduce that $\lim_{t \rightarrow \infty} F(t)^q$ exists, and hence $\lim_{t \rightarrow \infty} F(t)$ exists. Since $F \in L^p([0, \infty[)$, this forces this limit to be zero, which ends the proof. \square

In order to prove the weak convergence of the trajectories of system (2.2.45a)-(2.2.45b), we need the classical Opial's lemma [14] that we recall in its continuous form; see also [15], who initiated the use of this argument to analyze the asymptotic convergence of nonlinear contraction semigroups in Hilbert spaces.

Lemma 2.2.14 *Let S be a non empty subset of H and $x : [0, +\infty) \rightarrow H$ a map. Assume that*

- (i) *for every $z \in S$, $\lim_{t \rightarrow +\infty} \|x(t) - z\|$ exists;*
- (ii) *every weak sequential cluster point of the map x belongs to S .*

Then

$$w - \lim_{t \rightarrow +\infty} x(t) = x_\infty \quad \text{exists, for some element } x_\infty \in S.$$

We can now state our main convergence result.

Theorem 2.2.15 *Suppose that $\arg \min (\varphi + \psi) \neq \emptyset$, $\int_0^{+\infty} \frac{1}{\lambda(t)} dt = +\infty$, and there exists $\varepsilon > 0$ such that, for almost all $t > 0$*

$$1 + \frac{\dot{\lambda}(t)}{\lambda(t)} \geq \varepsilon. \quad (2.2.81)$$

Then $x(\cdot)$ converges weakly to a minimizer of $\varphi + \psi$.

Moreover, for any $z \in \arg \min (\varphi + \psi)$,

$$\frac{1}{\lambda} g_z \in L^1(0, \infty), \quad (2.2.82)$$

$$\lim_{t \rightarrow \infty} \frac{1}{\lambda(t)} g_z(t) = 0, \quad (2.2.83)$$

Proof. First note that all the assumptions of Proposition 2.2.10 and Theorem 2.2.11 are supposed to hold. Take $z \in \arg \min (\varphi + \psi)$. Using (2.2.68), Proposition 2.2.8 and (2.2.81) we conclude that

$$\frac{d}{dt} \Gamma_z(t) \leq - \left\{ \frac{1}{\lambda(t)} [(\varphi + \psi)(x(t)) - (\varphi + \psi)(z)] + \frac{\varepsilon}{\lambda(t)} g_z(t) + \frac{1}{\lambda(t)} h_z(t) \right\} \leq - \frac{\varepsilon}{\lambda(t)} g_z(t)$$

which, in view of the nonnegativity of Γ_z and g_z proves (2.2.82).

To prove (2.2.83), first use definition (2.2.67) to conclude that

$$\frac{d}{dt} \Gamma_z(t) = \langle \dot{x}(t), x(t) - z \rangle + \frac{d}{dt} \frac{1}{\lambda(t)} g_z(t) \leq 0,$$

where the inequality follows from item 1 of Proposition 2.2.10. Using item 2 of Proposition 2.2.10 we have

$$M = \sup_{t \geq 0} \|x(t) - z\| < \infty.$$

Combining the two above inequalities we conclude that

$$\frac{d}{dt} \frac{1}{\lambda} g_z(t) \leq M \|\dot{x}(t)\|.$$

According to item 3 of Proposition 2.2.10, $\|\dot{x}\| \in L^2$. To end the proof of (2.2.83), combine this inclusion with the above inequality, the non-negativity of g_z/λ and Lemma 2.2.13.

Using item 1 of Proposition 2.2.10, Γ_z is a non-increasing function. Hence the following limits exist

$$\lim_{t \rightarrow \infty} \Gamma_z(t) = \lim_{t \rightarrow \infty} \frac{1}{2} \|x(t) - z\|^2 + \lambda(t)^{-1} g_z(t) \in [0, \infty).$$

Combining the above result with (2.2.83) (which has just been proved), we conclude that there exists $\lim_{t \rightarrow \infty} \|x(t) - z\|$. As z is an arbitrary point in the solution set, x is bounded, and all its weak sequential cluster points belongs to the solution set. Using Opial's Lemma 2.2.14 we conclude that x is weakly convergent toward a particular point of the solution set. Note that our proof can be viewed as an adaptation (with Newton metric aspects) of the Bruck's proof of convergence for the generalized continuous steepest descent, see [15]. \square

Theorem 2.2.16 *If $\arg \min (\varphi + \psi) \neq \emptyset$, λ is bounded, and there exists $\varepsilon > 0$ such that, for almost all $t \geq 0$*

$$1 + \frac{\dot{\lambda}(t)}{\lambda(t)} \geq \varepsilon,$$

then $x(\cdot)$ converges weakly to a minimizer of $\varphi + \psi$, and

$$\lim_{t \rightarrow \infty} \|v(t) + \nabla\psi(x(t))\| = 0.$$

Proof. We already know :

1. x is bounded and converges weakly to a solution,
2. $(\varphi + \psi)(x)$ converges to the minimal value of $\varphi + \psi$,
3. $\langle \dot{x}, \dot{v} \rangle \geq 0$,
4. $\dot{x} \in L^2(0, \infty)$.

We want to show that

$$v(t) + \nabla\psi(x(t)) \rightarrow 0 \quad \text{as } t \rightarrow \infty. \quad (2.2.84)$$

Since $x(\cdot)$ is bounded, $\nabla\psi$ is L -Lipschitz continuous (for some positive L) on a closed ball which contains $x(\cdot)$. In particular $t \mapsto \nabla\psi(x(t))$ is absolutely continuous on bounded intervals, and for almost all $t \geq 0$

$$\left\| \frac{d}{dt} \nabla\psi(x) \right\| \leq L \|\dot{x}\|, \quad \left\langle \frac{d}{dt} \nabla\psi(x), \dot{x} \right\rangle \geq 0 \quad (2.2.85)$$

where the second inequality follows from the convexity of ψ . Therefore, for almost all $t \geq 0$

$$\begin{aligned} \frac{d}{dt} \frac{1}{2} \|v + \nabla\psi(x)\|^2 &= \left\langle \dot{v} + \frac{d}{dt} \nabla\psi(x), v + \nabla\psi(x) \right\rangle \\ &= \left\langle \dot{v} + \frac{d}{dt} \nabla\psi(x), -\lambda\dot{x} - \dot{v} \right\rangle \\ &= -\lambda \langle \dot{v}, \dot{x} \rangle - \|\dot{v}\|^2 - \lambda \left\langle \frac{d}{dt} \nabla\psi(x), \dot{x} \right\rangle - \left\langle \frac{d}{dt} \nabla\psi(x), \dot{v} \right\rangle \end{aligned}$$

where the second equality follows from (2.2.45b). Combining the above equation with item 3 and (2.2.85) we conclude that

$$\frac{d}{dt} \frac{1}{2} \|v + \nabla\psi(x)\|^2 \leq -\|\dot{v}\|^2 + L \|\dot{x}\| \|\dot{v}\| \quad (2.2.86)$$

$$\leq -\|\dot{v}\|^2 + \frac{1}{2} (L^2 \|\dot{x}\|^2 + \|\dot{v}\|^2) \quad (2.2.87)$$

$$= -\frac{1}{2} \|\dot{v}\|^2 + \frac{L^2}{2} \|\dot{x}\|^2. \quad (2.2.88)$$

Therefore, using the above inequality, item 4 and the non-negativity of $\|v + \nabla\psi(x)\|$ we conclude that

- a** there exists $\lim_{t \rightarrow \infty} \|v(t) + \nabla\psi(x(t))\|$;
- b** $\|\dot{v}\|^2 \in L^1[0, \infty)$;

where item **a** follows from the above inequality and Lemma 2.2.12. If λ is bounded, by using equation (2.2.45b), $\|v + \nabla\psi(x)\|$ is in L^2 , hence converges to 0. \square

Remark 2.2.17 One can reasonably conjecture that convergence should be strong provided one of the functions is strongly convex, if the sum is even, or if the solution set has nonempty interior. These are the classical assumptions insuring strong convergence of the trajectories of the continuous steepest descent, see [6, 15].

Indeed, under the assumptions of Theorem 2.2.11, we know that for each trajectory x of our system

$$\lim_{t \rightarrow +\infty} (\varphi + \psi)(x(t)) = \inf (\varphi + \psi).$$

This allows to conclude that the property of strong convergence is true in the following cases :

a) If one of the functions is strongly convex, then $\varphi + \psi$ is strongly convex. By a classical result, we know that, in that case, every minimizing sequence converges strongly to the unique minimizer, see [16, Theorem 1.31] for example, whence the result.

b) If one of the functions is boundedly inf-compact, we easily deduce that the trajectory remains in a fixed level set of this function, whence the result.

The general study of the strong convergence property is an interesting subject, but out of the scope of the paper.

2.2.6 Gradient-projection and forward-backward methods

Gradient-Projection Methods

Let us apply the results of the preceding sections to the case $A = \partial\delta_C$, where δ_C is the indicator function of a closed convex nonempty set $C \subset H$; we recall that $\delta_C : H \rightarrow \mathbb{R} \cup \{+\infty\}$ is equal to 0 on C and $+\infty$ elsewhere. The subdifferential of δ_C is the normal cone mapping $x \mapsto A(x) = N_C(x)$. For any $x \in H$, and $\lambda > 0$ the resolvent and Yosida approximation of A are equal to

$$\begin{aligned} J_\lambda^A(x) &= P_C(x) \\ A_\lambda(x) &= \frac{1}{\lambda}(x - P_C(x)), \end{aligned}$$

where P_C is the projection operator onto C (best approximation operator).

The operator $B = \nabla\psi$ is the gradient of a convex, continuously differentiable function $\psi : H \rightarrow \mathbb{R}$, and we assume that $\nabla\psi$ is Lipschitz continuous on bounded sets.

In this setting, the (x, z) dynamical system (2.2.23a)-(2.2.23b), which has been analyzed in Theorem 2.2.4 (recall that $\mu(t) = \frac{1}{\lambda(t)}$), can be written as

$$\begin{cases} x(t) = P_C(z(t)), \\ \dot{z}(t) + \left(1 - \frac{\dot{\mu}(t)}{\mu(t)}\right) (z(t) - P_C(z(t))) + \mu(t)\nabla\psi(P_C(z(t))) = 0. \end{cases} \quad (2.2.89)$$

The results obtained in the preceding section give in this particular situation the following

Theorem 2.2.18 *Let C be a closed convex nonempty set in H , and $\psi : H \rightarrow \mathbb{R}$ a convex, differentiable function whose gradient is Lipschitz continuous on bounded sets. Let $\lambda : [0, +\infty[\rightarrow]0, +\infty[$ be absolutely continuous on each interval $[0, b]$, $0 < b < +\infty$. Set $\mu(t) = \frac{1}{\lambda(t)}$. Then,*

- *for any $z_0 \in H$ there exists a unique global solution (x, z) of (2.2.89) such that $z(0) = z_0$ (and $x(0) = P_C(z_0)$).*
- *for each $t \geq 0$, $x(t) \in C$, and $t \mapsto \psi(x(t))$ is a non-increasing function.*
- *if λ is nonincreasing then $\lim_{t \rightarrow +\infty} \psi(x(t)) = \inf_C(\psi)$.*
- *Suppose that $\int_0^{+\infty} \frac{1}{\lambda(t)} dt = +\infty$, and for almost all $t > 0$, $1 + \frac{\dot{\lambda}(t)}{\lambda(t)} \geq 0$, then*

$$\lim_{t \rightarrow +\infty} \psi(x(t)) = \inf_C(\psi).$$
- *Suppose that $\arg \min_C \psi \neq \emptyset$, $\int_0^{+\infty} \frac{1}{\lambda(t)} dt = +\infty$, and there exists $\varepsilon > 0$ such that, for almost all $t > 0$*

$$1 + \frac{\dot{\lambda}(t)}{\lambda(t)} \geq \varepsilon.$$

Then $x(\cdot)$ converges weakly to a minimizer of ψ on C .

In the case μ constant, we have $\dot{\mu} = 0$, and system (2.2.89) takes the simpler form

$$\begin{cases} x(t) = P_C(z(t)), \\ \dot{z}(t) + z(t) - P_C(z(t)) + \mu \nabla \psi(P_C(z(t))) = 0. \end{cases} \quad (2.2.90)$$

Clearly, all the conditions of Theorem 2.2.18 are satisfied. Hence each trajectory of (2.2.90) is minimizing, and weakly converges to an optimal solution when $\arg \min_C \psi \neq \emptyset$.

Systems (2.2.89) and (2.2.90) are seemingly new. They share some similarities with the continuous gradient-projection method

$$\dot{x}(t) + x(t) - P_C[x(t) - \mu \nabla \psi(x(t))] = 0, \quad (2.2.91)$$

which has been studied by Antipin [17], and Bolte [18]. It is reasonable to conjecture that the time discretization of systems (2.2.89) and (2.2.90) provides new gradient-projection algorithms, which share the good convergence properties of the continuous dynamics. See [1] for a general discussion concerning the link between continuous and discrete dynamics, from the point of view of their convergence. For example, time discretization of (2.2.90) with step size $h_k > 0$ gives

$$\begin{cases} x_k = P_C(z_k), \\ \frac{z_{k+1} - z_k}{h_k} + z_k - P_C(z_k) + \mu \nabla \psi(P_C(z_k)) = 0. \end{cases} \quad (2.2.92)$$

Equivalently

$$\begin{cases} x_k = P_C(z_k), \\ z_{k+1} = (1 - h_k)z_k + h_k(x_k - \mu \nabla \psi(x_k)). \end{cases} \quad (2.2.93)$$

By taking $h_k = 1$, we recover the classical gradient-projection method

$$x_{k+1} = P_C(x_k - \mu \nabla \psi(x_k)). \quad (2.2.94)$$

Forward-Backward Methods

Consider now the more general situation where the operator $A = \partial \varphi$ is the subdifferential of a convex lower semicontinuous proper function $\varphi : H \rightarrow \mathbb{R} \cup \{+\infty\}$. For any $x \in H$ and

$\mu > 0$, the resolvent and Yosida approximation of A are given by

$$\begin{aligned} J_\mu^A(x) &= \text{prox}_{\mu\phi}(x) \\ A_\mu(x) &= \frac{1}{\mu}(x - \text{prox}_{\mu\phi}(x)), \end{aligned}$$

where $\text{prox}_{\mu\phi}$ is the proximal mapping of $\mu\phi$, namely

$$\text{prox}_{\mu\phi}(x) = \arg \min \left\{ \mu\phi(\xi) + \frac{1}{2}\|x - \xi\|^2 : \xi \in H \right\}.$$

The operator $B = \nabla\psi$ is the gradient of a convex, continuously differentiable function $\psi : H \rightarrow \mathbb{R}$, and we assume that $\nabla\psi$ is Lipschitz continuous on bounded sets.

In this context, we make an analysis very similar to the gradient-projection method. For convenience of the reader, we give a self-contained formulation.

The (x, z) dynamical system (2.2.23a)-(2.2.23b), which has been analyzed in Theorem 2.2.4 (recall that $\mu(t) = \frac{1}{\lambda(t)}$), can be written as

$$\begin{cases} x(t) = \text{prox}_{\mu(t)\phi}(z(t)), \\ \dot{z}(t) + \left(1 - \frac{\dot{\mu}(t)}{\mu(t)}\right) (z(t) - \text{prox}_{\mu(t)\phi}(z(t))) + \mu(t)\nabla\psi(\text{prox}_{\mu(t)\phi}(z(t))) = 0 \end{cases} \quad (2.2.95)$$

Theorem 2.2.19 *Let $\varphi : H \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex lower semicontinuous proper function, and $\psi : H \rightarrow \mathbb{R}$ a convex, differentiable function whose gradient is Lipschitz continuous on bounded sets. Let $\lambda : [0, +\infty[\rightarrow]0, +\infty[$ be absolutely continuous on each interval $[0, b]$, $0 < b < +\infty$. Set $\mu(t) = \frac{1}{\lambda(t)}$. Then,*

- for any $z_0 \in H$ there exists a unique global solution (x, z) of (2.2.95) such that $z(0) = z_0$ (and $x(0) = \text{prox}_{\mu(0)\phi}(z_0)$).
- $t \mapsto \phi(x(t)) + \psi(x(t))$ is a non-increasing function.
- if λ is nonincreasing then $\lim_{t \rightarrow +\infty} (\phi(x(t)) + \psi(x(t))) = \inf_H(\phi + \psi)$.
- Suppose that $\int_0^{+\infty} \frac{1}{\lambda(t)} dt = +\infty$ and for almost all $t > 0$ $1 + \frac{\dot{\lambda}(t)}{\lambda(t)} \geq 0$, then $\lim_{t \rightarrow +\infty} (\phi(x(t)) + \psi(x(t))) = \inf_H(\phi + \psi)$.
- Suppose that $\arg \min_H(\phi + \psi) \neq \emptyset$, $\int_0^{+\infty} \frac{1}{\lambda(t)} dt = +\infty$ and there exists $\varepsilon > 0$ such that, for almost all $t > 0$

$$1 + \frac{\dot{\lambda}(t)}{\lambda(t)} \geq \varepsilon.$$

Then $x(\cdot)$ converges weakly to a minimizer of $\phi + \psi$ on H .

In the case μ constant, we have $\dot{\mu} = 0$, and system (2.2.95) takes the simpler form

$$\begin{cases} x(t) = \text{prox}_{\mu\phi}(z(t)), \\ \dot{z}(t) + z(t) - \text{prox}_{\mu\phi}(z(t)) + \mu\nabla\psi(\text{prox}_{\mu\phi}(z(t))) = 0. \end{cases} \quad (2.2.96)$$

All the conditions of Theorem 2.2.19 are satisfied. Hence each trajectory of (2.2.96) is minimizing, and weakly converges to an optimal solution when $\arg \min_H(\phi + \psi) \neq \emptyset$.

As an interesting feature, time discretization of systems (2.2.95) and (2.2.96) suggest new forward-backward algorithms. For example, time discretization of (2.2.96) with step size h_k gives

$$\begin{cases} x_k = \text{prox}_{\mu\phi}(z_k), \\ \frac{z_{k+1} - z_k}{h_k} + z_k - \text{prox}_{\mu\phi}(z_k) + \mu\nabla\psi(\text{prox}_{\mu\phi}(z_k)) = 0. \end{cases} \quad (2.2.97)$$

Equivalently

$$\begin{cases} x_k = \text{prox}_{\mu\phi}(z_k), \\ z_{k+1} = (1 - h_k)z_k + h_k(x_k - \mu\nabla\psi(x_k)). \end{cases} \quad (2.2.98)$$

By taking $h_k = 1$, we recover the classical forward-backward algorithm

$$x_{k+1} = \text{prox}_{\mu\phi}(x_k - \mu\nabla\psi(x_k)). \quad (2.2.99)$$

Algorithm (2.2.98) has some similarities with the classical relaxed forward-backward algorithm (see [7, Theorem 25.8]) :

$$\begin{cases} y_k = x_k - \mu\nabla\Psi(x_k), \\ x_{k+1} = (1 - h_k)x_k + h_k\text{prox}_{\mu\phi}(y_k). \end{cases} \quad (2.2.100)$$

It involves the same basic blocks, but it differs in the order of the operations : in (2.2.98) the relaxation is done inside the proximal loop, while, in the classical one, the relaxation is done outside. Of course, when $\text{prox}_{\mu\phi}$ is linear, the two algorithms coincide.

It is reasonable to expect that algorithm (2.2.98) (and its companions, with varying regularization parameter μ) share the good convergence properties of the continuous dynamics. Their detailed analysis is beyond the scope of this article. It is an interesting research topic for the future.

2.2.7 Conclusions

We introduced a Newton-like continuous dynamic which aims at solving structured monotone inclusions in Hilbert spaces. Given a maximal monotone operator $M = A + B$, with A maximal monotone, B monotone and Lipschitz continuous, the dynamic can be rewritten as a classical differential equation involving separately the resolvents of A (implicit), and evaluations of B (explicit). In the convex subdifferential case, with a Levenberg-Marquardt regularization parameter that is allowed to tend to zero (roughly speaking no faster than e^{-t}), we prove a descent minimizing property, and weak convergence to equilibria of the trajectories. The nice convergence properties of the continuous dynamics may serve as a guideline and an incentive to study some related questions for discrete dynamics. Indeed, time discretization of these continuous dynamics naturally suggests new explicit-implicit (forward-backward) algorithms, and gradient-projection methods for solving structured monotone inclusions.

Further research may also concern :

- (a) Convergence analysis with $A = \partial\phi$, and B a cocoercive operator (see [19], [20]), for example $B = I - T$ with T a nonexpansive map.
- (b) Study of the closed-loop regularized dynamic with two operators A and B , in the line of Attouch-Redont-Svaiter [3].
- (c) Compare and possibly make connection with some related second-order differential systems involving inertia and Hessian driven damping, see [21], [22], with in view accelerated methods, like Beck-Teboulle [4].
- (d) Application to signal/imaging, sparse optimization, and PDE's.

Bibliographie

- [1] Peypouquet, J., Sorin, S. : Evolution equations for maximal monotone operators : asymptotic analysis in continuous and discrete time. *J. of Convex Analysis* **17**, 1113-1163 (2010).
- [2] Attouch, H., Svaiter, B.F. : A continuous dynamical Newton-like approach to solving monotone inclusions. *SIAM J. Control Optim.* **49**, 574–598 (2011).
- [3] Attouch, H., Redont, P., Svaiter, B.F. : Global convergence of a closed-loop regularized Newton method for solving monotone inclusions in Hilbert spaces. *J. Optim. Theory Appl.* **157**, issue 3, 624–650 (2013).
- [4] Beck, A., Teboulle, M. : Gradient-based algorithms with applications in signal recovery problems. In : Palomar, D., Eldar, Y. (eds.) : *Convex Optimization in Signal Processing and Communications*, pp. 33–88. Cambridge University Press, (2010).
- [5] Combettes, P.L., Pesquet, J.-C. : Proximal splitting methods in signal processing. In : Bauschke, H., Burachik, R., Combettes, P.L., Elser, V., Luke, D.R., Wolkowicz, H. (eds.) : *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pp. 185-212. Springer, New York, (2011).
- [6] Brézis, H. : *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert*. North-Holland/Elsevier, New-York (1973).
- [7] Bauschke, H., Combettes, P.L. : *Convex Analysis and Monotone Operator Theory*. CMS books in Mathematics, Springer (2011).
- [8] Zeidler, E. : *Nonlinear Functional Analysis and its Applications, Part II : Monotone Operators*. Springer-Verlag, New-York (1990).
- [9] Minty, G.J. : Monotone (nonlinear) operators in Hilbert spaces. *Duke Mathematical Journal* **29**, 341–346 (1962).
- [10] Rockafellar, R.T. : Maximal monotone relations and the second derivatives of nonsmooth functions. *Ann. Inst. Henri Poincaré* **2**, 167–184 (1985).
- [11] Haraux, A. : *Systèmes dynamiques dissipatifs et applications*. RMA 17, Masson, Paris, (1991).
- [12] Sontag, E.D. : *Mathematical Control Theory*, second edition. Springer-Verlag, New-York (1998).
- [13] Brézis, H. : *Analyse Fonctionnelle*. Masson, Paris (1983).
- [14] Opial, Z. : Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Amer. Math. Soc.* **73**, 591–597 (1967).
- [15] Bruck, R. E. : Asymptotic convergence of nonlinear contraction semigroups in Hilbert spaces. *J. Funct. Anal.* **18**, 15-26 (1975).

- [16] Bonnans, F. : Optimisation Continue, Mathématiques appliquées pour le Master/SMAL, Dunod, Paris, (2006).
- [17] Antipin, A.S. : Minimization of convex functions on convex sets by means of differential equations. *Differential Equations*. **30**, 1365–1375 (1994).
- [18] Bolte, J. : Continuous gradient projection method in Hilbert spaces. *J. Optim. Theory Appl.* **119**, 235–259 (2003).
- [19] Attouch, H., Briceno, L., Combettes, P.L. : A parallel splitting method for coupled monotone inclusions. *SIAM J. Control Optim.* **48**, 3246–3270 (2010).
- [20] Zhu, D.L., Marcotte, P. : Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities. *J. Optim.* **6**, 714—726 (1996).
- [21] Attouch, H., Maingé, P.-E., Redont, P. : A second-order differential system with Hessian-driven damping ; application to non-elastic shock laws. *Differ. Equ. Appl.* **4**, 27–65 (2012).
- [22] Alvarez, F., Attouch, H., Bolte, J., Redont, P. : A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics. *J. Math. Pures Appl.* **81**, 747–779 (2002).

Chapitre 3

Stabilité de la méthode de Newton régularisée avec deux potentiels

3.1 Description et résultats principaux :

Dans ce chapitre, nous étudions la stabilité structurelle de la méthode de Newton régularisée avec deux potentiels. Rappelons que cette méthode, introduite dans le chapitre précédent, vise à résoudre des inclusions monotones structurées. Le terme de régularisation de type Levenberg-Marquard agit en boucle ouverte. Dans ce qui suit, on se place dans le cas où les opérateurs dérivent de potentiels convexes, le problème d'équilibre correspondant est le problème de minimisation suivant

Problème 3.1.1

$$\min \{ \varphi(x) + \psi(x) : x \in H \} .$$

Dans ce problème :

- i)* $\varphi : H \rightarrow \mathbb{R} \cup \{+\infty\}$ est une fonction convexe, semi-continue inférieurement, et propre.
- ii)* $\psi : H \rightarrow \mathbb{R}$ est une fonction convexe, continue et différentiable, et $\nabla\psi$ est localement Lipschitzien.
- iii)* $\varphi + \psi$ est minorée sur H .

Dans Abbas, Attouch, et Svaiter, JOTA (2014), afin de résoudre le problème (3.1.1), on a introduit le système dynamique continu de type Newton régularisé suivant :

$$v(t) \in \partial\varphi(x(t)) \tag{3.1.1}$$

$$\lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) + \nabla\psi(x(t)) = 0, \tag{3.1.2}$$

$$x(0) = x_0, v(0) = v_0, v_0 \in \partial\varphi(x_0), \tag{3.1.3}$$

avec les hypothèses suivantes sur la fonction $\lambda(\cdot)$

$$\lambda : [0, +\infty[\rightarrow]0, +\infty[\text{ est continue,} \tag{3.1.4}$$

$$\text{et absolument continue sur tout intervalle } [0, b], 0 < b < +\infty. \tag{3.1.5}$$

On désignera les bornes inférieure et supérieures de $\lambda(\cdot)$ sur $[0, b]$ par

$$\forall t \in [0, b], \quad 0 < \lambda_{b, \min} \leq \lambda(t) \leq \lambda_{b, \max} < +\infty. \quad (3.1.6)$$

Nous avons vu dans le chapitre précédent qu'il existe une solution unique globale forte du problème de Cauchy (3.1.1)-(3.1.3), et que lorsque $\lambda(\cdot)$ ne tend pas trop vite vers zéro, toute trajectoire $x(\cdot)$ converge faiblement vers une solution du problème de minimisation (3.1.1).

A noter qu'il est important, pour des motifs numériques, d'étudier la stabilité des trajectoires du problème de Cauchy (3.1.1)-(3.1.3) sous l'effet de perturbations portant sur les données. Nous porterons une attention particulière aux perturbations portant sur les données initiales, et aux perturbations portant sur le terme de régularisation $\lambda(\cdot)$. Ce dernier joue un rôle essentiel dans l'analyse de la convergence des trajectoires vers des équilibres. A noter que d'autres types de perturbations seraient aussi intéressants à considérer, comme les perturbations portant sur les potentiels φ et ψ , un sujet d'étude pour le futur.

Nous établissons tout d'abord des estimations du type énergie pour les trajectoires.

Dans le théorème suivant 3.1.2, nous établissons la dépendance lipschitzienne des trajectoires par rapport au terme de régularisation $\lambda(\cdot)$.

Théorème 3.1.2 *Supposons que $\lambda, \eta : [0, T] \rightarrow [c_0, +\infty[$ sont deux fonctions absolument continues, avec $T > 0$, et il existe une constante positive c_0 tel que $0 < c_0 \leq \lambda(t)$, pour tout $t \in [0, T]$. Soit $(x, v), (y, w) : [0, T] \rightarrow H \times H$ les solutions respectives des inclusions différentielles :*

$$\begin{aligned} \lambda \dot{x} + \dot{v} + v + \nabla \psi(x) &= 0, \quad v \in \partial \varphi(x), \quad x(0) = x_0, \quad v(0) = v_0 \\ \eta \dot{y} + \dot{w} + w + \nabla \psi(y) &= 0, \quad w \in \partial \varphi(y), \quad y(0) = y_0, \quad w(0) = w_0. \end{aligned}$$

Soit $\theta : [0, T] \rightarrow \mathbb{R}$, définie pour tout $t \in [0, T]$ par

$$\theta(t) = \sqrt{c_0^2 \|x(t) - y(t)\|^2 + \|v(t) - w(t)\|^2}.$$

Alors

$$\|\theta\|_{L^\infty([0, T])} \leq \left[\frac{\lambda(0) + \eta(0)}{2} \|x_0 - y_0\| + \|v_0 - w_0\| + \frac{C}{2} \|\lambda - \eta\|_{L^1([0, T])} \right] \times \exp \left(\frac{\|\dot{\lambda} + \dot{\eta}\|_{L^1}}{2c_0} + T \left(1 + \frac{L_\psi}{c_0} \right) \right), \quad (3.1.7)$$

avec,

$$\begin{aligned} C &= \frac{\|v_0\| + \|w_0\|}{c_0} + \frac{1 + \sqrt{2T}}{c_0} (\|\nabla \psi(x_0)\| + \|\nabla \psi(y_0)\|) + \\ &\frac{(\sqrt{2T} + \sqrt{T})L_\psi}{c_0^{\frac{3}{2}}} \left(\left((\varphi + \psi)(x_0) - \inf_H (\varphi + \psi) \right)^{\frac{1}{2}} + \left((\varphi + \psi)(y_0) - \inf_H (\varphi + \psi) \right)^{\frac{1}{2}} \right). \end{aligned}$$

En particulier, si $x_0 = y_0, v_0 = w_0$, alors

$$\|\theta\|_{L^\infty([0, T])} \leq \frac{C}{2} \|\lambda - \eta\|_{L^1([0, T])} \times \exp \left(\frac{\|\dot{\lambda} + \dot{\eta}\|_{L^1([0, T])}}{2c_0} + T \left(1 + \frac{L_\psi}{c_0} \right) \right), \quad (3.1.8)$$

avec

$$C = \frac{2\|v_0\|}{c_0} + 2 \frac{1 + \sqrt{2T}}{c_0} \|\nabla \psi(x_0)\| + 2 \frac{(\sqrt{2T} + \sqrt{T})L_\psi}{c_0^{\frac{3}{2}}} \left((\varphi + \psi)(x_0) - \inf_H (\varphi + \psi) \right)^{\frac{1}{2}}.$$

Par un argument de prolongement par continuité, le résultat précédent nous permet d'étendre notre étude au cas où le coefficient $\lambda(\cdot)$ est à variation bornée. A cet effet, pour étudier l'existence et l'unicité, nous utilisons une méthode de régularisation par convolution, qui permet de se ramener au cas $\lambda(\cdot)$ absolument continu, puis on passe à la limite dans les équations.

Dans le théorème suivant nous établissons par cette méthode l'existence et l'unicité d'une solution globale forte du problème (3.1.1)-(3.1.3), dans le cas où λ est une fonction à variation bornée.

Théorème 3.1.3 *Soit $\lambda : [0, T] \rightarrow]0, +\infty[$ à variation bornée sur $[0, T]$. Supposons que $c_0 = \inf_{t \in [0, T]} \lambda(t) > 0$. Soit $x_0 \in \text{dom } \partial\varphi$ et $v_0 \in \partial\varphi(x_0)$, $v_0 \neq 0$. Alors il existe une solution unique globale forte sur $[0, T]$ du problème de Cauchy*

$$v(t) \in \partial\varphi(x(t)), \quad 0 \leq t \leq T \quad (3.1.9)$$

$$\lambda(t) \dot{x}(t) + \dot{v}(t) + v(t) + \nabla\psi(x(t)) = 0, \quad \text{ppt. } 0 \leq t \leq T, \quad (3.1.10)$$

$$x(0) = x_0, \quad v(0) = v_0. \quad (3.1.11)$$

3.2 Article en anglais

Stability of a Regularized Newton Method with two Potentials¹

Abstract

In a Hilbert space setting, we study the stability properties of the regularized continuous Newton method with two potentials, which aims at solving inclusions governed by structured monotone operators. The Levenberg-Marquardt regularization term acts in an open loop way. As a byproduct of our study, we can take the regularization coefficient of bounded variation. These stability results are directly related to the study of numerical algorithms that combine forward-backward and Newton's methods.

3.2.1 Introduction

Throughout this paper H is a real Hilbert space with scalar product $\langle \cdot, \cdot \rangle$ and norm $\|x\| = \sqrt{\langle x, x \rangle}$ for $x \in H$. We make the following standing assumptions : φ and ψ are two functions which act on H and satisfy

- $\varphi : H \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex lower semicontinuous, and proper ;
- $\psi : H \rightarrow \mathbb{R}$ is convex differentiable, and $\nabla\psi$ is Lipschitz continuous on the bounded subsets of H .
- $\varphi + \psi$ is bounded from below on H .

We are concerned with the study of Newton-like continuous and discrete dynamics attached to solving the the structured minimization problem

$$(\mathcal{P}) \quad \min \{\varphi(x) + \psi(x) : x \in H\}.$$

1. B. Abbas, Stability of a Regularized Newton Method with two Potentials, arXiv :1505.05418v1 [math.OC] 20 May 2015

Note the asymmetry between φ , which may be nonsmooth, with extended real values, and ψ which is continuously differentiable, whence the structured property of the above problem. Indeed, we wish to design continuous and discrete dynamics which exploit this particular structure and involve φ via implicit operations (like resolvent or proximal operators) and ψ via explicit operations (typically gradient-like methods). So doing we expect obtaining new forward-backward splitting methods involving Riemannian metric aspects, and which are close to the Newton method. This approach has been delineated in a series of recent papers, [1], [2], [6], [7]. In this paper we are concerned with the stability properties with respect to the data $(\lambda, x_0, v_0, \dots)$ of the strong solutions of the differential inclusion

$$v(t) \in \partial\varphi(x(t)) \quad (3.2.1a)$$

$$\lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) + \nabla\psi(x(t)) = 0 \quad (3.2.1b)$$

$$x(0) = x_0, v(0) = v_0. \quad (3.2.1c)$$

Let us now make our standing assumption on function $\lambda(\cdot)$:

$$\lambda : [0, +\infty[\longrightarrow]0, +\infty[\text{ is continuous,} \quad (3.2.2)$$

$$\text{and absolutely continuous on each interval } [0, b], \quad 0 < b < +\infty. \quad (3.2.3)$$

Hence $\dot{\lambda}(t)$ exists for almost every $t > 0$, and $\dot{\lambda}(\cdot)$ is Lebesgue integrable on each bounded interval $[0, b]$. We stress the fact that we assume $\lambda(t) > 0$, for any $t \geq 0$. By continuity of $\lambda(\cdot)$, this implies that, for any $b > 0$, there exists some positive finite lower and upper bounds for $\lambda(\cdot)$ on $[0, b]$, i.e., for any $t \in [0, b]$

$$0 < \lambda_{b, \min} \leq \lambda(t) \leq \lambda_{b, \max} < +\infty. \quad (3.2.4)$$

Our main interest is to allow $\lambda(t)$ to go to zero as $t \rightarrow +\infty$. This makes the corresponding Levenberg-Marquardt regularization method asymptotically close to the Newton's method. Let us summarize the results obtained in [2]. Under the above assumptions, for any Cauchy data $x_0 \in \text{dom}\partial\varphi$ and $v_0 \in \partial\varphi(x_0)$, there exists a unique strong global solution $(x(\cdot), v(\cdot)) : [0, +\infty[\rightarrow H \times H$ of the Cauchy problem (3.2.1a)-(3.2.1c). Assuming that the solution set is nonempty, if $\lambda(t)$ tends to zero not too fast, as $t \rightarrow +\infty$, then $v(t) \rightarrow 0$ strongly, and $x(t)$ converges weakly to some equilibrium which is a solution of the minimization problem (\mathcal{P}) . By Minty representation of $\partial\varphi$, the solution pair $(x(\cdot), v(\cdot))$ of (3.2.1a)-(3.2.1c) can be represented as follows : set $\mu(t) = \frac{1}{\lambda(t)}$, then for any $t \in [0, +\infty)$,

$$x(t) = \text{prox}_{\mu(t)\varphi}(z(t)); \quad (3.2.5)$$

$$v(t) = \nabla\varphi_{\mu(t)}(z(t)), \quad (3.2.6)$$

where $z(\cdot) : [0, +\infty[\rightarrow H$ is the unique strong global solution of the Cauchy problem

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t))\nabla\varphi_{\mu(t)}(z(t)) + \mu(t)\nabla\psi(\text{prox}_{\mu(t)\varphi}(z(t))) = 0 \quad (3.2.7)$$

$$z(0) = x_0 + \mu(0)v_0. \quad (3.2.8)$$

Let us recall that $\text{prox}_{\mu\varphi}$ is the proximal mapping associated to $\mu\varphi$. Equivalently, $\text{prox}_{\mu\varphi} = (I + \mu\partial\varphi)^{-1}$ is the resolvent of index $\mu > 0$ of the maximal monotone operator $\partial\varphi$, and $\nabla\varphi_{\mu}$ is its Yosida approximation of index $\mu > 0$.

Let us stress the fact that, for each $t > 0$, the operators $\text{prox}_{\mu(t)\varphi} : H \rightarrow H, \nabla\varphi_{\mu(t)} : H \rightarrow H$ are everywhere defined and Lipschitz continuous, which makes this system relevant to the Cauchy–Lipschitz theorem in the nonautonomous case, which naturally suggests good stability results of the solution of (3.2.1a)-(3.2.1c) with respect to the data.

This paper is organized as follows : We first establish a priori energy estimates on the trajectories. Then we consider the case where λ is locally absolutely continuous. Note that it is important, for numerical reasons, to study the stability of the solution with respect to perturbations of the data, and in particular of λ which plays a crucial role in the regularization process. In Theorem 3.2.5 we prove the Lipschitz continuous dependence of the solution with respect to λ . Moreover, the Lipschitz constant only depends on the L^1 norm of the time derivative of λ . Finally, we extend our analysis to the case where λ is a function with bounded variation (possibly involving jumps). We use a regularization by convolution method in order to reduce to the smooth case, and then pass to the limit in the equations. So doing, in Theorem 3.2.11 and Corollary 3.2.12, we prove the existence and uniqueness of a strong solution for (1a)-(1c), in the case where λ is a function with bounded variation.

A priori estimates

The linear space $H \times H$ is equipped with its usual Hilbertian norm $\|(\xi, \zeta)\| = \sqrt{\|\xi\|^2 + \|\zeta\|^2}$. In this section we work on a fixed bounded interval $[0, T]$, and following assumption (3.2.2), we suppose that there exists some positive constant c_0 such that

$$0 < c_0 \leq \lambda(t) \quad \text{for all } t \in [0, T]. \quad (3.2.9)$$

We will also assume that $\nabla\psi$ is L_ψ -Lipschitz continuous. Indeed, this is not a restrictive assumption since one can reduce the study to trajectories belonging to a fixed ball in H . We will often omit the time variable t and write x, v, \dots for $x(t), v(t), \dots$ when no ambiguity arises. In the following two Propositions we denote by $(x(\cdot), v(\cdot)) : [0, +\infty) \rightarrow H \times H$ the unique strong global solution of the Cauchy problem (3.2.1a)-(3.2.1c).

Proposition 3.2.1 *Let (x, v) be the strong solution of system (3.2.1a)-(3.2.1c) on $[0, T]$, $T > 0$. Then*

$$\int_0^T \|\dot{x}(t)\|^2 \leq \frac{1}{c_0} \left((\varphi + \psi)(x_0) - \inf_H (\varphi + \psi) \right). \quad (3.2.10)$$

$$\|x\|_{L^\infty(0, T; H)} \leq \|x_0\| + \sqrt{\frac{T}{c_0}} \left((\varphi + \psi)(x_0) - \inf_H (\varphi + \psi) \right)^{\frac{1}{2}}. \quad (3.2.11)$$

Proof. a) For almost every $t > 0$, $\dot{x}(t)$ and $\dot{v}(t)$ are well defined, thus

$$\langle \dot{x}(t), \dot{v}(t) \rangle = \lim_{h \rightarrow 0} \frac{1}{h^2} \langle x(t+h) - x(t), v(t+h) - v(t) \rangle.$$

By equation (3.2.1a), we have $v(t) \in \partial\varphi(x(t))$. Since $\partial\varphi : H \rightrightarrows H$ is monotone

$$\langle x(t+h) - x(t), v(t+h) - v(t) \rangle \geq 0.$$

Dividing by h^2 and passing to the limit preserves the inequality, which yields

$$\langle \dot{x}(t), \dot{v}(t) \rangle \geq 0. \quad (3.2.12)$$

By taking the inner product of both sides of (3.2.1b) by $\dot{x}(t)$ we obtain

$$\lambda(t)\|\dot{x}(t)\|^2 + \langle \dot{x}(t), \dot{v}(t) \rangle + \langle v(t), \dot{x}(t) \rangle + \langle \nabla\psi(x(t)), \dot{x}(t) \rangle = 0.$$

By (3.2.12) we infer

$$\lambda(t)\|\dot{x}(t)\|^2 + \langle v(t), \dot{x}(t) \rangle + \langle \nabla\psi(x(t)), \dot{x}(t) \rangle \leq 0. \quad (3.2.13)$$

Noticing that x and v are continuous on $[0, T]$, hence bounded, and that λ is bounded from below on $[0, T]$ by a positive number, one easily gets from (3.2.13) that

$$\dot{x} \in L^2(0, T; H). \quad (3.2.14)$$

For our stability analysis, we now establish a precise estimate of the L^2 norm of \dot{x} . By the classical derivation chain rule

$$\frac{d}{dt}\psi(x(t)) = \langle \nabla\psi(x(t)), \dot{x}(t) \rangle. \quad (3.2.15)$$

We appeal to a similar formula which is still valid for a convex lower semicontinuous proper function $\varphi : H \rightarrow \mathbb{R} \cup \{+\infty\}$. Notice that

- i*) $v(t) \in \partial\varphi(x(t))$ for almost every $t \in [0, T]$;
- ii*) v is continuous on $[0, T]$, and hence belongs to $L^2(0, T; H)$;
- iii*) $\dot{x} \in L^2(0, T; H)$ by (3.2.14).

By *i*), *ii*), *iii*), conditions of [11, Lemma 3.3] are satisfied, which allows to deduce that $t \mapsto \varphi(x(t))$ is absolutely continuous on $[0, T]$, and for almost $t \in [0, T]$,

$$\frac{d}{dt}\varphi(x(t)) = \langle v(t), \dot{x}(t) \rangle. \quad (3.2.16)$$

Combining (3.2.13) with (3.2.15) and (3.2.16) we obtain

$$\lambda(t)\|\dot{x}(t)\|^2 + \frac{d}{dt}(\varphi(x(t)) + \psi(x(t))) \leq 0. \quad (3.2.17)$$

By integrating the above inequality from 0 to T we obtain

$$\int_0^T \lambda(t)\|\dot{x}(t)\|^2 dt + (\varphi + \psi)(x(T)) \leq (\varphi + \psi)(x(0)). \quad (3.2.18)$$

Since $\varphi + \psi$ is bounded from below on H , and λ is minorized by the positive constant c_0 on $[0, T]$ (see (3.2.9)), we infer

$$\int_0^T \|\dot{x}(t)\|^2 dt \leq \frac{1}{c_0} \left((\varphi + \psi)(x_0) - \inf_H(\varphi + \psi) \right). \quad (3.2.19)$$

b) Since $x(\cdot)$ is absolutely continuous on bounded sets, for any $t \in [0, T]$

$$x(t) = x_0 + \int_0^t \dot{x}(\tau) d\tau. \quad (3.2.20)$$

Passing to the norm, and using Cauchy-Schwarz inequality yields

$$\|x(t)\| \leq \|x_0\| + \int_0^t \|\dot{x}(\tau)\| d\tau \quad (3.2.21)$$

$$\leq \|x_0\| + \sqrt{T} \left(\int_0^T \|\dot{x}(\tau)\|^2 d\tau \right)^{\frac{1}{2}}. \quad (3.2.22)$$

Combining the above inequality with (3.2.19) gives

$$\|x(t)\| \leq \|x_0\| + \sqrt{\frac{T}{c_0}} \left((\varphi + \psi)(x_0) - \inf_H(\varphi + \psi) \right)^{\frac{1}{2}}. \quad (3.2.23)$$

This being true for any $t \in [0, T]$

$$\|x\|_{L^\infty(0,T;H)} \leq \|x_0\| + \sqrt{\frac{T}{c_0}} \left((\varphi + \psi)(x_0) - \inf_H(\varphi + \psi) \right)^{\frac{1}{2}}. \quad (3.2.24)$$

□

Let us now exploit another a priori energy estimate.

Proposition 3.2.2 *Let (x, v) be the strong solution of system (3.2.1a)-(3.2.1c) on $[0, T]$, $T > 0$. Then*

$$\int_0^T \|\dot{v}(t)\|^2 dt \leq \|v_0\|^2 + 2T \|\nabla\psi(x_0)\|^2 + \frac{2T^2 L_\psi^2}{c_0} \left((\varphi + \psi)(x_0) - \inf_H(\varphi + \psi) \right) \quad (3.2.25)$$

$$\|v\|_{L^\infty(0,T;H)} \leq \|v_0\| + \sqrt{2T} \|\nabla\psi(x_0)\| + \sqrt{\frac{2}{c_0}} T L_\psi \left((\varphi + \psi)(x_0) - \inf_H(\varphi + \psi) \right)^{\frac{1}{2}}. \quad (3.2.26)$$

Proof. By taking the scalar product of (3.2.1b) by $\dot{v}(t)$ we obtain

$$\lambda(t) \langle \dot{x}(t), \dot{v}(t) \rangle + \|\dot{v}(t)\|^2 + \langle v(t), \dot{v}(t) \rangle + \langle \nabla\psi(x(t)), \dot{v}(t) \rangle = 0.$$

By (3.2.12) we infer

$$\|\dot{v}(t)\|^2 + \langle v(t), \dot{v}(t) \rangle + \langle \nabla\psi(x(t)), \dot{v}(t) \rangle \leq 0. \quad (3.2.27)$$

Hence

$$\|\dot{v}(t)\|^2 + \frac{1}{2} \frac{d}{dt} \|v(t)\|^2 \leq \|\nabla\psi(x(t))\| \|\dot{v}(t)\| \quad (3.2.28)$$

$$\leq \frac{1}{2} \|\nabla\psi(x(t))\|^2 + \frac{1}{2} \|\dot{v}(t)\|^2 \quad (3.2.29)$$

which implies

$$\|\dot{v}(t)\|^2 + \frac{d}{dt} \|v(t)\|^2 \leq \|\nabla\psi(x(t))\|^2. \quad (3.2.30)$$

By integrating the above inequality we deduce that, for any $t \in [0, T]$

$$\int_0^t \|\dot{v}(\tau)\|^2 d\tau + \|v(t)\|^2 \leq \|v_0\|^2 + \int_0^T \|\nabla\psi(x(\tau))\|^2 d\tau. \quad (3.2.31)$$

Since $\nabla\psi$ is L_ψ -Lipschitz continuous

$$\|\nabla\psi(x(\tau))\| \leq \|\nabla\psi(x_0)\| + L_\psi\|x(\tau) - x_0\|. \quad (3.2.32)$$

A careful look at the proof of (3.2.24) gives the more precise estimate

$$\|x - x_0\|_{L^\infty(0,T;H)} \leq \sqrt{\frac{T}{c_0}} \left((\varphi + \psi)(x_0) - \inf_H(\varphi + \psi) \right)^{\frac{1}{2}} \quad (3.2.33)$$

Hence for any $\tau \in [0, T]$

$$\|\nabla\psi(x(\tau))\| \leq \|\nabla\psi(x_0)\| + L_\psi\sqrt{\frac{T}{c_0}} \left((\varphi + \psi)(x_0) - \inf_H(\varphi + \psi) \right)^{\frac{1}{2}}. \quad (3.2.34)$$

Combining (3.2.31) with (3.2.34) gives

$$\int_0^t \|\dot{v}(\tau)\|^2 d\tau + \|v(t)\|^2 \leq \|v_0\|^2 + 2T\|\nabla\psi(x_0)\|^2 + \frac{2T^2L_\psi^2}{c_0} \left((\varphi + \psi)(x_0) - \inf_H(\varphi + \psi) \right). \quad (3.2.35)$$

As a consequence

$$\int_0^T \|\dot{v}(t)\|^2 dt \leq \|v_0\|^2 + 2T\|\nabla\psi(x_0)\|^2 + \frac{2T^2L_\psi^2}{c_0} \left((\varphi + \psi)(x_0) - \inf_H(\varphi + \psi) \right), \quad (3.2.36)$$

and

$$\|v\|_{L^\infty(0,T;H)} \leq \|v_0\| + \sqrt{2T}\|\nabla\psi(x_0)\| + \sqrt{\frac{2}{c_0}}TL_\psi \left((\varphi + \psi)(x_0) - \inf_H(\varphi + \psi) \right)^{\frac{1}{2}}. \quad (3.2.37)$$

which ends the proof. \square

We can now deduce from the two preceding propositions an a priori bound on the L^∞ norm of \dot{x} and \dot{v} .

Proposition 3.2.3

$$\|\dot{x}\|_{L^\infty(0,T;H)} \leq \frac{\|v_0\|}{c_0} + \frac{1 + \sqrt{2T}}{c_0}\|\nabla\psi(x_0)\| + \frac{(\sqrt{2T} + \sqrt{T})L_\psi}{c_0^{\frac{3}{2}}} \left((\varphi + \psi)(x_0) - \inf_H(\varphi + \psi) \right)^{\frac{1}{2}} \quad (3.2.38)$$

$$\|\dot{v}\|_{L^\infty(0,T;H)} \leq \|v_0\| + (1 + \sqrt{2T})\|\nabla\psi(x_0)\| + \frac{(\sqrt{2T} + \sqrt{T})L_\psi}{c_0^{\frac{1}{2}}} \left((\varphi + \psi)(x_0) - \inf_H(\varphi + \psi) \right)^{\frac{1}{2}} \quad (3.2.39)$$

Proof. a) Let us return to the equation obtained by taking the inner product of both sides of (3.2.1b) by $\dot{x}(t)$

$$\lambda(t)\|\dot{x}(t)\|^2 + \langle \dot{x}(t), \dot{v}(t) \rangle + \langle v(t), \dot{x}(t) \rangle + \langle \nabla\psi(x(t)), \dot{x}(t) \rangle = 0.$$

By (3.2.12), and λ is minorized by the positive constant c_0 on $[0, T]$, we infer

$$c_0 \|\dot{x}(t)\|^2 + \langle v(t) + \nabla\psi(x(t)), \dot{x}(t) \rangle \leq 0.$$

Hence

$$\begin{aligned} \|\dot{x}(t)\| &\leq \frac{1}{c_0} \|v(t) + \nabla\psi(x(t))\| \\ &\leq \frac{1}{c_0} (\|v(t)\| + \|\nabla\psi(x_0)\| + L_\psi \|x(t) - x_0\|). \end{aligned}$$

By (3.2.33) and (3.2.37) we deduce that

$$\begin{aligned} \|\dot{x}\|_{L^\infty(0,T;H)} &\leq \frac{1}{c_0} \left(\|v_0\| + \sqrt{2T} \|\nabla\psi(x_0)\| + \sqrt{\frac{2}{c_0}} T L_\psi \left((\varphi + \psi)(x_0) - \inf_H (\varphi + \psi) \right)^{\frac{1}{2}} \right) \\ &\quad + \frac{1}{c_0} \|\nabla\psi(x_0)\| + \frac{L_\psi}{c_0} \sqrt{\frac{T}{c_0}} \left((\varphi + \psi)(x_0) - \inf_H (\varphi + \psi) \right)^{\frac{1}{2}} \\ &\leq \frac{\|v_0\|}{c_0} + \frac{1 + \sqrt{2T}}{c_0} \|\nabla\psi(x_0)\| + \frac{(\sqrt{2T} + \sqrt{T}) L_\psi}{c_0^{\frac{3}{2}}} \left((\varphi + \psi)(x_0) - \inf_H (\varphi + \psi) \right)^{\frac{1}{2}}. \end{aligned}$$

b) Let us return to the equation obtained by taking the inner product of both sides of (3.2.1b) by $\dot{v}(t)$

$$\lambda(t) \langle \dot{x}(t), \dot{v}(t) \rangle + \|\dot{v}(t)\|^2 + \langle v(t), \dot{v}(t) \rangle + \langle \nabla\psi(x(t)), \dot{v}(t) \rangle = 0.$$

A similar argument as above yields

$$\|\dot{v}(t)\| \leq \|v(t) + \nabla\psi(x(t))\|$$

from which we deduce the result. □

Let us enunciate some straight consequences of Proposition 3.2.3 .

Corollary 3.2.4 *The following properties hold : for any $0 < T < +\infty$*

1. $t \mapsto x(t)$ is Lipschitz continuous on $[0, T]$ with constant

$$\frac{\|v_0\|}{c_0} + \frac{1 + \sqrt{2T}}{c_0} \|\nabla\psi(x_0)\| + \frac{(\sqrt{2T} + \sqrt{T}) L_\psi}{c_0^{\frac{3}{2}}} \left((\varphi + \psi)(x_0) - \inf_H (\varphi + \psi) \right)^{\frac{1}{2}}$$

2. $t \mapsto v(t)$ is Lipschitz continuous on $[0, T]$, with constant

$$\|v_0\| + (1 + \sqrt{2T}) \|\nabla\psi(x_0)\| + \frac{(\sqrt{2T} + \sqrt{T}) L_\psi}{c_0^{\frac{1}{2}}} \left((\varphi + \psi)(x_0) - \inf_H (\varphi + \psi) \right)^{\frac{1}{2}}$$

3.2.2 Stability results

In the next theorem we analyze the Lipschitz continuous dependence of the solution (x, v) of the Cauchy problem (3.2.1a)-(3.2.1c) with respect to the function λ and the initial point (x_0, v_0) . Our demonstration is based on that followed in [6, Theorem 3.1].

Theorem 3.2.5 *Suppose that $\lambda, \eta : [0, T] \rightarrow [c_0, +\infty[$ are absolutely continuous functions, with $T > 0$ and $c_0 > 0$. Let $(x, v), (y, w) : [0, T] \rightarrow H \times H$ be the respective strong solutions of the inclusions*

$$\lambda \dot{x} + \dot{v} + v + \nabla \psi(x) = 0, \quad v \in \partial \varphi(x), \quad x(0) = x_0, \quad v(0) = v_0, \quad (3.2.40)$$

$$\eta \dot{y} + \dot{w} + w + \nabla \psi(y) = 0, \quad w \in \partial \varphi(y), \quad y(0) = y_0, \quad w(0) = w_0. \quad (3.2.41)$$

Define $\theta : [0, T] \rightarrow \mathbb{R}$, for each $t \in [0, T]$, by

$$\theta(t) = \sqrt{c_0^2 \|x(t) - y(t)\|^2 + \|v(t) - w(t)\|^2}.$$

Then

$$\|\theta\|_{L^\infty([0, T])} \leq \left[\frac{\lambda(0) + \eta(0)}{2} \|x_0 - y_0\| + \|v_0 - w_0\| + \frac{C}{2} \|\lambda - \eta\|_{L^1([0, T])} \right] \times \exp \left(\frac{\|\dot{\lambda} + \dot{\eta}\|_{L^1}}{2c_0} + T \left(1 + \frac{L_\psi}{c_0} \right) \right), \quad (3.2.42)$$

with

$$C = \frac{\|v_0\| + \|w_0\|}{c_0} + \frac{1 + \sqrt{2T}}{c_0} (\|\nabla \psi(x_0)\| + \|\nabla \psi(y_0)\|) + \frac{(\sqrt{2T} + \sqrt{T})L_\psi}{c_0^{\frac{3}{2}}} \left(\left((\varphi + \psi)(x_0) - \inf_H (\varphi + \psi) \right)^{\frac{1}{2}} + \left((\varphi + \psi)(y_0) - \inf_H (\varphi + \psi) \right)^{\frac{1}{2}} \right).$$

In particular, if $x_0 = y_0$, $v_0 = w_0$, then

$$\|\theta\|_{L^\infty([0, T])} \leq \frac{C}{2} \|\lambda - \eta\|_{L^1([0, T])} \times \exp \left(\frac{\|\dot{\lambda} + \dot{\eta}\|_{L^1([0, T])}}{2c_0} + T \left(1 + \frac{L_\psi}{c_0} \right) \right), \quad (3.2.43)$$

with

$$C = \frac{2\|v_0\|}{c_0} + 2 \frac{1 + \sqrt{2T}}{c_0} \|\nabla \psi(x_0)\| + 2 \frac{(\sqrt{2T} + \sqrt{T})L_\psi}{c_0^{\frac{3}{2}}} \left((\varphi + \psi)(x_0) - \inf_H (\varphi + \psi) \right)^{\frac{1}{2}}.$$

Proof. To simplify the exposition, define $\gamma : [0, T] \rightarrow \mathbb{R}$

$$\gamma = \frac{\lambda + \eta}{2}.$$

Using the assumptions $\lambda, \eta \geq c_0$ and the monotonicity of $\partial \varphi$ we conclude that for any $t \in [0, T]$

$$c_0 \leq \gamma, \quad \langle x - y, v - w \rangle \geq 0.$$

Therefore

$$\theta \leq \sqrt{\gamma^2 \|x - y\|^2 + \|v - w\|^2} \leq \|\gamma(x - y) + (v - w)\|. \quad (3.2.44)$$

Let us show that θ satisfies a Gronwall inequality. Let us start from

$$\frac{d}{dt} [\gamma(x - y) + (v - w)] = \dot{\gamma}(x - y) + \gamma(\dot{x} - \dot{y}) + (\dot{v} - \dot{w}).$$

In view of (3.2.40) and (3.2.41),

$$(\dot{v} - \dot{w}) = -\lambda\dot{x} + \eta\dot{y} - (v - w) - (\nabla\psi(x) - \nabla\psi(y)).$$

Combining the two above relations gives

$$\begin{aligned} \frac{d}{dt} [\gamma(x - y) + (v - w)] &= \dot{\gamma}(x - y) + \gamma(\dot{x} - \dot{y}) - \lambda\dot{x} + \eta\dot{y} - (v - w) - (\nabla\psi(x) - \nabla\psi(y)), \\ &= \dot{\gamma}(x - y) + \frac{\eta - \lambda}{2}(\dot{x} + \dot{y}) - (v - w) - (\nabla\psi(x) - \nabla\psi(y)). \end{aligned}$$

Since γ, x, y, w, v are absolutely continuous, the function $\gamma(x - y) + (v - w)$ is also absolutely continuous. As a consequence, integration of the above inequality on $[0, s]$, for $s \in [0, T]$, yields

$$\begin{aligned} &[\gamma(x - y) + (v - w)](s) - [\gamma(x - y) + (v - w)](0) = \\ &\frac{1}{2} \int_0^s (\eta - \lambda)(\dot{x} + \dot{y}) dt + \int_0^s (\dot{\gamma}(x - y) - (v - w)) dt - \int_0^s (\nabla\psi(x) - \nabla\psi(y)) dt. \end{aligned}$$

Passing to the norm

$$\begin{aligned} \|\gamma(x - y) + (v - w)\|(s) &\leq \gamma(0) \|x_0 - y_0\| + \|v_0 - w_0\| \\ &+ \frac{1}{2} \int_0^s |\lambda - \eta| \|\dot{x} + \dot{y}\| dt + \int_0^s \|\dot{\gamma}(x - y) - (v - w)\| dt + \int_0^s \|\nabla\psi(x) - \nabla\psi(y)\| dt. \end{aligned} \quad (3.2.45)$$

By Lipschitz continuity property of $\nabla\psi$, and definition of θ , we have

$$\begin{aligned} \|\nabla\psi(x(t)) - \nabla\psi(y(t))\| &\leq L_\psi \|x(t) - y(t)\| \\ &\leq \frac{L_\psi}{c_0} \theta(t). \end{aligned} \quad (3.2.46)$$

On the other hand

$$\begin{aligned} \|\dot{\gamma}(x - y) - (v - w)\|^2 &= |\dot{\gamma}|^2 \|x - y\|^2 + \|v - w\|^2 + 2 \left\langle c_0(x - y), \frac{\dot{\gamma}}{c_0}(v - w) \right\rangle \\ &\leq |\dot{\gamma}|^2 \|x - y\|^2 + \|v - w\|^2 + c_0^2 \|x - y\|^2 + \frac{|\dot{\gamma}|^2}{c_0^2} \|v - w\|^2 \\ &= \left(\frac{|\dot{\gamma}|^2}{c_0^2} + 1 \right) \theta^2, \end{aligned} \quad (3.2.47)$$

Combining (3.2.45), (3.2.46), (3.2.47) and $\theta \leq \|\gamma(x - y) + (v - w)\|$, we obtain

$$\begin{aligned} \theta(s) &\leq \gamma(0) \|x_0 - y_0\| + \|v_0 - w_0\| + \frac{1}{2} \int_0^s |\lambda - \eta| \|\dot{x} + \dot{y}\| dt + \int_0^s \theta(t) \left[\sqrt{\frac{|\dot{\gamma}|^2}{c_0^2} + 1} + \frac{L_\psi}{c_0} \right] dt \\ &\leq \gamma(0) \|x_0 - y_0\| + \|v_0 - w_0\| + \frac{1}{2} \|\lambda - \eta\|_{L^1([0,T])} \|\dot{x} + \dot{y}\|_{L^\infty([0,T])} + \int_0^s \theta(t) \left[\frac{|\dot{\gamma}|}{c_0} + 1 + \frac{L_\psi}{c_0} \right] dt. \end{aligned}$$

Applying Gronwall's inequality yields

$$\theta(s) \leq \left[\gamma(0) \|x_0 - y_0\| + \|v_0 - w_0\| + \frac{1}{2} \|\lambda - \eta\|_{L^1([0,T])} \|\dot{x} + \dot{y}\|_{L^\infty([0,T])} \right] \times \exp \int_0^s \left[\frac{|\dot{\gamma}|}{c_0} + 1 + \frac{L_\psi}{c_0} \right] dt.$$

Combining this estimation with the bound on the L^∞ norm of \dot{x} and \dot{y} (see Proposition (3.2.3)) gives

$$\|\theta\|_{L^\infty([0,T])} \leq \left[\gamma(0) \|x_0 - y_0\| + \|v_0 - w_0\| + \frac{C}{2} \|\lambda - \eta\|_{L^1([0,T])} \right] \times \exp \int_0^T \left[\frac{|\dot{\gamma}|}{c_0} + 1 + \frac{L_\psi}{c_0} \right] dt$$

with

$$\begin{aligned} C &= \frac{\|v_0\| + \|w_0\|}{c_0} + \frac{1 + \sqrt{2T}}{c_0} (\|\nabla\psi(x_0)\| + \|\nabla\psi(y_0)\|) + \\ &\quad \frac{(\sqrt{2T} + \sqrt{T})L_\psi}{c_0^{\frac{3}{2}}} \left(\left((\varphi + \psi)(x_0) - \inf_H(\varphi + \psi) \right)^{\frac{1}{2}} + \left((\varphi + \psi)(y_0) - \inf_H(\varphi + \psi) \right)^{\frac{1}{2}} \right). \end{aligned}$$

Equivalently

$$\begin{aligned} \|\theta\|_{L^\infty([0,T])} &\leq \left[\frac{\lambda(0) + \eta(0)}{2} \|x_0 - y_0\| + \|v_0 - w_0\| + \frac{C}{2} \|\lambda - \eta\|_{L^1([0,T])} \right] \times \\ &\quad \exp \left(\frac{\|\dot{\lambda} + \dot{\eta}\|_{L^1([0,T])}}{2c_0} + T \left(1 + \frac{L_\psi}{c_0} \right) \right) \end{aligned}$$

□

Remark 3.2.6 In the cas $\psi = 0$ we recover exactly the same stability results as [6, Theorem 3.1].

Remark 3.2.7 It is worth noticing that, besides the Lipschitz continuous dependence with respect to λ of the solution (x, v) of (3.2.1a)–(3.2.1c), Theorem (3.2.5) also provides its continuous dependence with respect to the initial data (x_0, v_0) . More precisely, if (x_n, v_n) (resp. (x, v)) is the solution of (3.2.1a)–(3.2.1c) corresponding to the Cauchy data (x_{0n}, v_{0n}) (resp. (x_0, v_0)), as a direct consequence of (3.2.42) we obtain for all $T > 0$

$$(\partial\varphi \ni (x_{0n}, v_{0n}) \longrightarrow (x_0, v_0)) \implies (x_n, v_n) \longrightarrow (x, v) \text{ uniformly on } [0, T].$$

Note also that, by contrast with semigroup generated by a subdifferential convex [11], there is no regularizing effect on the initial condition : there is no way to define a solution of (3.2.1a)–(3.2.1c) for $x_0 \in \overline{\text{dom}\partial\varphi} \setminus \text{dom}\partial\varphi$ since in that case, for any approximation sequence $x_{0n} \in \text{dom}\partial\varphi$ and $v_{0n} \in \partial\varphi(x_{0n})$, one has $\lim_n \|v_{0n}\| = +\infty$, which would imply the blow-up of the sequence (x_n, v_n) as $n \longrightarrow +\infty$, on any finite interval.

Remark 3.2.8 Another approach is to study the equivalent problem (3.2.7)-(3.2.8), based on the known stability results for the Cauchy-Lipschitz problem. Although conceptually simple, this approach seems more technical.

3.2.3 Bounded Variation Regularization Coefficient $\lambda(\cdot)$

Let us suppose that $\lambda(\cdot) : [0, T] \rightarrow]0, \infty[$ is of bounded variation on $[0, T]$, where $T > 0$. That is $\text{TV}(\lambda, [0, T]) < +\infty$, where $\text{TV}(\lambda, [0, T])$ is the total variation of λ on $[0, T]$:

$$\text{TV}(\lambda, [0, T]) = \sup \sum_{i=1}^p |\lambda(\tau_i) - \lambda(\tau_{i-1})|,$$

the supremum being taken over all $p \in \mathbb{N}$ and all strictly increasing sequences $\tau_0 < \tau_1 < \dots < \tau_p$ of points of $[0, T]$. Function λ may involve jumps. We also suppose that λ is bounded away from 0 :

$$\inf \lambda([0, T]) > 0.$$

The following lemmas which are proved in [6, Lemma 3.1, Lemma 3.2] gather some classical facts concerning the approximation of functions of bounded variation by smooth functions together with some technical results useful for sequel.

Lemma 3.2.9 *Let $\lambda : [0, T] \rightarrow]0, \infty[$ be of bounded variation on $[0, T]$. Then there exists a sequence $(\lambda_n)_{n \in \mathbb{N}}$, with $\lambda_n \in C^\infty([0, T])$ for each $n \in \mathbb{N}$, such that*

$$i) \inf \lambda([0, T]) \leq \lambda_n(t) \leq \sup \lambda([0, T]), \quad \forall t \in [0, T], \quad \forall n \in \mathbb{N}.$$

In particular, $\lambda_n \geq 0$ if $\lambda \geq 0$;

$$ii) \lambda_n \rightarrow \lambda \text{ in } L^p(0, T) \text{ for any } 1 \leq p < \infty;$$

$$iii) \text{TV}(\lambda_n, [0, T]) = \int_0^T |\dot{\lambda}_n(t)| dt \leq \text{TV}(\lambda, [0, T]).$$

Lemma 3.2.10 *Let $z_n, z \in C([0, T], H)$ be such that $z_n \rightarrow z$ uniformly and $(z_n)_n$ is L -Lipschitz continuous for some positive constant L independent of $n \in \mathbb{N}$. Let $\lambda_n \rightarrow \lambda$ in $L^2(0, T)$. Then $\lambda_n \dot{z}_n$ converges weakly to $\lambda \dot{z}$ in $L^2(0, T; H)$.*

We can now state the main result of this section.

Theorem 3.2.11 *Let $\lambda : [0, T] \rightarrow]0, \infty[$ be of bounded variation on $[0, T]$, and suppose $c_0 = \inf \lambda([0, T]) > 0$. Let $x_0 \in \text{dom} \partial \varphi$ and $v_0 \in \partial \varphi(x_0)$, $v_0 \neq 0$. Then there is existence and uniqueness of a strong solution $(x, v) : [0, T] \rightarrow H \times H$ of the Cauchy problem*

$$v(t) \in \partial \varphi(x(t)), \quad 0 \leq t \leq T \tag{3.2.48a}$$

$$\lambda(t) \dot{x}(t) + \dot{v}(t) + v(t) + \nabla \psi(x(t)) = 0, \quad \text{a.e. } 0 \leq t \leq T, \tag{3.2.48b}$$

$$x(0) = x_0, \quad v(0) = v_0. \tag{3.2.48c}$$

Proof. Existence : According to Lemma 3.2.9, there exists a sequence of functions $(\lambda_n)_{n \in \mathbb{N}}$ in $C^\infty([0, T])$, with $\lambda_n(t) \geq c_0$, which converges to λ in $L^p(0, T)$ for $p \geq 1$, and satisfies the condition

$$\int_0^T |\dot{\lambda}_n(t)| dt \leq \text{TV}(\lambda). \quad (3.2.49)$$

For each $n \in \mathbb{N}$ there exists a unique (x_n, v_n) solution of the differential inclusion

$$v_n(t) \in \partial\varphi(x_n(t)), \quad 0 \leq t \leq T, \quad (3.2.50a)$$

$$\lambda_n(t) \dot{x}_n(t) + \dot{v}_n(t) + v_n(t) + \nabla\psi(x_n(t)) = 0, \quad a.e. \quad 0 \leq t \leq T, \quad (3.2.50b)$$

$$x_n(0) = x_0, \quad v_n(0) = v_0. \quad (3.2.50c)$$

We will show that (x_n, v_n) converges uniformly to a solution of (3.2.48a-3.2.48c).

Consider, in $C([0, T], H \times H)$, the norm

$$\|(z, w)\|_{c_0} = \max_{t \in [0, T]} \sqrt{c_0^2 \|z(t)\|^2 + \|w(t)\|^2}.$$

It is equivalent to the *sup* norm in $C([0, T], H \times H)$. Using Theorem 3.2.5, (3.2.43) and (3.2.49) we deduce the existence of a constant C (which is independant of n) such that for any n, m we have

$$\begin{aligned} \|(x_n, v_n) - (x_m, v_m)\|_{c_0} &\leq C \exp\left(\frac{\|\dot{\lambda}_n + \dot{\lambda}_m\|_{L^1([0, T])}}{2c_0} + T\left(1 + \frac{L_\psi}{c_0}\right)\right) \|\lambda_n - \lambda_m\|_{L^1([0, T])} \\ &\leq C \exp\left(\frac{\text{TV}(\lambda)}{c_0} + T\left(1 + \frac{L_\psi}{c_0}\right)\right) \|\lambda_n - \lambda_m\|_{L^1([0, T])}. \end{aligned}$$

Since the sequence $(\lambda_n)_{n \in \mathbb{N}}$ converges to λ in $L^1(0, T)$, we deduce that (x_n, v_n) is a Cauchy sequence with respect to the *sup* norm. Therefore, (x_n, v_n) converges uniformly to some continuous $(x, v) : [0, T] \rightarrow H \times H$. Since $\nabla\psi$ is continuous we also obtain

$$\nabla\psi(x_n) \rightarrow \nabla\psi(x) \quad \text{uniformly on } [0, T].$$

Moreover, Lemma 3.2.10 ensures that $\lambda_n \dot{x}_n$ and \dot{v}_n converge weakly to $\lambda \dot{x}$ and \dot{v} in $L^2(0, T; H)$. Hence letting $n \rightarrow +\infty$ in (3.2.50b) gives

$$\lambda \dot{x} + \dot{v} + v + \nabla\psi(x) = 0,$$

that's (3.2.48b). Finally, since the graph of $\partial\varphi$ is closed we obtain, for all $t \in [0, T]$, $v(t) \in \partial\varphi(x(t))$ that's (3.2.48a). Hence (x, v) is a solution of (3.2.48a)-(3.2.48c).

Uniqueness : We adapt the proof of [6], using differential and integral calculus for BV functions which involves differential measures.

Define $\lambda^- : [0, T] \mapsto [c_0, +\infty[$ by

$$\begin{aligned} \lambda^-(0) &= \lambda(0), \\ 0 < t \leq T : \lambda^-(t) &= \lim_{\varepsilon > 0, \varepsilon \rightarrow 0} \lambda(t - \varepsilon). \end{aligned}$$

Let $(x, v), (y, w) : [0, T] \rightarrow H \times H$ be two strong solutions of (3.2.48a)-(3.2.48b)-(3.2.48c). Explicitly

$$\begin{aligned}\lambda \dot{x} + \dot{v} + v + \nabla \psi(x) &= 0 \text{ a.e.}; v(t) \in \partial \varphi(x(t)) \quad \forall t; x(0) = x_0, v(0) = v_0, \\ \lambda \dot{y} + \dot{w} + w + \nabla \psi(y) &= 0 \text{ a.e.}; w(t) \in \partial \varphi(y(t)) \quad \forall t; y(0) = x_0, w(0) = v_0.\end{aligned}$$

Since $\lambda = \lambda^-$ a.e., we also have

$$\lambda^- \dot{x} + \dot{v} + v + \nabla \psi(x) = 0 \text{ and } \lambda^- \dot{y} + \dot{w} + w + \nabla \psi(y) = 0 \text{ a.e.}$$

and consequently

$$\lambda^- (\dot{x} - \dot{y}) + (\dot{v} - \dot{w}) + v - w + \nabla \psi(x) - \nabla \psi(y) = 0 \text{ a.e.} \quad (3.2.51)$$

In terms of differential measures on $[0, T]$ we have [22, Proposition 11.1]

$$d[\lambda^-(x - y) + (v - w)] = \lambda^- d(x - y) + (x - y) d\lambda^- + d(v - w). \quad (3.2.52)$$

Integrating the left hand term on $[0, s[$ and taking the initial condition into account, we obtain for $s \in [0, T]$ [22, Corollary 8.2]

$$\int_{[0, s[} d[\lambda^-(x - y) + (v - w)] = \lambda^-(s)(x(s) - y(s)) + (v(s) - w(s)). \quad (3.2.53)$$

Now integrating the right hand term of (3.2.52) on $[0, s)$ and taking (3.2.51) into account, we get

$$\begin{aligned}\int_{[0, s[} [\lambda^- d(x - y) + (x - y) d\lambda^- + d(v - w)] \\ &= \int_{[0, s[} [\lambda^- d(x - y) + d(v - w)] + \int_{[0, s[} (x - y) d\lambda^- \\ &= \int_{[0, s[} [\lambda^- (\dot{x} - \dot{y}) + \dot{v} - \dot{w}] dt + \int_{[0, s[} (x - y) d\lambda^- \\ &= - \int_{[0, s[} [v - w + \nabla \psi(x) - \nabla \psi(y)] dt + \int_{[0, s[} (x - y) d\lambda^-. \quad (3.2.54)\end{aligned}$$

From (3.2.52, 3.2.53, 3.2.54) we deduce

$$\lambda^-(s)(x(s) - y(s)) + (v(s) - w(s)) = \int_{[0, s[} (x - y) d\lambda^- - \int_{[0, s[} [v - w + \nabla \psi(x) - \nabla \psi(y)] dt.$$

Whence

$$\|\lambda^-(s)(x(s) - y(s)) + (v(s) - w(s))\| \leq \int_{[0, s[} \|x - y\| |d\lambda^-| + \int_{[0, s[} \|v - w\| dt + \int_{[0, s[} \|\nabla \psi(x) - \nabla \psi(y)\| dt \quad (3.2.55)$$

Define $\theta(s) = (c_0^2 \|x(s) - y(s)\|^2 + \|v(s) - w(s)\|^2)^{1/2}$. The same reasoning as in Theorem 3.2.5 yields

$$\theta(s) \leq \|\lambda^-(s)(x(s) - y(s)) + (v(s) - w(s))\|, \quad \forall s \in [0, T].$$

Besides we also have $c_0 \|x(s) - y(s)\| \leq \theta(s)$, $\|v(s) - w(s)\| \leq \theta(s)$, and we have that $\nabla\psi$ is the gradient of a convex, continuously differentiable function $\psi : H \rightarrow \mathbb{R}$, and by Lipschitz continuity property of $\nabla\psi$, and definition of θ , we have

$$\begin{aligned} \|\nabla\psi(x(t)) - \nabla\psi(y(t))\| &\leq L_\psi \|x(t) - y(t)\| \\ &\leq \frac{L_\psi}{c_0} \theta(t). \end{aligned} \quad (3.2.56)$$

Hence, with (3.2.55)

$$\theta(s) \leq \frac{1}{c_0} \int_{[0,s[} \theta |d\lambda^-| + \int_{[0,s[} \theta dt + \frac{1}{c_0} L_\psi \int_{[0,s[} \theta dt = \int_{[0,s[} \theta d\mu, \quad (3.2.57)$$

where $d\mu$ denotes the nonnegative measure $\frac{1}{c_0} |d\lambda^-| + \left(1 + \frac{1}{c_0} L_\psi\right) dt$.

If $\theta \not\equiv 0$ on $[0, T]$, define $t_0 = \inf \{t \in [0, T], \theta(t) > 0\}$. Note $t_0 < T$ and $\theta(t_0) = 0$, since θ is continuous. With (3.2.57) we then have

$$\theta(s) \leq \int_{]t_0, s[} \theta d\mu, \quad t_0 < s \leq T. \quad (3.2.58)$$

In view of $\int_{]t_0, t_0]} d\mu = 0$ and of the right continuity at t_0 of $t \rightarrow \int_{]t_0, t]} d\mu$, [22, Proposition 9.1] there exists some $t_1 \in]t_0, T]$ such that $\int_{]t_0, t_1]} d\mu < 1/2$. Let M be an upper bound of θ on $[0, t_1]$; from (3.2.58) we deduce, for $s \in]t_0, t_1]$

$$\theta(s) \leq M \int_{]t_0, s[} d\mu \leq M \int_{]t_0, t_1[} d\mu \leq \frac{M}{2}.$$

Hence $M/2$ is also an upper bound of θ on $[0, t_1]$, which necessarily entails $M = 0$ and $\theta \equiv 0$ on $[0, t_1]$. But this is contradiction with the definition of t_0 . Hence $\theta \equiv 0$ and $(x, v) \equiv (y, w)$ on $[0, T]$. \square

Theorem 3.2.11 has a natural global version formulation :

Corollary 3.2.12 *Suppose that $\lambda : [0, \infty[\rightarrow]0, \infty[$ is of bounded variation on $[0, T]$ and $\inf \lambda([0, T]) > 0$ for any $T < \infty$. Let $v_0 \in \partial\varphi(x_0)$ and $v_0 \neq 0$. Then there is existence and uniqueness of a strong solution $(x, v) : [0, \infty[\rightarrow H \times H$ of the Cauchy problem*

$$\lambda \dot{x} + \dot{v} + v + \nabla\psi(x) = 0, \quad v(t) \in \partial\varphi(x(t)) \quad x(0) = x_0, \quad v(0) = v_0 \quad (3.2.59)$$

where the first equality holds for almost all $t \in [0, \infty[$, and the inclusion holds for all $t \in [0, \infty[$.

Remark 3.2.13 The results obtained in this paper still hold if $B = \nabla\psi$ the gradient of a convex, continuously differentiable function $\psi : H \rightarrow \mathbb{R}$ is replaced by a maximal monotone cocoercive operator $B : H \rightarrow H$ (see [1]).

3.2.4 Perspective

Let us list some interesting questions to be examined in the future :

1. Study the asymptotic stability properties, corresponding to the case where $T = +\infty$, in connection with the convergence results of [2] and [7].
2. Extend the results to the case where the Levenberg-Marquart regularization term is given in a closed-loop form, $\lambda(t) = \alpha(\|\dot{x}(t)\|)$ as in [6].
3. Study the same questions for the corresponding forward-backward algorithms, see [1].

Bibliographie

- [1] B. Abbas, H. Attouch, Dynamical systems and forward-backward algorithms associated with the sum of a convex subdifferential and a monotone cocoercive operator, *Optimization*, (2014) <http://dx.doi.org/10.1080/02331934.2014.971412>.
- [2] B. Abbas, H. Attouch, B. F. Svaiter, Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces, *J. Optim. Theory Appl.*, **161** (2014), No. 2, pp. 331-360.
- [3] Alvarez, F., Attouch, H., Bolte, J., Redont, P. : A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics. *J. Math. Pures Appl.* **81**, 747–779 (2002).
- [4] Antipin, A.S. : Minimization of convex functions on convex sets by means of differential equations. *Differential Equations.* **30**, 1365–1375 (1994).
- [5] Attouch, H., Briceno, L., Combettes, P.L. : A parallel splitting method for coupled monotone inclusions. *SIAM J. Control Optim.* **48**, 3246–3270 (2010).
- [6] H. Attouch, P. Redont, B. F. Svaiter, Global convergence of a closed-loop regularized Newton method for solving monotone inclusions in Hilbert spaces, *J. Optim. Theory Appl.*, 157 (2013), No. 3, pp. 624–650.
- [7] Attouch, H., Svaiter, B.F. : A continuous dynamical Newton-like approach to solving monotone inclusions. *SIAM J. Control Optim.* **49**, 574–598 (2011).
- [8] Bauschke, H., Combettes, P.L. : *Convex Analysis and Monotone Operator Theory*. CMS books in Mathematics, Springer (2011).
- [9] Beck, A., Teboulle, M. : Gradient-based algorithms with applications in signal recovery problems. In : Palomar, D., Eldar, Y. (eds.) : *Convex Optimization in Signal Processing and Communications*, pp. 33–88. Cambridge University Press, (2010).
- [10] Bolte, J. : Continuous gradient projection method in Hilbert spaces. *J. Optim. Theory Appl.* **119**, 235–259 (2003).
- [11] Brézis, H. : *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert*. North-Holland/Elsevier, New-York (1973).
- [12] Brézis, H. : *Analyse Fonctionnelle*. Masson, Paris (1983).
- [13] Bruck, R. E. : Asymptotic convergence of nonlinear contraction semigroups in Hilbert spaces. *J. Funct. Anal.* **18**, 15-26 (1975).
- [14] Combettes, P.L., Pesquet, J.-C. : Proximal splitting methods in signal processing. In : Bauschke, H., Burachik, R., Combettes, P.L., Elser, V., Luke, D.R., Wolkowicz, H. (eds.) : *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pp. 185-212. Springer, New York, (2011).

- [15] Dennis, J. E., Schnabel, R.B. : Numerical Methods for Unconstrained Minimization. Prentice-Hall, Englewood Cliffs, NJ, 1983. Reprinted by SIAM publications (1993).
- [16] Dunn, J.C., Bertsekas, D.P. : Efficient dynamic programming implementations of Newton's method for unconstrained optimal control problems. *J. Optim. Theory Appl.* **63**, pp. 23–38 (1999).
- [17] Ferreira, O.P., Svaiter, B.F. : Kantorovich's theorem on Newton's method on Riemannian manifolds. *Journal of Complexity* **18**, 304–329 (2002).
- [18] Güler, O. : On the convergence of the proximal point algorithm for convex minimization, *SIAM J. Control Optim.* **29**, 403–419 (1991).
- [19] Haraux, A. : Systèmes dynamiques dissipatifs et applications. RMA 17, Masson, Paris, (1991).
- [20] Iusem, A.N., Svaiter, B.F., Da Cruz Neto, J.X. : Central paths, generalized proximal point methods, and Cauchy trajectories in riemannian manifolds. *SIAM J. Control Optim.* **37**, 566–588 (1989).
- [21] Minty, G.J. : Monotone (nonlinear) operators in Hilbert spaces. *Duke Mathematical Journal* **29**, 341–346 (1962).
- [22] Moreau, J.J. : Bounded variation in time. In : Moreau, J.J., Panagiotopoulos, P.D., Strang, G. (eds.) *Topics in Nonsmooth Mechanics*, pp. 1–76. Birkhauser, Basel (1988)
- [23] Nocedal, J., Wright, S. : Numerical Optimization. Springer Series in Operations Research, Springer-Verlag, New-York (1999).
- [24] Opial, Z. : Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Amer. Math. Soc.* **73**, 591–597 (1967).
- [25] Peypouquet, J., Sorin, S. : Evolution equations for maximal monotone operators : asymptotic analysis in continuous and discrete time, *J. of Convex Analysis* **17**, 1113–1163 (2010).
- [26] Rockafellar, R.T. : Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* **14**, 877–898 (1976).
- [27] Rockafellar, R.T. : Maximal monotone relations and the second derivatives of nonsmooth functions. *Ann. Inst. Henri Poincaré* **2**, 167–184 (1985).
- [28] Sontag, E.D. : *Mathematical Control Theory*, second edition. Springer-Verlag, New-York (1998).
- [29] Solodov, M.V., Svaiter, B.F. : A globally convergent inexact Newton method for systems of monotone equations. In : Fukushima, M., Qi, L., (eds) : *Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, pp. 355–369. Kluwer Academic Publishers, (1999).
- [30] Zeidler, E. : *Nonlinear Functional Analysis and its Applications, Part II : Monotone Operators*. Springer-Verlag, New-York (1990).
- [31] Zhu, D.L., Marcotte, P., Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities. *J. Optim.* **6**, 714–726 (1996).

Chapitre 4

Systèmes dynamiques et algorithmes forward-backward associés à la somme d'un sous-différentiel convexe et d'un opérateur monotone cocoercif

4.1 Description et résultats principaux :

Dans ce chapitre, nous étudions différentes classes de systèmes dynamiques visant à résoudre les inclusions monotones gouvernées par un opérateur maximal monotone structuré $M = \partial\Phi + B$. On se place dans un espace de Hilbert H , $\partial\Phi$ désigne le sous différentiel d'une fonction convexe, semi-continue inférieurement, et propre $\Phi : H \rightarrow \mathbb{R} \cup \{+\infty\}$, et $B : H \rightarrow H$ est un opérateur monotone cocoercif, i.e. il existe $\beta > 0$ tel que

$$\forall x, y \in \mathcal{H} \quad \langle Bx - By, x - y \rangle \geq \beta \|Bx - By\|^2.$$

Le problème s'écrit

Problème 4.1.1 *Soit H un espace de Hilbert réel*

$$\text{Trouver } x \in H \text{ tel que } \partial\Phi(x) + B(x) \ni 0.$$

Le formalisme utilisé dans le problème (4.1.1) permet de couvrir une grande classe de problèmes, allant des sciences de la décision à l'ingénierie. Sur un plan mathématique, il est directement lié à deux domaines importants, à savoir l'optimisation convexe (lorsque $B = 0$), et les problèmes de point fixe pour les contractions (lorsque $\Phi = 0$, et $B = I - T$, avec T contraction).

Nous étudions diverses dynamiques visant à la résolution du problème d'équilibre correspondant (4.1.1). Nous commençons par l'étude des dynamiques continues et discrètes de type Newton régularisé, qui ont été introduites dans Attouch et Svaiter SIAM 2011, et développées dans Abbas, Attouch, et Svaiter JOTA, 2014.

Le problème de Cauchy pour ce système de Newton régularisé s'écrit sous la forme du système en (x, v) suivant

$$v(t) \in \partial\Phi(x(t)), \tag{4.1.1}$$

$$\lambda \dot{x}(t) + \dot{v}(t) + v(t) + B(x(t)) = 0, \tag{4.1.2}$$

$$x(0) = x_0, v(0) = v_0, v_0 \in \partial\Phi(x_0). \tag{4.1.3}$$

Dans (4.1.2) λ , le terme de régularisation de type Levenberg - Marquardt, est une constante positive.

En s'appuyant sur la représentation de Minty de l'opérateur maximal monotone $\partial\Phi$, nous reformulons le système (4.1.1)-(4.1.3) sous la forme équivalente suivante :

$$\begin{cases} x(t) = \text{prox}_{\mu(t)\Phi}(y(t)), \\ \dot{y}(t) + y(t) - \text{prox}_{\mu(t)\Phi}(y(t)) + \mu B(\text{prox}_{\mu\Phi}(y(t))) = 0, \end{cases} \quad (4.1.4)$$

avec $\mu = 1/\lambda$, $y : [0, +\infty[\rightarrow H$, $y(\cdot) = x(\cdot) + \mu v(\cdot)$, et $\text{prox}_{\mu\Phi} : H \rightarrow H$ est l'opérateur proximal associé à $\mu\Phi$, qui est la résolvante de l'opérateur $\partial\Phi$

$$J_{\mu}^{\partial\Phi} x = \text{prox}_{\mu\Phi} x = \text{argmin} \left\{ \Phi(z) + \frac{1}{2\mu} \|x - z\|^2 : z \in H \right\}$$

Dans le théorème suivant, nous montrons un résultat d'existence et d'unicité :

Théorème 4.1.2 *Soit λ une constante positive. Supposons que $\partial\Phi$ est le sous différentiel d'une fonction convexe, semi-continue inférieurement, et propre $\Phi : H \rightarrow \mathbb{R} \cup \{+\infty\}$, et B est un opérateur monotone cocoercif sur H . Soit $(x_0, v_0) \in H \times H$ et $v_0 \in \partial\Phi(x_0)$. Alors, il existe une solution unique globale forte $(x(\cdot), v(\cdot)) : [0, +\infty[\rightarrow H \times H$ du problème de Cauchy (4.1.1)-(4.1.3).*

Puis, nous étudions les propriétés de convergence des trajectoires lorsque t tend vers $+\infty$. A cet effet, nous nous appuyons sur une analyse asymptotique de type Lyapunov, et sur le lemme d'Opial. Nous utilisons les fonctions de Lyapunov suivantes : pour tout $z \in S = \{z \in H; \partial\Phi(z) + B(z) \ni 0\}$, pour tout $t \geq 0$

$$\begin{aligned} g_z(t) &:= \Phi(z) - [\Phi(x(t)) + \langle z - x(t), v(t) \rangle] \\ \Gamma_z(t) &:= \frac{1}{2} \|x(t) - z\|^2 + \mu g_z(t), \end{aligned}$$

Le théorème suivant est notre résultat principal concernant la convergence des trajectoires.

Théorème 4.1.3 *Supposons que $S = \{z \in H; \partial\Phi(z) + B(z) \ni 0\} \neq \emptyset$. Alors, pour toute trajectoire $x(\cdot)$ du système (4.1.1)-(4.1.3), les propriétés de convergence suivantes sont satisfaites, lorsque t tend vers $+\infty$:*

- (1) $\lim_{t \rightarrow +\infty} \|v(t) + B(x(t))\| = 0$;
- (2) $B(x(\cdot))$ converge fortement vers $B(z)$, où $B(z)$ est défini de manière unique pour $z \in S$.
- (3) $v(\cdot)$ converge fortement vers $-B(z)$, où $B(z)$ est défini de manière unique pour $z \in S$.
- (4) $x(\cdot)$ converge faiblement vers un élément de S .

Puis, nous étudions les méthodes Backward-Forward associées à la méthode de Newton régularisée : la discrétisation explicite en temps du système (4.1.4) débouche sur le nouvel algorithme backward-forward :

$$(BF) \quad \begin{cases} x_k = \text{prox}_{\mu\Phi}(y_k) \\ y_{k+1} = (1-h)y_k + h(x_k - \mu B(x_k)) \end{cases} \quad (4.1.5)$$

Les propriétés asymptotiques du système dynamique continu (4.1.4) (et ses fonctions de Lyapunov) nous donnent un guide pour l'étude des propriétés asymptotiques de l'algorithme (4.1.5).

Hypothèses **H**

H_Φ : $\Phi : H \longrightarrow \mathbb{R} \cup \{+\infty\}$ est une fonction convexe, semi-continue inférieurement, et propre.

H_B : $B : H \longrightarrow H$ est un opérateur monotone cocoercif.

H_{μ,β} : les paramètres h, μ vérifient $0 < h < \delta := \frac{1}{2} + \inf \left\{ 1, \frac{\beta}{\mu} \right\}$, et $0 < \mu < 2\beta$.

H_S : l'ensemble des solutions $S = \{z \in H; \partial\Phi(z) + B(z) \ni 0\}$ n'est pas vide.

Sous l'Hypothèse **H**, nous démontrons le résultat de convergence suivant :

Théorème 4.1.4 *Supposons les hypothèses **H** vérifiées. Soit (x_k, y_k) une suite générée par (BF). Alors les propriétés suivantes sont satisfaites :*

(a) (y_k) converge faiblement vers un élément \bar{y} , avec $\text{prox}_{\mu\Phi}(\bar{y}) \in S$;

(b) (x_k) converge faiblement vers $\bar{x} = \text{prox}_{\mu\Phi}(\bar{y})$, $\bar{x} \in S$;

(c) $B(x_k)$ converge fortement vers $B(\bar{x})$;

(d) la vitesse est de carré sommable, i.e. $\sum_k \|x_k - x_{k-1}\|^2 < \infty$, $\sum_k \|y_k - y_{k-1}\|^2 < \infty$;

(e) $y_k - x_k$ converge fortement dans \mathcal{H} .

Les algorithmes forward-backward sont des outils d'optimisation très utiles relevant des méthodes d'éclatement (splitting). Ils permettent de réduire la taille des problèmes étudiés et ont été appliquées avec succès à de nombreux problèmes allant du traitement du signal, image, information, aux méthodes de décomposition de domaine dans les EDP.

Les méthodes forward-backward classiques s'appuient sur la discrétisation naturelle explicite-implicite du semi-groupe de contractions généré par $(-\partial\Phi + B)$ (cf. paragraphe suivant). Une nouveauté de ce travail est que nous montrons que la discrétisation de la méthode de Newton régularisé, débouche sur de nouveaux algorithmes, dits backward-forward, comprenant les mêmes blocs de bases, mais organisés différemment.

Notons qu'on obtient la convergence forte des trajectoires si on suppose que Φ est inf-compact, ou fortement convexe.

Puis, dans la section suivante nous considérons un autre système dynamique attaché à la résolution de (4.1.1), qui est le semi-groupe de contractions générée par $(-\partial\Phi + B)$

$$\dot{x}(t) + \partial\Phi(x(t)) + B(x(t)) \ni 0. \quad (4.1.6)$$

Cette dynamique relève de la théorie générale des semi-groupes de contractions dans les Hilbert, dont les générateurs sont précisément les opérateurs maximaux monotones. La théorie générale correspondante, nous donne l'existence et l'unicité d'une solution globale forte pour le problème de Cauchy. Pour chaque $x_0 \in \text{dom}\partial\Phi$, il existe une unique solution globale forte de (4.1.6) qui vérifie $x(0) = x_0$.

Supposons que $S \neq \emptyset$, où S désigne toujours l'ensemble des zéros de l'opérateur $(\partial\Phi + B)$. Baillon-Brézis ont étudié le comportement asymptotique des trajectoires, et ont montré que chaque trajectoire du système (4.1.6) converge faiblement de façon ergodique vers un point d'équilibre qui est un élément de S .

Bruck a étudié les propriétés de convergence faible (au lieu de convergence ergodique faible). La théorie de la convergence de Bruck est applicable séparément à $\partial\Phi$ et B qui sont des opérateurs demi-positifs. Mais on ne sait pas si la somme de ces deux opérateurs reste un opérateur demi-positif, et donc la théorie de Bruck n'est pas directement applicable à notre cadre. Une nouveauté de notre travail est que nous montrons que toute trajectoire du système (4.1.6) converge faiblement vers un élément de S .

Dans le théorème suivant, nous montrons que (4.1.6) a des propriétés de convergence similaires à celles présentées dans la méthode de Newton régularisée.

Théorème 4.1.5 *Supposons que $S = \{z \in H; \partial\Phi(z) + B(z) \ni 0\} \neq \emptyset$. Alors, pour toute trajectoire $x(\cdot)$ du (4.1.6), les propriétés de convergence suivantes sont satisfaites :*

- (1) $\int_0^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty$, $x(\cdot)$ est à énergie finie.
- (2) $x(\cdot)$ converge faiblement vers un élément de S .
- (3) $B(x(\cdot))$ converge fortement vers $B(z)$, où $B(z)$ est défini de manière unique pour $z \in S$.

La discrétisation en temps du système (4.1.6) débouche sur l'algorithme forward-backward classique, où h est un pas du temps constant,

$$x_{k+1} = (I + h\partial\Phi)^{-1}(x_k - hB(x_k)) \quad (4.1.7)$$

et qui génère des suites (x_k) qui convergent faiblement vers un élément de S , avec $0 < h < 2\beta$.

Finalement, nous considérons le système dynamique proximal-gradient associé à la reformulation de (4.1.1) comme un problème de point fixe

$$\dot{x}(t) + x(t) - \text{prox}_{\mu\Phi}(x(t) - \mu B(x(t))) = 0. \quad (4.1.8)$$

Insistons sur le fait que $\text{prox}_{\mu\Phi}$ est lipschitzien, ce système relève donc du théorème de Cauchy-Lipschitz.

Les propriétés de la convergence ont été étudiées d'abord par Antipin et Bolte dans le cas particulier où Φ est la fonction indicatrice d'un ensemble convexe fermé C , et B désigne le gradient d'une fonction convexe différentiable. Dans ce cas, le système (4.1.8) se ramène à la méthode de gradient-projeté suivante :

$$\dot{x}(t) + x(t) - \text{proj}_C(x(t) - \mu\nabla\Psi(x(t))) = 0.$$

Nos résultats de convergence pour la dynamique gradient-proximal sont contenus dans le théorème suivant :

Théorème 4.1.6 *Soit $\Phi : H \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe, semi-continue inférieurement, et propre, et $B : H \rightarrow H$ est un opérateur monotone cocoercif. Supposons que $S = \{z \in H; \partial\Phi(z) + B(z) \ni 0\} \neq \emptyset$. Pour chaque $x_0 \in H$, soit $x : [0, +\infty[\rightarrow H$ la solution unique globale forte du problème de Cauchy pour le système proximal gradient :*

$$\begin{aligned} \dot{x}(t) + x(t) - \text{prox}_{\mu\Phi}(x(t) - \mu B(x(t))) &= 0 \\ x(0) &= x_0. \end{aligned}$$

Alors les propriétés suivantes sont satisfaites :

- (1) *Supposons que $0 < \mu < 4\beta$, alors*
 - (i) $x(t)$ converge faiblement vers $x_\infty \in S$, lorsque $t \rightarrow +\infty$.
 - (ii) $B(x(\cdot))$ converge fortement vers $B(z)$, où $B(z)$ est définie de manière unique pour $z \in S$.
 - (iii) $\lim_{t \rightarrow +\infty} \dot{x}(t) = 0$, $\int_0^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty$.
- (2) *Supposons que $B = \nabla\Psi$, Ψ est différentiable convexe. Alors pour tout $\mu > 0$, (i), (ii), et (iii) sont satisfaites.*

La discrétisation en temps du système (4.1.8) fournit l'algorithme forward-backward relaxé suivant :

$$x_{k+1} = (1 - h)x_k + h \operatorname{prox}_{\mu\Phi}(x_k - \mu B(x_k)). \quad (4.1.9)$$

La convergence faible des suites (x_k) générées par (4.1.9), vers un élément de S , est obtenue sous l'hypothèse

$$0 < \mu < 2\beta, 0 < h < \delta := \frac{1}{2} + \inf \left\{ 1, \frac{\beta}{\mu} \right\}.$$

La section 4.3 est consacrée à une illustration numérique visant à comparer, dans un cas modèle, les comportements asymptotiques des algorithmes forward-backward associés aux dynamiques considérées précédemment, à savoir la dynamique Newton régularisé, le semi-groupe de contractions généré par $-(\partial\varphi + B)$, et le système dynamique proximal-gradient. On travaille dans le cadre simple suivant :

- $H = \mathbb{R}^n$.
- Φ est la fonction indicatrice d'un ensemble convexe fermé C . Dans ce cas, l'opérateur proximal associé à Φ n'est rien d'autre que la projection sur l'ensemble C . Dans notre exemple, on a pris C égal à un hyperplan, i.e., $C = \{x \in \mathbb{R}^n; \langle x, a \rangle = b\}$, où $a \in \mathbb{R}^n \setminus \{0\}$ et $b \in \mathbb{R}$.
- B désigne le gradient d'une fonction convexe différentiable Ψ , où $\Psi(x) = \frac{1}{2} \|Lx - d\|^2$, avec $L \in \mathbb{R}^{m \times n} \setminus \{0\}$, $d \in \mathbb{R}^m$.

Nous mettons en œuvre nos algorithmes à l'aide du logiciel Scilab, avec L généré de manière aléatoire dans $\mathbb{R}^m \times \mathbb{R}^n$ $n = m = 20$, $d = I_m$, et à partir de points de départ générés aléatoirement dans $[-2, 2]^{20}$. On considère l'hyperplane $C = \{x \in \mathbb{R}^n; \langle x, a \rangle = b\}$, avec $a = \frac{1}{\sqrt{n}} \cdot I_n = \frac{1}{\sqrt{n}} \cdot \text{ones}(n, 1)$, $b = 3$.

Nous comparons les propriétés de convergence des algorithmes forward-backward associés aux dynamiques considérées précédemment, pour différents choix des paramètres μ , β , h , et pour 5000 itérations.

Nous remarquons que les trois algorithmes ont des performances voisines. Les algorithmes associés à la dynamique de Newton régularisé (backward-forward) et le gradient-proximal sont très proches. Lorsque $h > 1$, ces derniers sont plus efficaces que la méthode forward-backward classique associée au semi-groupe de contractions. De plus, la convergence des trajectoires vers la solution est plus rapide que dans le cas $h < 1$, où l'algorithme associé au semi-groupe de contractions généré par $-M$ est le plus efficace. En outre, lorsque les paramètres satisfont $\beta \leq \mu < 2\beta$, la convergence vers la solution est plus rapide que le cas $0 < \mu < \beta$.

4.2 Article en anglais

Dynamical systems and forward-backward algorithms associated with the sum of a convex subdifferential and a monotone cocoercive operator ¹

abstract

In a Hilbert framework, we introduce continuous and discrete dynamical systems which aim at solving inclusions governed by structured monotone operators $A = \partial\Phi + B$, where $\partial\Phi$ is the

1. B. Abbas, H. Attouch, Dynamical systems and forward-backward algorithms associated with the sum of a convex subdifferential and a monotone cocoercive operator, Optimization, (2014) <http://dx.doi.org/10.1080/02331934.2014.971412>.

subdifferential of a convex lower semicontinuous function Φ , and B is a monotone cocoercive operator. We first consider the extension to this setting of the regularized Newton dynamic with two potentials which was considered in Abbas, Attouch, Svaiter JOTA, 2014. Then, we revisit some related dynamical systems, namely the semigroup of contractions generated by A , and the continuous gradient projection dynamic. By a Lyapunov analysis, we show the convergence properties of the orbits of these systems, thereby extending the known results. The time discretization of these dynamics gives various forward-backward splitting methods (some new) for solving structured monotone inclusions involving non-potential terms. The convergence of these algorithms is obtained under classical step size limitation. Perspectives are given in the field of numerical splitting methods for optimization, and multi-criteria decision processes.

4.2.1 Introduction

Throughout this paper, \mathcal{H} is a real Hilbert space with scalar product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. We are going to study some continuous and discrete dynamics which aim at solving structured monotone inclusions of the following type

$$\partial\Phi(x) + Bx \ni 0 \tag{4.2.1}$$

where $\partial\Phi$ is the subdifferential of a convex lower semicontinuous function $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$, and B is a monotone cocoercive operator. Recall that a monotone operator $B : \mathcal{H} \rightarrow \mathcal{H}$ is cocoercive if there exists a constant $\beta > 0$ such that for all $x, y \in \mathcal{H}$

$$\langle Bx - By, x - y \rangle \geq \beta \|Bx - By\|^2.$$

The abstract formulation (4.2.1) covers a wide variety of problems in decision sciences and engineering, see for example [4], [9], [19], [34], and the discussion at the end of the paper. It is directly connected to two important areas, namely convex optimization (take $B = 0$), and the theory of fixed point for nonexpansive mappings (take $\Phi = 0$, and $B = I - T$ with T a nonexpansive mapping). It comes naturally into play when we consider both aspects within a decision process.

By a classical result, the two operators $\partial\Phi$ and B are maximal monotone, as well as their sum $A = \partial\Phi + B$. We will exploit the structure of the maximal monotone operator A , first to develop continuous dynamics, and then, by time discretization, splitting forward-backward algorithms that aim to solve (4.2.1). As a common characteristic of these dynamics, they are first-order evolution equations, whose stationary points are precisely the zeroes of the operator $A = \partial\Phi + B$. Among these dynamics and algorithms some are new, and for others it is an opportunity to revisit, and extend some convergence results with a unifying perspective.

1. Our first concern is the Newton-like dynamic approach to solving monotone inclusions which was introduced in [13]. To adapt it to structured monotone inclusions and splitting methods, this study was developed in [1], where the operator is the sum of the subdifferential of a convex lower semicontinuous function, and the gradient of a convex differentiable function. We wish to extend this study to a non potential case, and so enlarge its range of applications. Specifically, our analysis focuses on the convergence properties (as $t \rightarrow +\infty$) of the orbits of the system (4.2.2)-(4.2.3)

$$v(t) \in \partial\Phi(x(t)) \tag{4.2.2}$$

$$\lambda \dot{x}(t) + \dot{v}(t) + v(t) + B(x(t)) = 0. \tag{4.2.3}$$

In (4.2.3), λ is a positive constant which acts as a Levenberg-Marquard regularization parameter. When λ is small, and $B = 0$, the system is close to the continuous Newton method for solving $\partial\Phi(x) \ni 0$. The x components of the stationary points of the (x, v) system (4.2.2)-(4.2.3) are precisely the zeroes of the operator $A = \partial\Phi + B$. The Cauchy problem for (4.2.2)-(4.2.3) is well-posed. Indeed, by introducing the new unknown function $y(\cdot) = x(\cdot) + \mu v(\cdot)$, and setting $\mu = \frac{1}{\lambda}$, (4.2.2)-(4.2.3) can be equivalent written as

$$\begin{cases} x(t) = \text{prox}_{\mu\Phi}(y(t)), \\ \dot{y}(t) + y(t) - \text{prox}_{\mu\Phi}(y(t)) + \mu B(\text{prox}_{\mu\Phi}(y(t))) = 0, \end{cases}$$

where $\text{prox}_{\mu\Phi}$ is the proximal mapping of $\mu\Phi$. Since $\text{prox}_{\mu\Phi}$ and B are Lipschitz continuous operators, the above differential equation (with respect to y) is relevant to Cauchy-Lipschitz theorem. Under the sole assumption that the solution set S of (4.2.1) is not empty, in Theorem 4.2.8 we will show that, for any orbit of system (4.2.2)-(4.2.3), $x(\cdot)$ converges weakly to an element of S . Strong convergence is obtained under the assumption Φ inf-compact, or strongly convex.

The above system is regular with respect to the new variable y . Its explicit discretization gives (with a constant step size $h > 0$) the following backward-forward algorithm : $(x_k, y_k) \rightarrow (x_{k+1}, y_{k+1})$

$$\text{(BF)} \quad \begin{cases} x_k = \text{prox}_{\mu\Phi}(y_k), \\ y_{k+1} = (1 - h)y_k + h(x_k - \mu B(x_k)). \end{cases}$$

We will show in Theorem 4.2.20 that, under the assumption, $0 < h \leq 1$, and $0 < \mu < 2\beta$, (BF) generates sequences that converge weakly to equilibria. Indeed, this is a limitation of the step size very similar that of the classical forward-backward algorithm. Note that, when $h \neq 1$, the algorithm (BF) can be viewed as a dual form of the classical forward-backward algorithm, with reversal of the order in the composition of two basic blocks $\text{prox}_{\mu\Phi}$ and $I - \mu B$.

2. Then, we consider a naturally related dynamical system, which is the semigroup of contractions generated by $-A$, $A = \partial\Phi + B$, whose orbits are the solution trajectories of the differential inclusion

$$\dot{x}(t) + \partial\Phi(x(t)) + B(x(t)) \ni 0. \quad (4.2.4)$$

In Theorem 4.2.24, we show the weak convergence of the orbits of (4.2.4) to solutions of (4.2.1), a property which surprisingly has not been systematically studied before, and which completes Bruck's convergence theory for semigroups of contractions [24]. Explicit-implicit time discretization of (4.2.4) gives the classical forward-backward algorithm.

3. Finally, we consider the dynamic which is associated to the reformulation of (4.2.1) as a fixed point problem :

$$\dot{x}(t) + x(t) - \text{prox}_{\mu\Phi}(x(t) - \mu B(x(t))) = 0. \quad (4.2.5)$$

It is a regular dynamic which is relevant of Cauchy-Lipschitz theorem. Its convergence properties have been first investigated by Antipin [2] and Bolte [22] in the particular case where Φ is the indicator function of a closed convex set C , and B is the gradient of a convex differentiable function. In that case, the above system specializes to the continuous gradient projection method. In Theorem 4.2.28 we extend these convergence results to our more general setting. The explicit time discretization of (4.2.5) gives the relaxed forward-backward algorithm

$$x_{k+1} = (1 - h)x_k + h\text{prox}_{\mu\Phi}(x_k - \mu B(x_k)).$$

A thorough comparative study of the forward-backward algorithms provided by discretization of these various related systems is an important issue from a numerical point of view. It is a subject of ongoing study (see [11]), which is beyond the scope of this document.

The paper is organized as follows : In Section 4.2.2, we study the convergence properties of the orbits of the continuous dynamical system (4.2.2)-(4.2.3). In Section 4.2.3, we show the convergence properties of the backward-forward (BF) algorithm which is obtained by time discretization of (4.2.2)-(4.2.3). In Section 4.2.4, we examine the convergence properties of the orbits of the semigroup generated by $-(\partial\Phi + B)$, and make the link with the classical (FB) algorithm. In Section 4.2.5, we introduce the proximal-gradient dynamical system, study its convergence properties, and make the link with the relaxed (FB) algorithm. We complete this study by some perspectives in the realm of numerical optimization, and multi-criteria decision processes.

4.2.2 The continuous regularized Newton-like dynamic

Definition, global existence

By applying the Minty transformation to $\partial\Phi$, system (4.2.2)-(4.2.3) can be reformulated in a form which is relevant to the Cauchy-Lipschitz theorem, see [1], [12], [13]. First set $\mu = \frac{1}{\lambda}$ and rewrite (4.2.3) as

$$\dot{x}(t) + \mu\dot{v}(t) + \mu v(t) + \mu B(x(t)) = 0. \quad (4.2.6)$$

Let us introduce the new unknown function $y(\cdot) = x(\cdot) + \mu v(\cdot)$. Since $v(\cdot) \in \partial\Phi(x(\cdot))$, we have $x(\cdot) = \text{prox}_{\mu\Phi}y(\cdot)$, where $\text{prox}_{\mu\Phi}$ is the proximal mapping associated to $\mu\Phi$. Recall that $\text{prox}_{\mu\Phi} = (I + \mu\partial\Phi)^{-1}$, where $(I + \mu\partial\Phi)^{-1}$ is the resolvent of index $\mu > 0$ of the maximal monotone operator $\partial\Phi$. We obtain the equivalent dynamic

$$\begin{cases} \dot{x}(t) = \text{prox}_{\mu\Phi}(y(t)) \\ \dot{y}(t) + y(t) - \text{prox}_{\mu\Phi}(y(t)) + \mu B(\text{prox}_{\mu\Phi}(y(t))) = 0, \end{cases} \quad (4.2.7)$$

which makes use only of the proximal mapping associated to $\mu\Phi$, and B , which are both Lipschitz continuous operators. Indeed, for any $\mu > 0$, the operator $\text{prox}_{\mu\Phi}$ is firmly nonexpansive, see [19, Proposition 12.27]. When Φ is equal to the indicator function of a closed convex set $C \subset \mathcal{H}$, $\text{prox}_{\mu\Phi}$ is independent of μ , and is equal to proj_C , the projection operator on C (whence the proximal terminology, introduced by Moreau).

By Lemma 4.2.4 below, B is a maximal monotone Lipschitz continuous operator. Thus, by specializing Theorem 3.1. of [1] to our situation, we obtain that the Cauchy problem for (4.2.2)-(4.2.3) is well-posed. More precisely,

Theorem 4.2.1 *Let $\lambda > 0$ be a positive constant. Suppose that $\partial\Phi$ is the subdifferential of a convex lower semicontinuous proper function $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$, and that $B : \mathcal{H} \rightarrow \mathcal{H}$ is a cocoercive operator on \mathcal{H} . Let $(x_0, v_0) \in \mathcal{H} \times \mathcal{H}$ be such that $v_0 \in \partial\Phi(x_0)$.*

Then, there exists a unique strong global solution $(x(\cdot), v(\cdot)) : [0, +\infty[\rightarrow \mathcal{H} \times \mathcal{H}$ of the Cauchy

problem

$$v(t) \in \partial\Phi(x(t)); \quad (4.2.8)$$

$$\lambda \dot{x}(t) + \dot{v}(t) + v(t) + B(x(t)) = 0; \quad (4.2.9)$$

$$x(0) = x_0, v(0) = v_0. \quad (4.2.10)$$

In the above statement, we use the following notion of strong solution, as defined in [1], and [13].

Definition 4.2.2 *We say that the pair $(x(\cdot), v(\cdot))$ is a strong global solution of (4.2.8)-(4.2.9)-(4.2.10) iff the following properties are satisfied :*

(i) $x(\cdot), v(\cdot) : [0, +\infty[\rightarrow \mathcal{H}$ are absolutely continuous on each interval $[0, b]$, $0 < b < +\infty$;

(ii) $v(t) \in \partial\Phi(x(t))$ for all $t \in [0, +\infty[$;

(iii) $\lambda \dot{x}(t) + \dot{v}(t) + v(t) + B(x(t)) = 0$ for almost all $t \in [0, +\infty[$;

(iv) $x(0) = x_0, v(0) = v_0$.

Equivalent systems (4.2.2)-(4.2.3) and (4.2.7) provide a dynamic whose time discretization yields a new class of forward-backward algorithms.

Remark 4.2.3 For sake of simplicity, we have taken the regularization parameters λ and μ constant. Indeed, the conclusion of Theorem 4.2.1 still holds true, just assuming that $\lambda : [0, +\infty[\rightarrow]0, +\infty[$ is absolutely continuous on each bounded interval $[0, b]$, $0 < b < +\infty$ (indeed, it is enough assuming that λ is locally of bounded variation). Taking λ varying and asymptotically vanishing provides a dynamic which is asymptotically close the Newton dynamic associated to Φ , see [1], [12], [13]. This is an important issue for fast converging methods, a subject for further studies.

Cocoercive operators

We collect some facts that will be useful.

Lemma 4.2.4 *Let $B : \mathcal{H} \rightarrow \mathcal{H}$ be a β -cocoercive operator. Then, B is $\frac{1}{\beta}$ -Lipschitz continuous.*

Proof. Let $x, y \in \mathcal{H}$. Since B is β -cocoercive, by Cauchy-Schwarz inequality we have

$$\begin{aligned} \beta \|Bx - By\|^2 &\leq \langle Bx - By, x - y \rangle \\ &\leq \|Bx - By\| \|x - y\|. \end{aligned}$$

Hence

$$\|Bx - By\| \leq \frac{1}{\beta} \|x - y\|,$$

which expresses that B is $\frac{1}{\beta}$ - Lipschitz continuous. \square

Let us list some important classes of cocoercive operators.

- $B = I - T$ where $T : \mathcal{H} \rightarrow \mathcal{H}$ is a contraction. One can easily verify that B is $\frac{1}{2}$ -cocoercive.

- $B = M_\lambda$ where M_λ (with parameter $\lambda > 0$) is the Yosida approximation of a general maximal monotone operator $M : \mathcal{H} \rightarrow 2^{\mathcal{H}}$, (see [23]). One can easily verify that M_λ is λ -cocoercive.

By Lemma 4.2.4, if B is β -cocoercive, then it is β^{-1} -Lipschitz continuous. The next lemma, which provides a converse implication, supplies us with another important instance of cocoercive operator.

Lemma 4.2.5 [18, Corollaire 10] *Let $\Psi : \mathcal{H} \rightarrow \mathbb{R}$ be a differentiable convex function and let $\tau > 0$. Suppose that $\nabla\Psi$ is τ -Lipschitz continuous. Then $\nabla\Psi$ is τ^{-1} -cocoercive.*

Lemma 4.2.6 [26, Lemma 2.3] *Let $B : \mathcal{H} \rightarrow \mathcal{H}$ be a β -cocoercive operator, and let $\mu \in]0, 2\beta[$. Then $Id - \mu B$ is nonexpansive.*

Because of the cocoercive property of B , its inverse operator B^{-1} is strongly monotone. Hence, even if the primal problem (4.2.1) has multiple solutions, the Attouch-Théra dual problem ([14])

$$B^{-1}\xi - \partial\Phi^*(-\xi) \ni 0$$

has a unique solution (with $\xi = Bz$, and z solution of the primal problem). Returning to the primal problem (4.2.1), this gives the following result (we give below another direct proof which does not use a duality argument).

Lemma 4.2.7 *Let B be a maximal monotone operator which is cocoercive, and let $\partial\Phi$ be the subdifferential of a convex lower semicontinuous proper function $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$. Set $S = \{z \in \mathcal{H}; \partial\Phi(z) + Bz \ni 0\}$ be the solution set of (4.2.1). Then Bz is a constant vector, as z varies over S .*

Proof. Let z_1 and z_2 be two elements of S . Hence $-Bz_1 \in \partial\Phi(z_1)$ and $-Bz_2 \in \partial\Phi(z_2)$. By the monotonicity property of $\partial\Phi$

$$\langle -Bz_1 + Bz_2, z_1 - z_2 \rangle \geq 0.$$

Equivalently

$$0 \geq \langle Bz_1 - Bz_2, z_1 - z_2 \rangle.$$

Combining this inequality with the cocoercive property of B ,

$$\langle Bz_1 - Bz_2, z_1 - z_2 \rangle \geq \beta \|Bz_1 - Bz_2\|^2,$$

we obtain $0 \geq \beta \|Bz_1 - Bz_2\|^2$, that is $Bz_1 = Bz_2$. □

Convergence of the regularized Newton-like dynamic

We will study the convergence properties of the orbits of system (4.2.8)-(4.2.9), whose existence is guaranteed by Theorem 4.2.1. We call

$$S = \{z \in \mathcal{H}; \partial\Phi(z) + Bz \ni 0\}$$

the solution set of problem (4.2.1), and we assume that $S \neq \emptyset$. Let $(x(\cdot), v(\cdot)) : [0, +\infty[\rightarrow \mathcal{H} \times \mathcal{H}$ be the solution of the Cauchy problem (4.2.8)-(4.2.9)-(4.2.10). Equivalently with $\mu = \frac{1}{\lambda}$

$$v(t) \in \partial\Phi(x(t)); \tag{4.2.11}$$

$$\dot{x}(t) + \mu\dot{v}(t) + \mu v(t) + \mu B(x(t)) = 0; \tag{4.2.12}$$

$$x(0) = x_0, v(0) = v_0. \tag{4.2.13}$$

We will use the following functions : for any $z \in S$, for any $t \geq 0$

$$g_z(t) := \Phi(z) - [\Phi(x(t)) + \langle z - x(t), v(t) \rangle]$$

$$\Gamma_z(t) := \frac{1}{2} \|x(t) - z\|^2 + \mu g_z(t).$$

Note that $\Gamma_z(t)$ is a Bregman distance between $x(t)$ and z . It is associated with the convex function $x \mapsto \frac{1}{2}\|x\|^2 + \mu\Phi(x)$. In our nonsmooth setting, it combines the metric of \mathcal{H} with the metric associated to the ‘‘Hessian’’ of Φ . Our proof of the convergence is based on Lyapunov analysis, and the fact that $t \mapsto \Gamma_z(t)$ is a decreasing function. Let us state our main convergence result.

Theorem 4.2.8 *Suppose $S \neq \emptyset$. Then for all $x(\cdot)$ orbit of the system (4.2.8)-(4.2.9), the following convergence properties are satisfied, when t tends to infinity :*

1. $\lim_{t \rightarrow +\infty} \|v(t) + B(x(t))\| = 0;$
2. $B(x(\cdot))$ converges strongly to Bz , where Bz is uniquely defined for $z \in S$.
3. $v(\cdot)$ converges strongly to $-Bz$, where Bz is uniquely defined for $z \in S$.
4. $x(\cdot)$ converges weakly to an element of S .

The proof of Theorem 4.2.8 has been extended to the end of this section. We collect first some preliminary technical lemma, then we conduct a Lyapunov-type analysis, and finally prove Theorem 4.2.8 and some convergence results which are connected.

A. Preliminary results We will frequently use the following derivation chain rule for a convex lower semicontinuous function $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$, see [23, Lemma 3.3].

Lemma 4.2.9 *Suppose that the assumptions i), ii), iii) are satisfied :*

i) $v(t) \in \partial\Phi(x(t))$ for almost every $t \in [0, b];$

ii) v belongs to $L^2(0, b; \mathcal{H});$

iii) $\dot{x} \in L^2(0, b; \mathcal{H}).$

Then, $t \mapsto \Phi(x(t))$ is absolutely continuous on $[0, b]$, and, for almost every $t \in [0, b]$,

$$\frac{d}{dt}\Phi(x(t)) = \langle v(t), \dot{x}(t) \rangle. \tag{4.2.14}$$

In order to prove the weak convergence of the trajectories of system (4.2.8)-(4.2.9), we will use the Opial's lemma [33] that we recall in its continuous form ; see also [24], who initiated the use of this argument to analyze the asymptotic convergence of nonlinear contraction semigroups in Hilbert spaces.

Lemma 4.2.10 *Let S be a non empty subset of \mathcal{H} and $x : [0, +\infty[\rightarrow \mathcal{H}$ a map. Assume that*

- (i) *for every $z \in S$, $\lim_{t \rightarrow +\infty} \|x(t) - z\|$ exists;*
- (ii) *every weak sequential cluster point of the map x belongs to S .*

Then

$$w - \lim_{t \rightarrow +\infty} x(t) = x_\infty \quad \text{exists, for some element } x_\infty \in S.$$

We will also need the following lemma from [1].

Lemma 4.2.11 *Suppose that $1 \leq p < \infty$, $1 \leq r \leq \infty$, $F \in L^p([0, \infty[)$ is a locally absolutely continuous nonnegative function, $G \in L^r([0, \infty[)$ and for almost all t*

$$\frac{d}{dt}F(t) \leq G(t). \tag{4.2.15}$$

Then $\lim_{t \rightarrow \infty} F(t) = 0$.

B. Lyapunov analysis As a main ingredient of our convergence proof, we are going to show that Γ_z is a strict Lyapunov function. More precisely,

Proposition 4.2.12 *Suppose that $S \neq \emptyset$. Then, for any $z \in S$, Γ_z is a decreasing nonnegative function, and hence $\lim_{t \rightarrow +\infty} \Gamma_z(t)$ exists. Moreover*

1. $\|B(x) - Bz\| \in L^2([0, +\infty[)$;
2. x is bounded;
3. $\|\dot{x}\| \in L^2([0, +\infty[)$;
4. $\|\dot{v} + v + Bz\| \in L^2([0, +\infty[)$;
5. $\|\dot{v}\| \in L^2([0, +\infty[)$;
6. $\|v\| \in L^\infty([0, +\infty[)$.
7. $\lim_{t \rightarrow +\infty} (\Phi(x(t)) + \langle x(t), Bz \rangle)$ exists.

In order to prove Proposition 4.2.12, let us first establish some technical results.

Lemma 4.2.13 *For any $t \geq 0$ and $z \in S$*

$$g_z(t) \geq 0$$

and for almost all $t \geq 0$

$$\frac{d}{dt}g_z(t) = \langle x(t) - z, \dot{v}(t) \rangle.$$

Proof. The first inequality $g_z(t) \geq 0$, follows from the subdifferential inequality for Φ at $x(t)$, and $v(t) \in \partial\Phi(x(t))$. By the derivation chain rule in the nonsmooth convex case, see Lemma 4.2.9, and $v(t) \in \partial\Phi(x(t))$, we have $\frac{d}{dt}\Phi(x(t)) = \langle v(t), \dot{x}(t) \rangle$. Hence

$$\begin{aligned} \frac{d}{dt}g_z(t) &= -\frac{d}{dt}\Phi(x(t)) + \langle \dot{x}(t), v(t) \rangle + \langle x(t) - z, \dot{v}(t) \rangle \\ &= -\langle \dot{x}(t), v(t) \rangle + \langle \dot{x}(t), v(t) \rangle + \langle x(t) - z, \dot{v}(t) \rangle \\ &= \langle x(t) - z, \dot{v}(t) \rangle. \end{aligned}$$

□

The following result from [13] will also be useful.

Lemma 4.2.14 *For almost every $t > 0$ the following property holds :*

$$\langle \dot{x}(t), \dot{v}(t) \rangle \geq 0. \quad (4.2.16)$$

Proof. For almost every $t > 0$, $\dot{x}(t)$ and $\dot{v}(t)$ are well defined, thus

$$\langle \dot{x}(t), \dot{v}(t) \rangle = \lim_{h \rightarrow 0} \frac{1}{h^2} \langle x(t+h) - x(t), v(t+h) - v(t) \rangle.$$

By equation (4.2.9), we have $v(t) \in \partial\Phi(x(t))$. Since $\partial\Phi : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ is monotone

$$\langle x(t+h) - x(t), v(t+h) - v(t) \rangle \geq 0.$$

Dividing by h^2 , and passing to the limit preserves the inequality, which yields (4.2.16). □

We can now proceed with the proof of Proposition 4.2.12.

Proof. By definition of Γ_z , and Lemma 4.2.13

$$\begin{aligned} \frac{d}{dt}\Gamma_z(t) &= \langle \dot{x}(t), x(t) - z \rangle + \mu \langle \dot{v}(t), x(t) - z \rangle \\ &= \langle x(t) - z, \dot{x}(t) + \mu\dot{v}(t) \rangle. \end{aligned} \quad (4.2.17)$$

From (4.2.12) and (4.2.17) we deduce that

$$\frac{d}{dt}\Gamma_z(t) + \mu \langle x(t) - z, v(t) + B(x(t)) \rangle = 0. \quad (4.2.18)$$

Since $z \in S$, we have $\partial\Phi(z) + Bz \ni 0$. Equivalently, there exists some $\xi \in \partial\Phi(z)$ such that $\xi + Bz = 0$. By monotonicity of $\partial\Phi$, and $v(t) \in \partial\Phi(x(t))$ we have

$$\langle x(t) - z, v(t) - \xi \rangle \geq 0. \quad (4.2.19)$$

Let us rewrite (4.2.18) as

$$\frac{d}{dt}\Gamma_z(t) + \mu \langle x(t) - z, v(t) - \xi \rangle + \mu \langle x(t) - z, \xi + B(x(t)) \rangle = 0,$$

which, from (4.2.19), gives

$$\frac{d}{dt}\Gamma_z(t) + \mu \langle x(t) - z, \xi + B(x(t)) \rangle \leq 0.$$

From $\xi + Bz = 0$, we deduce that

$$\frac{d}{dt}\Gamma_z(t) + \mu \langle x(t) - z, B(x(t)) - Bz \rangle \leq 0.$$

By the cocoercive property of B we infer

$$\frac{d}{dt}\Gamma_z(t) + \mu\beta \|B(x(t)) - Bz\|^2 \leq 0. \quad (4.2.20)$$

From (4.2.20), we readily obtain that Γ_z is a decreasing function. Being nonnegative, it converges to a finite value. By integration of the above inequality, and using that Γ_z is nonnegative we obtain

$$\int_0^\infty \|B(x(t)) - Bz\|^2 dt < +\infty,$$

that's item 1. Since g_z is nonnegative, and Γ_z is bounded from above, we deduce from the definition of Γ_z that $\|x(t) - z\|^2$ is bounded, which implies that the orbit x is bounded, that's item 2.

To prove item 3., we return to (4.2.20), and combine it with (4.2.9), $\lambda\dot{x}(t) + \dot{v}(t) + v(t) + B(x(t)) = 0$, to obtain

$$\frac{d}{dt}\Gamma_z(t) + \mu\beta \|\lambda\dot{x}(t) + \dot{v}(t) + v(t) + Bz\|^2 \leq 0.$$

After developing, we obtain

$$\frac{d}{dt}\Gamma_z(t) + \mu\beta\lambda^2 \|\dot{x}(t)\|^2 + \mu\beta \|\dot{v}(t) + v(t) + Bz\|^2 + 2\mu\beta\lambda \langle \dot{x}(t), \dot{v}(t) + v(t) + Bz \rangle \leq 0. \quad (4.2.21)$$

By examining the last term of the left member. We have

$$\langle \dot{x}(t), \dot{v}(t) \rangle \geq 0 \text{ by Lemma 4.2.14.}$$

$$\langle \dot{x}(t), v(t) \rangle = \frac{d}{dt}\Phi(x(t)) \text{ by Lemma 4.2.9.}$$

$$\langle \dot{x}(t), Bz \rangle = \frac{d}{dt}\langle x(t), Bz \rangle..$$

Combining (4.2.21), $\mu\lambda = 1$, and the above formulas we obtain

$$\frac{d}{dt}[\Gamma_z(t) + 2\beta(\Phi(x(t)) + \langle x(t), Bz \rangle)] + \beta\lambda \|\dot{x}(t)\|^2 + \mu\beta \|\dot{v}(t) + v(t) + Bz\|^2 \leq 0. \quad (4.2.22)$$

From this, we directly obtain that

$$G_z(t) := \Gamma_z(t) + 2\beta[\Phi(x(t)) + \langle x(t), Bz \rangle]$$

is a decreasing function. Since the orbit x is bounded, it follows that G_z is bounded from below (use that Γ_z is nonnegative, and Φ admits a continuous affine minorant). Hence $\lim_{t \rightarrow +\infty} G_z(t)$ exists. Since $\lim_{t \rightarrow +\infty} \Gamma_z(t)$ also exists, we infer

$$\lim_{t \rightarrow +\infty} (\Phi(x(t)) + \langle x(t), Bz \rangle) \text{ exists.}$$

By integration of (4.2.22), and using that G_z is bounded from below, we obtain

$$\int_0^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty \quad (4.2.23)$$

$$\int_0^{+\infty} \|\dot{v}(t) + v(t) + Bz\|^2 dt < +\infty. \quad (4.2.24)$$

Let us now establish estimations on \dot{v} . Let us start from (4.2.24), and develop it. Equivalently, there exists some positive constant M such that for any $0 < T < \infty$

$$\int_0^T (\|\dot{v}(t)\|^2 + \|v(t) + Bz\|^2 + 2\langle \dot{v}(t), v(t) + Bz \rangle) dt \leq M.$$

From

$$2\langle \dot{v}(t), v(t) + Bz \rangle = 2\langle \dot{v}(t), v(t) \rangle + 2\langle \dot{v}(t), Bz \rangle = \frac{d}{dt} (\|v(t)\|^2 + 2\langle v(t), Bz \rangle)$$

we deduce that

$$\int_0^T \|\dot{v}(t)\|^2 dt + \|v(T)\|^2 + 2\langle v(T), Bz \rangle \leq \|v_0\|^2 + 2\langle v_0, Bz \rangle + M.$$

This being valid for any $0 < T < \infty$, we immediately obtain

$$\int_0^{+\infty} \|\dot{v}(t)\|^2 dt < +\infty$$

and

$$\|v\| \in L^\infty([0, +\infty[),$$

which completes the proof of Proposition 4.2.12. \square

C. Proof of convergence 1. By Proposition 4.2.12 item 3 and 5, we have $\dot{x} \in L^2([0, +\infty[)$ and $\dot{v} \in L^2([0, +\infty[)$. Hence

$$\lambda\dot{x}(\cdot) + \dot{v}(\cdot) \in L^2([0, +\infty[). \quad (4.2.25)$$

By combining (4.2.9), $\lambda\dot{x}(t) + \dot{v}(t) + v(t) + B(x(t)) = 0$, with (4.2.25) we obtain

$$v + B(x) \in L^2([0, +\infty[). \quad (4.2.26)$$

Let us apply Lemma 4.2.11 with $F(t) = \frac{1}{2}\|v(t) + B(x(t))\|^2$. By (4.2.26) we have $F \in L^1([0, +\infty[)$.

Let us show that $\frac{d}{dt}F \in L^1([0, +\infty[)$. Indeed, it follows easily from the Lipschitz property of B that $B(x)$ is absolutely continuous on any bounded set, and that for almost all $t > 0$

$$\left\| \frac{d}{dt}B(x(t)) \right\| \leq \frac{1}{\beta} \|\dot{x}(t)\|. \quad (4.2.27)$$

From

$$\frac{d}{dt}F(t) = \left\langle v(t) + B(x(t)), \dot{v}(t) + \frac{d}{dt}B(x(t)) \right\rangle \quad (4.2.28)$$

by Cauchy-Schwarz inequality and (4.2.27), we deduce that

$$\left| \frac{d}{dt}F(t) \right| \leq \|v(t) + B(x(t))\| \left(\|\dot{v}(t)\| + \frac{1}{\beta} \|\dot{x}(t)\| \right). \quad (4.2.29)$$

Since $v + B(x)$, \dot{x} and \dot{v} belong to $L^2([0, +\infty[)$, we obtain $\frac{d}{dt}F \in L^1([0, +\infty[)$. By applying Lemma 4.2.11, we obtain $\lim_{t \rightarrow +\infty} F(t) = 0$, which proves the first item of Theorem 4.2.8.

2. and 3. Let us apply Lemma 4.2.11 with $F_1(t) = \frac{1}{2} \|B(x(t)) - Bz\|^2$. By Proposition 4.2.12, item 1, we have $F_1 \in L^1([0, +\infty[)$. Moreover, by using (4.2.27), and a similar argument as above, we have

$$\left| \frac{d}{dt}F_1(t) \right| \leq \frac{1}{\beta} \|\dot{x}(t)\| \times \|B(x(t)) - Bz\|. \quad (4.2.30)$$

Combining $\|\dot{x}\| \in L^2([0, +\infty[)$ with $\|B(x) - Bz\| \in L^2([0, +\infty[)$ we deduce from (4.2.30) that $\frac{d}{dt}F_1 \in L^1([0, +\infty[)$. By Lemma 4.2.11, we obtain $\lim F_1(t) = 0$, which is our claim. Item 3. is a straightforward consequence of the two previous items.

4. Let us verify that the conditions of Opial's lemma 4.2.10 are satisfied, taking S equal to the solution set of (4.2.1).

(i) Let \bar{x} be a weak sequential cluster point of x , i.e., $\bar{x} = w - \lim x(t_n)$ for some sequence $t_n \rightarrow +\infty$. The operator $A = \partial\Phi + B$ is maximal monotone, and hence is demi-closed. From $v(t_n) + B(x(t_n)) \rightarrow 0$ strongly, $x(t_n) \rightharpoonup \bar{x}$ weakly, and $v(t_n) + B(x(t_n)) \in A(x(t_n))$, we deduce that $A(\bar{x}) = \partial\Phi(\bar{x}) + B(\bar{x}) \ni 0$, that is $\bar{x} \in S$.

(ii) Let us recall the definition of

$$\Gamma_z(t) := \frac{1}{2} \|x(t) - z\|^2 + \mu g_z(t)$$

$g_z(t) := \Phi(z) - [\Phi(x(t)) + \langle z - x(t), v(t) \rangle]$. By Proposition 4.2.12, for any $z \in S$, Γ_z is a decreasing nonnegative function, and hence converges. Hence, in order to prove that $\lim_{t \rightarrow +\infty} \|x(t) - z\|$ exists, we just need to show that $\lim_{t \rightarrow +\infty} g_z(t)$ exists. To that end, we use Lemma 4.2.11 with $F(t) = g_z$.

By Lemma 4.2.13, for almost all $t \geq 0$

$$\frac{d}{dt}g_z(t) = \langle x(t) - z, \dot{v}(t) \rangle.$$

Hence

$$\left| \frac{d}{dt}g_z(t) \right| \leq \|\dot{v}(t)\| \times \|x(t) - z\|. \quad (4.2.31)$$

By Proposition 4.2.12 item 5, $\|\dot{v}\| \in L^2([0, +\infty[)$, and by Proposition 4.2.12 item 2., x is bounded. Hence $\frac{d}{dt}g_z \in L^2([0, +\infty[)$. Let us now prove that $g_z \in L^2([0, +\infty[)$. Since g_z is nonnegative, we just need to majorize it by a square integrable function. By the convex subdifferential inequality, and $-Bz \in \partial\Phi(z)$ we have

$$\Phi(x(t)) \geq \Phi(z) - \langle Bz, x(t) - z \rangle.$$

Equivalently

$$\Phi(z) - \Phi(x(t)) \leq \langle Bz, x(t) - z \rangle.$$

Combining this inequality with the definition of g_z , we obtain

$$0 \leq g_z(t) \leq \langle Bz, x(t) - z \rangle - \langle z - x(t), v(t) \rangle.$$

Equivalently

$$0 \leq g_z(t) \leq \langle Bz + v(t), x(t) - z \rangle.$$

By Cauchy-Schwarz inequality and the triangle inequality we deduce that

$$0 \leq g_z(t) \leq (\|Bz - Bx(t)\| + \|Bx(t) + v(t)\|) \|x(t) - z\|. \quad (4.2.32)$$

By Proposition 4.2.12 item 1, we have $\|B(x) - Bz\| \in L^2(0, +\infty)$.

By (4.2.26), $v + B(x) \in L^2([0, +\infty[)$. Since x is bounded, from (4.2.32) we obtain $g_z \in L^2([0, +\infty[)$. Thus g_z and $\frac{d}{dt}g_z$ belong to $L^2([0, +\infty[)$. By Lemma 4.2.11 we conclude that $\lim_{t \rightarrow +\infty} g_z(t)$ exists. Indeed the limit is equal to zero.

As a direct consequence of the above proof we have the following result.

Proposition 4.2.15 *Suppose that $S \neq \emptyset$. Then, for any $z \in S$,*

$$\Phi(z) - \Phi(x(t)) - \langle v(t), z - x(t) \rangle \longrightarrow 0 \quad \text{as } t \rightarrow +\infty.$$

Let us complete the above result by the following related convergence properties of the orbits x of system (4.2.8)-(4.2.9).

Proposition 4.2.16 *Suppose that $S \neq \emptyset$. Then, for any $z \in S$,*

$$\int_0^{+\infty} \Phi(x(t)) - \Phi(z) + \langle Bz, x(t) - z \rangle dt < +\infty,$$

and

$$\Phi(x(t)) - \Phi(z) + \langle Bz, x(t) - z \rangle \longrightarrow 0 \quad \text{as } t \rightarrow +\infty,$$

where Bz is the element which is uniquely defined for $z \in S$. In particular (take $z = x_\infty$),

$$\Phi(x(t)) \rightarrow \Phi(x_\infty) \quad \text{as } t \rightarrow +\infty,$$

where $x_\infty \in S$ is the weak limit of the trajectory $t \mapsto x(t)$.

Proof. Let us return to (4.2.18)

$$\frac{d}{dt}\Gamma_z(t) + \mu \langle x(t) - z, v(t) \rangle + \mu \langle x(t) - z, B(x(t)) \rangle = 0.$$

By $v(t) \in \partial\Phi(x(t))$, we have the subdifferential inequality

$$\Phi(z) \geq \Phi(x(t)) + \langle z - x(t), v(t) \rangle.$$

Combining the two above relations yields

$$\frac{d}{dt}\Gamma_z(t) + \mu [\Phi(x(t)) - \Phi(z)] + \mu \langle B(x(t)) - Bz, x(t) - z \rangle + \mu \langle Bz, x(t) - z \rangle \leq 0.$$

Since B is β -cocoercive, we have $\langle B(x(t)) - Bz, x(t) - z \rangle \geq \beta \|B(x(t)) - Bz\|^2$. Hence,

$$\frac{d}{dt}\Gamma_z(t) + \mu [\Phi(x(t)) - \Phi(z)] + \mu\beta \|B(x(t)) - Bz\|^2 + \mu \langle Bz, x(t) - z \rangle \leq 0.$$

As a consequence

$$\frac{d}{dt}\Gamma_z(t) + \mu [\Phi(x(t)) - \Phi(z) + \langle Bz, x(t) - z \rangle] \leq 0. \quad (4.2.33)$$

Since $-Bz \in \partial\Phi(z)$ we have $\Phi(x(t)) - \Phi(z) + \langle Bz, x(t) - z \rangle \geq 0$. By integration of (4.2.33), and Γ_z minorized, we obtain

$$\int_0^{+\infty} \Phi(x(t)) - \Phi(z) + \langle Bz, x(t) - z \rangle dt < +\infty. \quad (4.2.34)$$

Let us apply Lemma 4.2.11 with $F_2(t) = \Phi(x(t)) - \Phi(z) + \langle Bz, x(t) - z \rangle$. By (4.2.34), and F_2 nonnegative, we have $F_2 \in L^1([0, +\infty[)$. Moreover by using the derivation chain rule in the nonsmooth convex case, see Lemma 4.2.9, and $v(t) \in \partial\Phi(x(t))$, we have $\frac{d}{dt}\Phi(x(t)) = \langle v(t), \dot{x}(t) \rangle$. Hence

$$\frac{d}{dt}F_2(t) = \langle \dot{x}(t), v(t) + Bz \rangle, \quad (4.2.35)$$

which by Cauchy-Schwarz inequality yields

$$\left| \frac{d}{dt}F_2(t) \right| \leq \|\dot{x}(t)\| (\|v(t)\| + \|Bz\|). \quad (4.2.36)$$

By Proposition 4.2.12, item 3, $\|\dot{x}\| \in L^2([0, +\infty[)$. Moreover, by Proposition 4.2.12, item 5, $\|v\| \in L^\infty([0, +\infty[)$. Hence, by (4.2.36), we obtain $\frac{d}{dt}F_2 \in L^2([0, +\infty[)$. By Lemma 4.2.11, we deduce that $\lim_{t \rightarrow +\infty} F_2(t) = 0$, which is our claim.

Note that the same conclusion can be obtained, by using the relation

$$\Phi(x(t)) - \Phi(z) + \langle Bz, x(t) - z \rangle = -[\Phi(z) - \Phi(x(t)) - \langle v(t), z - x(t) \rangle] + \langle v(t) + Bz, x(t) - z \rangle,$$

Proposition 4.2.15, and Theorem 4.2.8, item 3. \square

Corollary 4.2.17 *Let us suppose that Φ is strongly convex². Then the solution set S is reduced to a single element \bar{z} , and any orbit $x(\cdot)$ of system (4.2.8)-(4.2.9) converges strongly to \bar{z} , as $t \rightarrow +\infty$.*

Proof. Since Φ is strongly convex, its subdifferential $\partial\Phi$ is strongly monotone, and so is the sum $A = \partial\Phi + B$. Thus the solution set is reduced to a single element, let \bar{z} .

Moreover, since Φ is strongly convex, and $-B(\bar{z}) \in \partial\Phi(\bar{z})$, we have the subdifferential inequality

$$\Phi(x(t)) - \Phi(\bar{z}) + \langle Bz, x(t) - \bar{z} \rangle \geq \gamma \|x(t) - \bar{z}\|^2$$

2. Indeed, the result is true under the more general assumption, Φ uniformly convex. We thank the referee for pointing this result

for some positive constant γ . By Proposition 4.2.16

$$\Phi(x(t)) - \Phi(\bar{z}) + \langle B\bar{z}, x(t) - \bar{z} \rangle \longrightarrow 0 \quad \text{as } t \rightarrow +\infty.$$

Hence $\lim \|x(t) - \bar{z}\| = 0$, which gives the claim. \square

Corollary 4.2.18 *Suppose that $S \neq \emptyset$. Let us suppose that Φ is boundedly inf-compact, i.e., the intersections of the sublevel sets of Φ with the closed balls of \mathcal{H} are relatively compact sets. Then, any orbit $x(\cdot)$ of system (4.2.8)-(4.2.9) converges strongly as $t \rightarrow +\infty$, and its limit belongs to S .*

Proof. We know that the orbit $x(\cdot)$ of system (4.2.8)-(4.2.9) converges weakly to some $\bar{z} \in S$. As a consequence it is bounded. Moreover by Proposition 4.2.16, $\Phi(x(t)) \rightarrow \Phi(x_\infty)$, and hence $x(\cdot)$ remains in a fixed sublevel set of Φ . Since Φ is boundedly inf-compact, we obtain that the trajectory is relatively compact in \mathcal{H} , and thus converges strongly. \square

Remark 4.2.19 By a similar argument, strong convergence of $x(\cdot)$ holds under the Kadec-Klee property : whenever (x_k) converges weakly to x , and $(\Phi(x_k))$ converges to $\Phi(x)$, then (x_k) converges strongly to x .

4.2.3 Backward-forward algorithms associated with the regularized Newton dynamic

Our algorithm is constructed using the following ideas. We rely on the equivalent formulation of the dynamic involving the new variable y

$$x(t) = \text{prox}_{\mu\Phi}(y(t)) \tag{4.2.37}$$

$$\dot{y}(t) + y(t) - \text{prox}_{\mu\Phi}(y(t)) + \mu B(\text{prox}_{\mu\Phi}(y(t))) = 0. \tag{4.2.38}$$

The differential equation (4.2.38) is governed by a Lipschitz continuous operator. Explicit discretization of (4.2.38) with respect to the time variable t , with constant step size $h > 0$, gives

$$\begin{cases} x_k = \text{prox}_{\mu\Phi}(y_k), \\ \frac{y_{k+1} - y_k}{h} + y_k - \text{prox}_{\mu\Phi}(y_k) + \mu B(\text{prox}_{\mu\Phi}(y_k)) = 0. \end{cases}$$

The algorithm can be equivalently written as $(x_k, y_k) \rightarrow (x_k, y_{k+1}) \rightarrow (x_{k+1}, y_{k+1})$,

$$\text{(BF)} \quad \begin{cases} x_k = \text{prox}_{\mu\Phi}(y_k), \\ y_{k+1} = (1 - h)y_k + h(x_k - \mu B(x_k)). \end{cases} \tag{4.2.39}$$

When $h = 1$ we recover the classical forward-backward algorithm

$$x_{k+1} = \text{prox}_{\mu\Phi}(x_k - \mu B(x_k)).$$

The backward-forward algorithm (BF) is closely related to the relaxed forward-backward algorithm ([19, Theorem 25.8]). It involves the same basic blocks but in a different order. When the prox is linear, then the operations commute, and we recover the classical relaxed forward-backward algorithm. But, in general, for nonlinear problems, and in the case λ non constant (which is of interest with respect to Newton method) this is not the case. As a guide for our study of the convergence of this algorithm, we use the Lyapunov functions that have been put to the fore in the study of the continuous dynamics.

As a standing assumption we make the following set of hypotheses :

Hypothesis H :

- \mathbf{H}_Φ : The function $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ is proper, lower-semicontinuous and convex.
- \mathbf{H}_B : The operator $B : \mathcal{H} \rightarrow \mathcal{H}$ is β -cocoercive.
- $\mathbf{H}_{h,\mu}$: Parameters h and μ satisfy $0 < h < \delta := \frac{1}{2} + \inf \left\{ 1, \frac{\beta}{\mu} \right\}$, and $0 < \mu < 2\beta$.
- \mathbf{H}_S : The solution set $S = \{z \in \mathcal{H}; \partial\Phi(z) + Bz \ni 0\}$ is nonempty.

Let us state our main convergence result.

Theorem 4.2.20 *Let Hypothesis H hold. Let (x_k, y_k) be a sequence generated by (BF). Then the following properties hold :*

- a) (y_k) converges weakly to an element \bar{y} , with $\text{prox}_{\mu\Phi}\bar{y} \in S$;
- b) (x_k) converges weakly to $\bar{x} = \text{prox}_{\mu\Phi}\bar{y}$, with $\bar{x} \in S$;
- c) $B(x_k)$ converges strongly to $B\bar{x}$;
- d) the velocity is square summable, i.e., $\sum_k \|x_k - x_{k-1}\|^2 < \infty$, and $\sum_k \|y_k - y_{k-1}\|^2 < \infty$; in particular $x_k - x_{k-1}$ and $y_k - y_{k-1}$ converge strongly to zero.
- e) $y_k - x_k$ converges strongly in \mathcal{H} .

Before proving Theorem 4.2.20, we review some classical results on α -averaged operators, which will be useful.

α -averaged operators

We will use the notion of α -averaged operator, see [19, Definition 4.23]. An operator $T : \mathcal{H} \rightarrow \mathcal{H}$ is α -averaged with constant $0 < \alpha < 1$, if there exists a nonexpansive operator $R : \mathcal{H} \rightarrow \mathcal{H}$ such that $T = (1 - \alpha)I + \alpha R$. The notions of cocoerciveness and α -averaged are intimately related. We collect below some classical facts that will be useful.

- $T : \mathcal{H} \rightarrow \mathcal{H}$ is β -cocoercive iff βT is $\frac{1}{2}$ -averaged, see [19, Remark 4.24].
- For any $\mu > 0$, the operator $\text{prox}_{\mu\Phi}$ is $\frac{1}{2}$ -averaged ([19, Corollary 23.8]).
- Suppose that $B : \mathcal{H} \rightarrow \mathcal{H}$ is β -cocoercive, and $0 < \mu < 2\beta$. Then, the operator $(I - \mu B)$ is $\frac{\mu}{2\beta}$ -averaged ([19, Proposition 4.33]). This result makes precise Lemma 4.2.6.

A major interest of the notion of α -averaged operator is that the composition of two such operators is still an averaged operator. This will be particularly useful when considering Krasnosel'skii-Mann iteration for the mapping $T = (I - \mu B) \circ (\text{prox}_{\mu\Phi})$. More precisely

Lemma 4.2.21 [19, Proposition 4.32] *Let $T_i : \mathcal{H} \rightarrow \mathcal{H}$ be a α_i -averaged operator, $i = 1, 2$. Then $T := T_1 \circ T_2$ is α -averaged with constant $\alpha = \frac{1}{\delta}$, and $\delta = \frac{1}{2} + \frac{1}{2} \inf \left\{ \frac{1}{\alpha_1}, \frac{1}{\alpha_2} \right\}$.*

Convergence of the algorithm (BF)

Proof. We will first study the convergence of the sequence (y_k) , and then of the sequence (x_k) .

Convergence of the sequence (y_k) . It will be obtained as a direct consequence of the convergence of the Krasnosel'skii-Mann iteration for nonexpansive mappings. Let $T : \mathcal{H} \rightarrow \mathcal{H}$ be the operator which is defined by : for any $\xi \in \mathcal{H}$,

$$T(\xi) = (I - \mu B) \circ (\text{prox}_{\mu\Phi}(\xi)). \quad (4.2.40)$$

The algorithm (BF) can be equivalently written as

$$\begin{cases} x_k = \text{prox}_{\mu\Phi}(y_k), \\ y_{k+1} = (1 - h)y_k + hT(y_k). \end{cases} \quad (4.2.41)$$

i) Let us first suppose that $0 < h < 1$. Let us verify that we are in the situation covered by Krasnosel'skii-Mann theorem. Since B is β -cocoercive, and $0 < \mu < 2\beta$, by Lemma 4.2.6, the operator $Id - \mu B$ is nonexpansive. Since $\text{prox}_{\mu\Phi}$ is nonexpansive, we deduce that the composition mapping $T = (I - \mu B) \circ (\text{prox}_{\mu\Phi})$ is nonexpansive. By Krasnosel'skii-Mann algorithm, [19, Theorem 5.14], for $0 < h < 1$, the sequence (y_k) converges weakly to a fixed point of T , let $y_k \rightarrow \bar{y}$ with $T(\bar{y}) = \bar{y}$. Equivalently, by definition of T , we have $\text{prox}_{\mu\Phi}(\bar{y}) - \mu B(\text{prox}_{\mu\Phi}(\bar{y})) = \bar{y}$. Set $\bar{z} := \text{prox}_{\mu\Phi}(\bar{y})$. We have $\bar{z} - \mu B(\bar{z}) = \bar{y}$, which gives

$$\frac{1}{\mu}(\bar{y} - \text{prox}_{\mu\Phi}(\bar{y})) + B(\bar{z}) = 0.$$

The extremality condition characterizing $\text{prox}_{\mu\Phi}(\bar{y})$ gives

$$\frac{1}{\mu}(\bar{y} - \text{prox}_{\mu\Phi}(\bar{y})) \in \partial\Phi(\text{prox}_{\mu\Phi}(\bar{y})).$$

Comparing the two above equations, we finally obtain

$$\partial\Phi(\bar{z}) + B(\bar{z}) \ni 0.$$

ii) Take now h possibly greater or equal than 1, but $h < \delta$. Let's analyze in more detail the algorithm

$$y_{k+1} = (1 - h)y_k + hT(y_k).$$

We rely on the notion of α -averaged operator that has been discussed in the previous subsection. Let us examine the operator $T = (I - \mu B) \circ (\text{prox}_{\mu\Phi})$. The operator $\text{prox}_{\mu\Phi}$ is $\frac{1}{2}$ -averaged ([19, Corollary 23.8]). The operator $(I - \mu B)$ is $\frac{\mu}{2\beta}$ -averaged ([19, Proposition 4.33]). Hence T is α -averaged with constant $\alpha = \frac{1}{\delta}$, and $\delta = \frac{1}{2} + \inf \left\{ 1, \frac{\beta}{\mu} \right\}$, see Lemma 4.2.21. By the condition $0 < \mu < 2\beta$, we have $\delta > 1$, and hence $0 < \alpha < 1$. By [19, Proposition 5.15], we deduce that, for any $0 < h < \delta$ the sequence (y_k) converges weakly to a fixed point of T , and

$$\sum_k \|Ty_k - y_k\|^2 < \infty.$$

Equivalently, by (4.2.41)

$$\sum_k \|y_{k+1} - y_k\|^2 < \infty. \quad (4.2.42)$$

Hence

$$y_{k+1} - y_k \rightarrow 0 \text{ strongly in } \mathcal{H}. \quad (4.2.43)$$

A Lyapunov-type sequence Take \bar{z} an arbitrary element in S . Since $x_k = \text{prox}_{\mu\Phi}(y_k)$, we have

$$v_k := \frac{1}{\mu}(y_k - x_k) \in \partial\Phi(x_k). \quad (4.2.44)$$

As a Lyapunov sequence take

$$A_k := \frac{1}{2\mu}\|x_k - \bar{z}\|^2 + g_{\bar{z}}^k, \quad (4.2.45)$$

with

$$g_{\bar{z}}^k := \Phi(\bar{z}) - (\Phi(x_k) + \langle \bar{z} - x_k, v_k \rangle). \quad (4.2.46)$$

a) The following equality is a direct consequence of the Hilbert structure of \mathcal{H} .

$$\|x_{k+1} - \bar{z}\|^2 = \|x_{k+1} - x_k + x_k - \bar{z}\|^2 \quad (4.2.47)$$

$$= \|x_{k+1} - x_k\|^2 + 2\langle x_{k+1} - x_k, x_k - \bar{z} \rangle + \|x_k - \bar{z}\|^2. \quad (4.2.48)$$

b) We have

$$\begin{aligned} g_{\bar{z}}^{k+1} - g_{\bar{z}}^k &= \Phi(\bar{z}) - \Phi(x_{k+1}) - \langle \bar{z} - x_{k+1}, v_{k+1} \rangle - \Phi(\bar{z}) + \Phi(x_k) + \langle \bar{z} - x_k, v_k \rangle \\ &= \Phi(x_k) - \Phi(x_{k+1}) + \langle x_{k+1} - x_k, v_k \rangle + \langle x_{k+1} - \bar{z}, v_{k+1} - v_k \rangle. \end{aligned}$$

By convexity of Φ and $v_k \in \partial\Phi(x_k)$, we have $\Phi(x_k) - \Phi(x_{k+1}) + \langle x_{k+1} - x_k, v_k \rangle \leq 0$, which gives

$$g_{\bar{z}}^{k+1} - g_{\bar{z}}^k \leq \langle x_{k+1} - \bar{z}, v_{k+1} - v_k \rangle. \quad (4.2.49)$$

c) Let us show that $A_k := \frac{1}{2\mu}\|x_k - \bar{z}\|^2 + g_{\bar{z}}^k$ is a Lyapunov sequence. By (4.2.44), (4.2.48), and (4.2.49) we have

$$\begin{aligned} A_{k+1} - A_k &\leq \frac{1}{2\mu} (\|x_{k+1} - x_k\|^2 + 2\langle x_{k+1} - x_k, x_k - \bar{z} \rangle) + \langle x_{k+1} - \bar{z}, v_{k+1} - v_k \rangle \\ &\leq \frac{1}{2\mu} \|x_{k+1} - x_k\|^2 + \frac{1}{\mu} (\langle x_{k+1} - x_k, x_k - \bar{z} \rangle + \langle x_{k+1} - \bar{z}, (y_{k+1} - y_k) - (x_{k+1} - x_k) \rangle) \\ &\leq \frac{1}{2\mu} \|x_{k+1} - x_k\|^2 + \frac{1}{\mu} (-\|x_{k+1} - x_k\|^2 + \langle x_{k+1} - \bar{z}, y_{k+1} - y_k \rangle) \\ &\leq -\frac{1}{\mu} \|x_{k+1} - x_k\|^2 + \frac{1}{\mu} \langle x_{k+1} - \bar{z}, y_{k+1} - y_k \rangle \\ &\leq -\frac{1}{\mu} \|x_{k+1} - x_k\|^2 + \frac{1}{\mu} \langle x_k - \bar{z}, y_{k+1} - y_k \rangle + \frac{1}{\mu} \langle x_{k+1} - x_k, y_{k+1} - y_k \rangle. \end{aligned}$$

Let us write (BF) algorithm as

$$y_{k+1} - y_k = h[(x_k - y_k) - \mu Bx_k]. \quad (4.2.50)$$

Replacing $y_{k+1} - y_k$ by this expression in the above inequality gives

$$A_{k+1} - A_k \leq -\frac{1}{\mu} \|x_{k+1} - x_k\|^2 + \frac{h}{\mu} \langle x_k - \bar{z}, (x_k - y_k) - \mu Bx_k \rangle + \frac{1}{\mu} \langle x_{k+1} - x_k, y_{k+1} - y_k \rangle.$$

Equivalently

$$A_{k+1} - A_k + \frac{1}{\mu} \|x_{k+1} - x_k\|^2 + h \langle Bx_k - B\bar{z}, x_k - \bar{z} \rangle + h \langle x_k - \bar{z}, \frac{1}{\mu} (y_k - x_k) + B\bar{z} \rangle \leq \frac{1}{\mu} \langle x_{k+1} - x_k, y_{k+1} - y_k \rangle. \quad (4.2.51)$$

Since B is β -cocoercive

$$\langle Bx_k - B\bar{z}, x_k - \bar{z} \rangle \geq \beta \|Bx_k - B\bar{z}\|^2. \quad (4.2.52)$$

By (4.2.44), we have $\frac{1}{\mu}(y_k - x_k) \in \partial\Phi(x_k)$. By definition of S , and $\bar{z} \in S$, we have $-B\bar{z} \in \partial\Phi(\bar{z})$. Hence, by monotonicity of $\partial\Phi$

$$\langle x_k - \bar{z}, \frac{1}{\mu}(y_k - x_k) + B\bar{z} \rangle \geq 0. \quad (4.2.53)$$

Combining (4.2.51), (4.2.52), and (4.2.53), we obtain

$$A_{k+1} - A_k + \frac{1}{\mu} \|x_{k+1} - x_k\|^2 + h\beta \|Bx_k - B\bar{z}\|^2 \leq \frac{1}{\mu} \langle x_{k+1} - x_k, y_{k+1} - y_k \rangle. \quad (4.2.54)$$

By using $\langle x_{k+1} - x_k, y_{k+1} - y_k \rangle \leq \frac{1}{2} \|x_{k+1} - x_k\|^2 + \frac{1}{2} \|y_{k+1} - y_k\|^2$, we obtain

$$A_{k+1} - A_k + \frac{1}{2\mu} \|x_{k+1} - x_k\|^2 + h\beta \|Bx_k - B\bar{z}\|^2 \leq \frac{1}{2\mu} \|y_{k+1} - y_k\|^2. \quad (4.2.55)$$

By (4.2.42), $\sum_k \|y_{k+1} - y_k\|^2 < \infty$, and A_k is nonnegative.. By a standard argument, from (4.2.55), we obtain

a) $\lim A_k$ exists. Since g^k is nonnegative, we have $A_k \geq \frac{1}{2\mu} \|x_k - \bar{z}\|^2$. As a consequence, the sequence (x_k) is bounded.

b) $\sum_k \|Bx_k - B\bar{z}\|^2 < \infty$. Hence

$$B(x_k) \rightarrow B\bar{z} \quad \text{strongly} \quad (4.2.56)$$

where $B\bar{z}$ is uniquely defined for $\bar{z} \in S$.

c) $\sum_k \|x_{k+1} - x_k\|^2 < \infty$.

This proves item c) and d) of Theorem 4.2.20.

Convergence of the sequence (x_k) . Let us write (BF) in the following form

$$\frac{1}{h\mu} (y_{k+1} - y_k) + \frac{1}{\mu} (y_k - x_k) + B(x_k) = 0, \quad (4.2.57)$$

with

$$v_k := \frac{1}{\mu} (y_k - x_k) \in \partial\Phi(x_k). \quad (4.2.58)$$

By (4.2.43), $y_{k+1} - y_k \rightarrow 0$ strongly in \mathcal{H} . By (4.2.56), $B(x_k) \rightarrow B\bar{z}$ strongly in \mathcal{H} . From (4.2.57) we deduce that

$$y_k - x_k \rightarrow -\mu B\bar{z}$$

strongly in \mathcal{H} . Note again that $B\bar{z}$ is uniquely defined when $\bar{z} \in S$. Since we have already obtained that the sequence (y_k) converges weakly, we deduce that the sequence (x_k) converges weakly, let $x_k \rightharpoonup \bar{x}$ weakly. From (4.2.57) and (4.2.58) we have

$$-\frac{1}{h\mu}(y_{k+1} - y_k) \in (\partial\Phi + B)(x_k).$$

The operator $A = \partial\Phi + B$ is maximal monotone, and hence is demi-closed. From $y_{k+1} - y_k \rightarrow 0$ strongly (4.2.43), $x_k \rightharpoonup \bar{x}$ weakly, we deduce that $A(\bar{x}) = \partial\Phi(\bar{x}) + B(\bar{x}) \ni 0$, that is $\bar{x} \in S$.

Let us make precise the relation between the respective limits of the sequences (y_k) and (x_k) . Let $y_k \rightharpoonup \bar{y}$ weakly. Since $\bar{x} \in S$ we have $B(x_k) \rightarrow B\bar{x}$ strongly in \mathcal{H} . From (4.2.57), we deduce that $\bar{y} - \bar{x} + \mu B\bar{x} = 0$. Since $B\bar{x} + \partial\Phi(\bar{x}) \ni 0$, we obtain $\bar{x} + \mu\partial\Phi(\bar{x}) \ni \bar{y}$. Hence $\bar{x} = \text{prox}_{\mu\Phi}\bar{y}$. This complete the proof of Theorem 4.2.20. \square

Remark 4.2.22 a) Clearly, since $\delta > 1$, we can take an arbitrary $0 < h \leq 1$. Indeed, the above analysis provides an over-relaxation result.

b) The above result can be readily extended to the case h_k varying with k . The convergence of (y_k) is satisfied under the assumption : there exists some $\epsilon > 0$, such that for all $k \in \mathbb{N}$, $0 < \epsilon \leq h_k \leq \delta - \epsilon$.

c) When the dimension of \mathcal{H} is finite, by continuity of $\text{prox}_{\mu\Phi}$, and $x_k = \text{prox}_{\mu\Phi}(y_k)$ we immediately obtain the convergence of the sequence (x_k) to an element of the solution set S . In the infinite dimensional case, this argument does not work anymore, because of the lack of continuity of the prox mapping for the weak topology.

d) By analogy with the continuous case, one can reasonably conjecture that the weak convergence of the sequence (x_k) holds under the weaker condition : $0 < h < 1$ and $h\mu < 2\beta$.

Remark 4.2.23 Comparing the numerical performance of the forward-backward algorithms provided by discretization of various dynamical systems is an important issue. This is a delicate question, directly related to obtaining rapid numerical methods, a subject of ongoing study, see [10], [11].

4.2.4 Semigroup generated by $-(\partial\Phi + B)$, and FB algorithms

Continuous case

Consider a closely related dynamical system, which is the semigroup generated by $-A$; $A = \partial\Phi + B$, whose orbits are the solution trajectories of the differential inclusion

$$\dot{x}(t) + \partial\Phi(x(t)) + B(x(t)) \ni 0. \quad (4.2.59)$$

Since the operator $A = \partial\Phi + B$ is maximal monotone, (4.2.59) is relevant to the general theory of semigroups generated by maximal monotone operators. For any Cauchy data $x_0 \in \text{dom}\partial\Phi$, there exists a unique strong solution of (4.2.59) which satisfies $x(0) = x_0$, see [23]. Moreover, by

a direct adaptation of the results of [23, Theorem 3.6], one can verify that there is a regularizing effect on the initial condition : for $x_0 \in \overline{\text{dom}\Phi}$, there exists a unique strong solution of (4.2.59) with Cauchy data $x(0) = x_0$, and which satisfies $x(t) \in \text{dom}\partial\Phi$ for all $t > 0$.

Let us suppose that $S \neq \emptyset$, where S still denotes the set of zeroes of $A = \partial\Phi + B$. Following Baillon-Brézis [16], each orbit of (4.2.59) converges weakly, in an ergodic way, to an equilibrium, which is an element of S . Note that the convergence theory of Bruck does apply separately to $\partial\Phi$ and B , which are demipositive, see [24]. But it is not known if the sum of the two operators $\partial\Phi + B$ is still demipositive. Indeed, it is not clear whether this notion is stable by sum. Thus, we are naturally led to perform a direct study of the convergence properties of the orbits of (4.2.59). Surprisingly, we have not found references to a previous systematic study of this question. Indeed, we are going to show that (4.2.59) has convergence properties which are similar to the regularized Newton-like dynamic. Then, we shall compare and show the differences between the two systems.

Theorem 4.2.24 *Suppose that $S \neq \emptyset$. Then, for any orbit $x(\cdot)$ of (4.2.59), the following properties hold :*

1. $\int_0^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty$, i.e., $x(\cdot)$ has a finite energy.
2. $x(\cdot)$ converges weakly to an element of S .
3. $B(x(\cdot))$ converges strongly to Bz , where Bz is uniquely defined for $z \in S$.

Proof. Let $x(\cdot) : [0, +\infty[\rightarrow \mathcal{H}$ be an orbit of (4.2.59). Equivalently, we set

$$v(t) \in \partial\Phi(x(t)) \tag{4.2.60}$$

$$\dot{x}(t) + v(t) + B(x(t)) = 0. \tag{4.2.61}$$

For any $z \in S$, let us define $h_z : [0, +\infty[\rightarrow \mathbb{R}^+$ by

$$h_z(t) := \frac{1}{2} \|x(t) - z\|^2.$$

Let us show that h_z is a Lyapunov function. By the classical derivation chain rule, and (4.2.61), for almost all $t \geq 0$

$$\frac{d}{dt} h_z(t) = \langle x(t) - z, \dot{x}(t) \rangle \tag{4.2.62}$$

$$= - \langle x(t) - z, v(t) + B(x(t)) \rangle. \tag{4.2.63}$$

Let us rewrite this last equality as

$$\frac{d}{dt} h_z(t) + \langle x(t) - z, v(t) + Bz \rangle + \langle x(t) - z, B(x(t)) - Bz \rangle = 0. \tag{4.2.64}$$

Since $z \in S$, we have $-Bz \in \partial\Phi(z)$. Moreover, $v(t) \in \partial\Phi(x(t))$. By monotonicity of $\partial\Phi$, this gives

$$\langle x(t) - z, v(t) + Bz \rangle \geq 0. \tag{4.2.65}$$

Combining (4.2.64) with (4.2.65) we obtain

$$\frac{d}{dt} h_z(t) + \langle x(t) - z, B(x(t)) - Bz \rangle \leq 0. \tag{4.2.66}$$

By cocoercivity of B , we deduce that

$$\frac{d}{dt}h_z(t) + \beta \|B(x(t)) - Bz\|^2 \leq 0. \quad (4.2.67)$$

From this, we readily obtain

$$t \mapsto h_z(t) \text{ is a decreasing function,} \quad (4.2.68)$$

and after integration of (4.2.67)

$$\int_0^{+\infty} \|B(x(t)) - Bz\|^2 dt < +\infty. \quad (4.2.69)$$

Let us return to (4.2.67). By (4.2.61), we equivalently have

$$\frac{d}{dt}h_z(t) + \beta \|\dot{x}(t) + v(t) + Bz\|^2 \leq 0. \quad (4.2.70)$$

After developing

$$\frac{d}{dt}h_z(t) + \beta \|\dot{x}(t)\|^2 + \beta \|v(t) + Bz\|^2 + 2\beta \langle \dot{x}(t), v(t) + Bz \rangle \leq 0. \quad (4.2.71)$$

By using the derivation chain rule for a convex lower semicontinuous function, see Lemma 4.2.9

$$\frac{d}{dt}\Phi(x(t)) = \langle v(t), \dot{x}(t) \rangle, \quad (4.2.72)$$

we can rewrite (4.2.71) as

$$\frac{d}{dt}[h_z(t) + 2\beta\Phi(x(t)) + 2\beta \langle x(t), Bz \rangle] + \beta \|\dot{x}(t)\|^2 + \beta \|v(t) + Bz\|^2 \leq 0. \quad (4.2.73)$$

Set

$$k_z(t) := \Phi(x(t)) - \Phi(z) + \langle Bz, x(t) - z \rangle. \quad (4.2.74)$$

Since $-Bz \in \partial\Phi(z)$, by the convex subdifferential inequality, we have $k_z(t) \geq 0$. Let us rewrite (4.2.73) as

$$\frac{d}{dt}[h_z(t) + 2\beta k_z(t)] + \beta \|\dot{x}(t)\|^2 + \beta \|v(t) + Bz\|^2 \leq 0. \quad (4.2.75)$$

Since $h_z + 2\beta k_z$ is nonnegative, by integration of (4.2.75) we infer

$$\int_0^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty. \quad (4.2.76)$$

That's item 1. In order to prove item 2., which is the weak convergence property of x , we use Opial's lemma 4.2.10, with S equal to the solution set of problem (4.2.1). By (4.2.68), $t \mapsto h_z(t)$ is a decreasing function, and hence $\lim \|x(t) - z\|$ exists. Let us complete the verification of the hypothesis of Opial's lemma, by showing that every weak sequential cluster point of x belongs to S . Let $\bar{x} = w - \lim x(t_n)$ for some sequence $t_n \rightarrow +\infty$. By the general theory of semigroups generated by maximal monotone operators, we have that

$$t \mapsto \|(\partial\Phi(x(t)) + B(x(t)))^0\| \quad (4.2.77)$$

is a nonincreasing function, where $(\partial\Phi(x(t)) + B(x(t)))^0$ is the element of minimal norm of the closed convex set $\partial\Phi(x(t)) + B(x(t))$. Since $\dot{x}(t) = -(\partial\Phi(x(t)) + B(x(t)))^0$ for almost all $t \geq 0$, we deduce from (4.2.76) that

$$\int_0^{+\infty} \|(\partial\Phi(x(t)) + B(x(t)))^0\|^2 dt < +\infty. \quad (4.2.78)$$

Since $t \mapsto \|(\partial\Phi(x(t)) + B(x(t)))^0\|$ is nonincreasing, it converges and, by (4.2.78) its limit is equal to zero. Thus, by taking $w(t) = (\partial\Phi(x(t)) + B(x(t)))^0$, we have obtained the existence of a mapping w which verifies : $w(t) \in (\partial\Phi + B)(x(t))$ for all $t > 0$, and $w(t) \rightarrow 0$ strongly in \mathcal{H} , as $t \rightarrow +\infty$. From $w(t_n) \in (\partial\Phi + B)(x(t_n))$, by the demiclosedness property of the maximal monotone operator $A = \partial\Phi + B$, we obtain $A(\bar{x}) = \partial\Phi(\bar{x}) + B(\bar{x}) \ni 0$, that is $\bar{x} \in S$.

Let us now prove item 3. Set $F_3(t) = \|B(x(t)) - Bz\|$. By (4.2.69) $F_3 \in L^2([0, +\infty[)$. Since B is Lipschitz continuous and $\dot{x} \in L^2([0, +\infty[)$ we obtain $\frac{d}{dt}F_3 \in L^2([0, +\infty[)$. By Lemma 4.2.11, we deduce that $\lim_{t \rightarrow +\infty} F_3(t) = 0$, which is our claim. \square

Remark 4.2.25 The strong convergence of orbits falls within the general theory of semigroup of contractions generated by a maximal monotone operator A . It is satisfied if A is strongly monotone, or Φ boundedly inf-compact (note that in the proof of Theorem 4.2.24 we have shown that $\Phi(x(t))$ converges, and thus is bounded). Also note that, following [23, Theorem 3.13], if $\text{int}S = \text{int}A^{-1}(0) \neq \emptyset$, then (4.2.59) has orbits whose total variation is bounded, and hence which converge strongly.

Remark 4.2.26 Let us compare the asymptotic behavior of the orbits of the semigroup generated by $-(\partial\Phi + B)$ with the orbits of the Newton-like regularized system. Since both converge weakly to equilibria, the point is compare their rate of convergence. For simplicity take $B = 0$, and Φ convex differentiable. Thus the point is : at which rate does $\nabla\Phi(x(t))$ converges to zero ?
a) For the semigroup, the standard estimation is the linear convergence :

$$\|\nabla\Phi(x(t))\| \leq \frac{C}{t}.$$

Indeed, without any further assumption on Φ or \mathcal{H} , this is the best known general estimate. Indeed, in infinite dimensional spaces one can exhibit orbits of the gradient flow which have infinite length, this is a consequence of Baillon counterexample [15]. Note that, in finite dimensional spaces, the corresponding result is not known [29].

b) For the Newton-like regularized system, $v(t) = \nabla\Phi(x(t))$ satisfies the differential equation

$$\lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) = 0.$$

By taking $\lambda(t) = ce^{-t}$, we have the following estimation (see [13, Proposition 5.1])

$$\|\nabla\Phi(x(t))\| \leq Ce^{-t}.$$

These results naturally suggest to extend our results to the case of a vanishing regularization parameter (see [12], [13] for some first results in this direction).

Implicit/explicit time discretization : FB algorithm

The discretization of (4.2.59) with respect to the time variable t , in an implicit way with respect to the nonsmooth term $\partial\Phi$, and explicit with respect to the smooth term B , and with constant step size $h > 0$, gives

$$\frac{x_{k+1} - x_k}{h} + \partial\Phi(x_{k+1}) + B(x_k) \ni 0. \quad (4.2.79)$$

Equivalently

$$x_{k+1} = (I + h\partial\Phi)^{-1}(x_k - hB(x_k)). \quad (4.2.80)$$

This is the classical forward-backward algorithm, whose convergence has been well established. The weak convergence of (x_k) to an element of S is obtained under the stepsize limitation : $0 < h < 2\beta$. One can consult [31], [19, Theorem 25.8], for the proof, and some further extensions of this result.

4.2.5 Proximal-gradient dynamics and relaxed FB algorithms

First recall some standard facts about the continuous gradient-projection system. This will lead us to consider a more general proximal-gradient dynamic. Then we will examine the corresponding relaxed FB algorithms, obtained by time discretization.

Gradient-projection dynamics

First take $\Phi = \delta_C$ equal to the indicator function of a closed convex set $C \subset \mathcal{H}$, and $B = \nabla\Psi$, the gradient of a convex differentiable function $\Psi : \mathcal{H} \rightarrow \mathbb{R}$. The semigroup of contractions, generated by $-A$, which has been studied in the previous section, specializes in gradient-projection system

$$\dot{x}(t) = \text{proj}_{T_C(x(t))}(-\nabla\Psi(x(t))), \quad (4.2.81)$$

where $T_C(x)$ is the tangent cone to C at $x \in C$. This is a direct consequence of the lazy property satisfied by the orbits of the semigroup of contractions, generated by $-A$, see [23], and of the Moreau decomposition theorem in a Hilbert space (with respect to the tangent cone $T_C(x)$ and its polar cone $N_C(x)$). From the perspective of optimization, this system has several drawbacks. The orbits ignore the constraint until they meet the boundary of C . Moreover, the vector field which governs the dynamic is discontinuous (at the boundary of the constraint). The following system first considered by Antipin [2], and Bolte [22] overcomes some of these difficulties :

$$\dot{x}(t) + x(t) - \text{proj}_C(x(t) - \mu\nabla\Psi(x(t))) = 0. \quad (4.2.82)$$

It can be introduced in a natural way, by rewriting the optimality condition

$$\nabla\Psi(x) + N_C(x) \ni 0 \quad (4.2.83)$$

as a fixed point problem

$$x - \text{proj}_C(x - \mu\nabla\Psi(x)) = 0, \quad (4.2.84)$$

where μ is a positive parameter (arbitrarily chosen). Note that the stationary points of (4.2.82) are precisely the solutions of (4.2.84). This dynamic is governed by a Lipschitz continuous vector field, and the orbits are classical solutions, i.e., continuously differentiable. Its properties are summarized in the following proposition, see [22].

Proposition 4.2.27 *Let $\Psi : \mathcal{H} \rightarrow \mathbb{R}$ be a convex differentiable function, whose gradient is Lipschitz continuous on bounded sets. Let C be a closed convex set in \mathcal{H} , and suppose that Ψ is bounded from below on C . Then, for any $x_0 \in \mathcal{H}$, there exists a unique classical global solution $x : [0, +\infty[\rightarrow \mathcal{H}$ of the Cauchy problem for the relaxed gradient-projection dynamical system*

$$\begin{cases} \dot{x}(t) + x(t) - \text{proj}_C(x(t) - \mu \nabla \Psi(x(t))) = 0; \\ x(0) = x_0. \end{cases} \quad (4.2.85)$$

The following asymptotic properties are satisfied :

- i) If $S = \text{argmin}_C \Psi$ is nonempty, then $x(t)$ converges weakly to some $x_\infty \in S$, as $t \rightarrow +\infty$.*
- ii) If moreover $x_0 \in C$, then $x(t) \in C$ for all $t \geq 0$, $\Psi(x(t))$ decreases to $\inf_C \Psi$ as t increases to $+\infty$, and*

$$\mu \frac{d}{dt} \Psi(x(t)) + \|\dot{x}(t)\|^2 \leq 0. \quad (4.2.86)$$

Proximal-gradient dynamics

Let us now return to our setting : $A = \partial\Phi + B$, where Φ is a closed convex proper function, and B is a monotone cocoercive operator (the previous case corresponds to $\Phi = \delta_C$ and $B = \nabla\Psi$, with $\nabla\Psi$ Lipschitz continuous). As a natural extension of (4.2.82), let us consider the differential system

$$\dot{x}(t) + x(t) - \text{prox}_{\mu\Phi}(x(t) - \mu B(x(t))) = 0. \quad (4.2.87)$$

We shall see that the explicit discretization of this system gives the relaxed FB algorithm. The vector field which governs (4.2.87) is Lipschitz continuous. Hence, for any $x_0 \in \mathcal{H}$, the corresponding Cauchy problem has a unique global classical solution. As far as we know, the convergence properties of this system have not been studied in this framework. Let us state our results.

Theorem 4.2.28 *Let $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex lower semicontinuous proper function, and B a maximal monotone operator which is β -cocoercive. Suppose that $S = \{z \in \mathcal{H}; \partial\Phi(z) + Bz \ni 0\}$, the solution set of (4.2.1), is nonempty.*

For any $x_0 \in \mathcal{H}$, let $x : [0, +\infty[\rightarrow \mathcal{H}$ be the unique classical global solution of the Cauchy problem for the proximal-gradient dynamical system

$$\begin{cases} \dot{x}(t) + x(t) - \text{prox}_{\mu\Phi}(x(t) - \mu B(x(t))) = 0; \\ x(0) = x_0. \end{cases} \quad (4.2.88)$$

Then, the following asymptotic properties are satisfied :

- 1. Suppose that $0 < \mu < 4\beta$, then*
 - i) $x(t)$ converges weakly to some $x_\infty \in S$, as $t \rightarrow +\infty$.*
 - ii) $B(x(t))$ converges strongly to Bz as $t \rightarrow +\infty$, where Bz is uniquely defined for $z \in S$. Moreover, $\int_0^\infty \|B(x(t)) - Bz\|^2 dt < +\infty$.*
 - iii) $\lim_{t \rightarrow +\infty} \dot{x}(t) = 0$, and $\int_0^\infty \|\dot{x}(t)\|^2 dt < +\infty$.*
- 2.. Suppose that $B = \nabla\Psi$, where Ψ is a convex differentiable function. Then, for arbitrary $\mu > 0$, the above properties i), ii), iii) are satisfied.*

Proof. We rely on a Lyapunov analysis. Take $z \in S$. Equivalently

$$-Bz \in \partial\Phi(z). \quad (4.2.89)$$

Set $\xi(t) := x(t) - \mu B(x(t))$. By definition of $\text{prox}_{\mu\Phi}$, we have

$$\frac{1}{\mu}(\xi(t) - \text{prox}_{\mu\Phi}\xi(t)) \in \partial\Phi(\text{prox}_{\mu\Phi}\xi(t)).$$

Since $\text{prox}_{\mu\Phi}\xi(t) = x(t) + \dot{x}(t)$, the above equation can be written in equivalent way

$$-B(x(t)) - \frac{1}{\mu}\dot{x}(t) \in \partial\Phi(x(t) + \dot{x}(t)). \quad (4.2.90)$$

By the monotonicity property of the operator $\partial\Phi$, and (4.2.89), (4.2.90), we obtain

$$0 \geq \left\langle x(t) - z + \dot{x}(t), B(x(t)) - Bz + \frac{1}{\mu}\dot{x}(t) \right\rangle. \quad (4.2.91)$$

Equivalently

$$0 \geq \frac{1}{2\mu} \frac{d}{dt} \|x(t) - z\|^2 + \frac{1}{\mu} \|\dot{x}(t)\|^2 + \langle B(x(t)) - Bz, x(t) - z \rangle + \langle \dot{x}(t), B(x(t)) - Bz \rangle. \quad (4.2.92)$$

1. Let us first examine the general case, B β -cocoercive. From (4.2.92), it follows that

$$0 \geq \frac{1}{2\mu} \frac{d}{dt} \|x(t) - z\|^2 + \frac{1}{\mu} \|\dot{x}(t)\|^2 + \beta \|B(x(t)) - Bz\|^2 + \langle \dot{x}(t), B(x(t)) - Bz \rangle. \quad (4.2.93)$$

Let us introduce $\alpha > 0$, a positive parameter. In order to estimate the last term in (4.2.93), we use Cauchy-Schwarz inequality, and the following elementary inequality

$$\|\dot{x}(t)\| \|B(x(t)) - Bz\| \leq \frac{1}{2\alpha} \|\dot{x}(t)\|^2 + \frac{\alpha}{2} \|B(x(t)) - Bz\|^2. \quad (4.2.94)$$

From (4.2.93) and (4.2.94) we deduce that

$$0 \geq \frac{1}{2\mu} \frac{d}{dt} \|x(t) - z\|^2 + \left(\frac{1}{\mu} - \frac{1}{2\alpha}\right) \|\dot{x}(t)\|^2 + \left(\beta - \frac{\alpha}{2}\right) \|B(x(t)) - Bz\|^2. \quad (4.2.95)$$

Choose α such that $\frac{1}{\mu} - \frac{1}{2\alpha} > 0$ and $\beta - \frac{\alpha}{2} > 0$. This is equivalent to find $\frac{\mu}{2} < \alpha < 2\beta$, which is possible iff $\mu < 4\beta$, that's precisely our condition on parameters μ and β . When this condition is satisfied, taking (for example) $\alpha = \frac{1}{2}(\frac{\mu}{2} + 2\beta) = \frac{\mu}{4} + \beta$ in (4.2.95), we obtain

$$0 \geq \frac{1}{2\mu} \frac{d}{dt} \|x(t) - z\|^2 + \frac{4\beta - \mu}{\mu(\mu + 4\beta)} \|\dot{x}(t)\|^2 + \frac{1}{8}(4\beta - \mu) \|B(x(t)) - Bz\|^2. \quad (4.2.96)$$

From (4.2.96), it follows that, for any $z \in S$, $t \mapsto \|x(t) - z\|$ is a decreasing function, and hence $\lim \|x(t) - z\|$ exists. Moreover, by integration of (4.2.96), we obtain

$$\int_0^\infty \|\dot{x}(t)\|^2 dt < +\infty, \quad (4.2.97)$$

$$\int_0^{\infty} \|B(x(t)) - Bz\|^2 dt < +\infty. \quad (4.2.98)$$

Since B is Lipschitz continuous, and $\|\dot{x}(\cdot)\|$ belongs to $L^2(0, +\infty)$, we have $\frac{d}{dt}B(x) \in L^2(0, +\infty)$. Hence, by (4.2.98), $B(x) - Bz$ and its derivative belong to $L^2(0, +\infty)$. By Lemma 4.2.11 we infer

$$\lim_{t \rightarrow +\infty} B(x(t)) = Bz \quad (4.2.99)$$

where Bz is uniquely defined for $z \in S$. On the other hand, by (4.2.87) and (4.2.97), \dot{x} and its derivative belong to $L^2(0, +\infty)$. By Lemma 4.2.11 we infer

$$\lim_{t \rightarrow +\infty} \dot{x}(t) = 0. \quad (4.2.100)$$

By Opial lemma 4.2.10, in order to obtain the weak convergence of the orbit x , we just need to prove that any weak sequential cluster point of x belongs to S . Let \bar{x} be a weak sequential cluster point of x , i.e., $\bar{x} = w - \lim x(t_n)$ for some sequence $t_n \rightarrow +\infty$. In order to pass to the limit on (4.2.87), we rewrite it as

$$x(t) - \mu B(x(t)) - \text{prox}_{\mu\Phi}(x(t) - \mu B(x(t))) = -\dot{x}(t) - \mu B(x(t)), \quad (4.2.101)$$

and use the demiclosedness property of the maximal monotone operator $I - \text{prox}_{\mu\Phi}$. Since $x(t_n) - \mu B(x(t_n)) \rightharpoonup \bar{x} - \mu Bz$, and $\dot{x}(t) + \mu B(x(t)) \rightarrow \mu Bz$ strongly, we obtain

$$\bar{x} - \mu Bz - \text{prox}_{\mu\Phi}(\bar{x} - \mu Bz) = -\mu Bz. \quad (4.2.102)$$

Equivalently,

$$\bar{x} = \text{prox}_{\mu\Phi}(\bar{x} - \mu Bz), \quad (4.2.103)$$

that is

$$\partial\Phi(\bar{x}) + Bz \ni 0.$$

Since B is maximal monotone, it is demiclosed, and hence $B\bar{x} = Bz$. Thus

$$\partial\Phi(\bar{x}) + B\bar{x} \ni 0,$$

and $\bar{x} \in S$, which completes the proof.

2. Now consider $B = \nabla\Psi$, where Ψ is a convex differentiable function (a special case of cocoercive operator), and show that we can conclude to the same convergence results, without making any restrictive assumption on $\mu > 0$. Let us return to (4.2.93). By using

$$\langle \dot{x}(t), \nabla\Psi(x(t)) - \nabla\Psi(z) \rangle = \frac{d}{dt}[\Psi(x(t)) - \langle \nabla\Psi(z), x(t) \rangle],$$

we can rewrite (4.2.93) as

$$0 \geq \frac{d}{dt} \left[\frac{1}{2\mu} \|x(t) - z\|^2 + \Psi(x(t)) - \Psi(z) - \langle \nabla\Psi(z), x(t) - z \rangle \right] + \frac{1}{\mu} \|\dot{x}(t)\|^2 + \beta \|B(x(t)) - Bz\|^2. \quad (4.2.104)$$

Since $\Psi(x(t)) - \Psi(z) - \langle \nabla\Psi(z), x(t) - z \rangle$ is nonnegative, by a similar argument as before we see that $B(x(\cdot))$ converges strongly to Bz , where Bz is uniquely defined for $z \in S$, $\lim_{t \rightarrow \infty} \dot{x}(t) = 0$, and $\|\dot{x}\| \in L^2(0, +\infty)$. Moreover, any weak sequential cluster point of x belongs to S . But

unlike the previous situation, we do not know if the limit of $\|x(t) - z\|$ exists. Instead we have $\lim E(t, z)$ exists for any $z \in S$, where

$$E(t, z) := \frac{1}{2\mu} \|x(t) - z\|^2 + \Psi(x(t)) - \Psi(z) - \langle \nabla \Psi(z), x(t) - z \rangle. \quad (4.2.105)$$

Following the arguments in [22], we will show that this implies that x has a unique weak sequential cluster point, which clearly implies the weak convergence of the whole sequence. Let z_1 and z_2 two weak sequential cluster points of x , i.e., $z_1 = w\text{-}\lim x(t_n)$, and $z_2 = w\text{-}\lim x(t'_n)$, for some sequences $t_n \rightarrow +\infty$ and $t'_n \rightarrow +\infty$. We already obtained that z_1 and z_2 belong to S . Hence $E(t, z_1)$ and $E(t, z_2)$ converge as $t \rightarrow +\infty$, as well as $E(t, z_1) - E(t, z_2)$. We deduce that the following limit exists

$$\lim_{t \rightarrow +\infty} \left[\frac{1}{\mu} \langle x(t), z_2 - z_1 \rangle + \langle \nabla \Psi(z_2) - \nabla \Psi(z_1), x(t) \rangle \right].$$

Thus the limits obtained by successively replacing t by t_n and t'_n are equal, which gives

$$\frac{1}{\mu} \langle z_1, z_2 - z_1 \rangle + \langle \nabla \Psi(z_2) - \nabla \Psi(z_1), z_1 \rangle = \frac{1}{\mu} \langle z_2, z_2 - z_1 \rangle + \langle \nabla \Psi(z_2) - \nabla \Psi(z_1), z_2 \rangle.$$

Equivalently

$$\frac{1}{\mu} \|z_2 - z_1\|^2 + \langle \nabla \Psi(z_2) - \nabla \Psi(z_1), z_2 - z_1 \rangle = 0,$$

which, by monotonicity of $\nabla \Psi$, gives $z_1 = z_2$. □

Remark 4.2.29 Under the more restrictive assumption, $0 < \mu < 2\beta$, using the results of section 4.2.2, the operator that governs the dynamical system is of the form $I - T$, where T is a contraction. Accordingly, the operator $I - T$ is demipositive, and the weak convergence of x is a direct consequence of Bruck Theorem [24]. It is an open question whether the convergence property is true for a general cocoercive operator B , without restriction on $\mu > 0$.

Relaxed forward-backward algorithms

The explicit discretization of the regular dynamic (4.2.87) with respect to the time variable t , with constant step size $h > 0$, gives

$$\frac{x_{k+1} - x_k}{h} + x_k - \text{prox}_{\mu\Phi}(x_k - \mu B(x_k)) = 0. \quad (4.2.106)$$

Equivalently

$$x_{k+1} = (1 - h)x_k + h \text{prox}_{\mu\Phi}(x_k - \mu B(x_k)).$$

This is the relaxed forward-backward algorithm, whose convergence properties are well known. The weak convergence of (x_k) to an element of S is obtained under the stepsize limitation : $0 < \mu < 2\beta$, and $0 < h \leq 1$. One can consult [19, Theorem 25.8], for the proof, and some further extensions of this result.

4.2.6 Perspective

Our work can be considered from two perspectives : numerical splitting methods in optimization, and modeling in decision sciences and engineering.

1. In recent years, there has been a great interest in the forward-backward methods, especially in the signal/image processing, and sparse optimization. A better understanding of these methods is a key to obtain further developments, and improvement of the methods : fast converging algorithms, nonconvex setting, multiobjective optimization, ... are crucial points to consider in the future. To cite some of these topics, in [3] the convergence of the classical forward-backward method, in a nonconvex nonsmooth framework, has been proved for functions satisfying the Kurdyka-Lojasiewicz inequality, a large class containing the semi-algebraic functions. The proof finds its roots in a dynamical argument. It is an open question to know if some other form of the FB algorithms works in this setting. Even for the relaxed FB algorithm this is an open question.

Similarly, the Nesterov method for obtaining convergence rate $O(\frac{1}{k^2})$, is known for the classical forward-backward algorithm in the case of convex minimization (for the exact scheme, and convergence of the values), see [32], [20] (FISTA method). It would be very interesting to know if the method can be adapted to other forms of these algorithms.

It turns out that there is a rich family of forward-backward algorithms. In this article, we have considered three classes of these algorithms. The link with the dynamical systems is a valuable tool for studying these algorithms, and to discover new one. The comparison between the algorithms that are obtained by time discretization of the continuous dynamics is a delicate subject, which is the subject of current research.

2. Many equilibrium problems in decision or physical sciences may be written either as convex minimization problem or as a search for a fixed point of a contraction. Often these two aspects are present simultaneously. For example, in game theory, agents may adopt strategies involving cooperative aspects (potential games) and noncooperative aspects. Nash equilibrium formulation can lead to a convex-concave saddle value problem, and non-potential monotone operators. An abundant literature has been devoted to finding common solutions of these problems. In contrast, our approach aims at finding a compromise solution of these two different types of problems. A basic ingredient is the resolution of $\partial\Phi(x) + Bx \ni 0$. An interesting direction for future research would be to consider a multicriteria dynamical process associated to the two operators $\partial\Phi$ and B , in line of the recent article [8].

The selection of equilibria with desirable properties is an important issue in decision sciences. With the introduction of regularization terms tending asymptotically to zero, not too quickly (eg Tikhonov type), the dynamic equilibrium approach provides an asymptotic hierarchical selection. There is an extensive literature on this topic, see [5], [6], [17], [21], [22], [25], [27], [28], [30], and references therein. It is an issue that is largely unexplored for the systems considered in this article.

Bibliographie

- [1] B. Abbas, H. Attouch, and B. F. Svaiter, Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces, *J. Optim. Theory Appl.*, **161**(2) (2014), pp. 331-360.
- [2] A.S. Antipin, Minimization of convex functions on convex sets by means of differential equations, *Differential Equations*, **30** (2010), pp. 3246–3270.
- [3] H. Attouch, J. Bolte, and B. F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems : proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods, *Mathematical Programming*, **137**(1) (2013), pp. 91–129.
- [4] H. Attouch, L. M. Briceño-Arias, and P. L. Combettes, A parallel splitting method for coupled monotone inclusions, *SIAM J. Control Optim.*, **48**(5) (2010), pp. 3246–3270.
- [5] H. Attouch, R. Cominetti, A dynamical approach to convex minimization coupling approximation with the steepest descent method, *J. Differential Equations*, **128**(2) (1996), pp. 519–540.
- [6] H. Attouch, M.-O. Czarnecki, Asymptotic behavior of coupled dynamical systems with multiscale aspects, *J. Differential Equations*, **248**(6) (2010), pp. 1315-1344.
- [7] H. Attouch, M.-O. Czarnecki, and J. Peypouquet, Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities, *SIAM J. Optim.*, **21**(4) (2011), pp. 1251–1274..
- [8] H. Attouch, X. Goudou, A continuous gradient-like dynamical approach to Pareto-optimization in Hilbert spaces, *Set-Valued and Variational Analysis*, **22**(1) (2014), pp. 189–219.
- [9] H. Attouch, P.E. Maingé, Asymptotic behavior of second order dissipative evolution equations combining potential with non-potential effects, *ESAIM Control Optim. Calc. of Var.*, **17**(3) (2011), pp. 836-857.
- [10] H. Attouch, J. Peypouquet, and P. Redont, A dynamical approach to an inertial forward-backward algorithm for convex minimization, *SIAM J. Optim.*, **24**(1) (2014), pp. 232–256.
- [11] H. Attouch, J. Peypouquet, and P. Redont, Forward-backward and backward-forward algorithms for structured monotone inclusions, working paper, (2014).
- [12] H. Attouch, P. Redont, and B. F. Svaiter, Global convergence of a closed-loop regularized Newton method for solving monotone inclusions in Hilbert spaces, *J. Optim. Theory Appl.*, **157** (2013), pp. 624–650.
- [13] H. Attouch, B. F. Svaiter, A continuous dynamical Newton-like approach to solving monotone inclusions, *SIAM J. Control Optim.*, **49** (2011), pp. 574–598.
- [14] H. Attouch, M. Théra, A general duality principle for the sum of two operators, *J. Convex Analysis*, **3** (1996), pp. 1–24.

- [15] J.-B. Baillon, *Un exemple concernant le comportement asymptotique de la solution du problème $\frac{du}{dt} + \partial\phi(u) \ni 0$* , J. Funct. Anal., **28** (1978), pp. 369–376.
- [16] J.-B. Baillon and H. Brézis, *Une remarque sur le comportement asymptotique des semi-groupes non linéaires*, Houston J. Math., **2** (1976), pp. 5–7.
- [17] J.-B. Baillon and R. Cominetti, *A convergence result for non-autonomous subgradient evolution equations and its application to the steepest descent exponential penalty trajectory in linear programming*, J. Funct. Anal., **187** (2001), pp. 263–273.
- [18] J.-B. Baillon and G. Haddad, *Quelques propriétés des opérateurs angle-bornés et n -cycliquement monotones*, Israël J. Math., **26** (1977), pp. 137–150.
- [19] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer, New York, (2011).
- [20] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM Journal on Imaging Sciences, **2**(1) (2009), pp. 183–202.
- [21] W. Bian and X. Xue, *Asymptotic behavior analysis on multivalued evolution inclusion with projection in Hilbert space*, Optimization, DOI : 10.1080/02331934.2013.811668, (2013).
- [22] J. Bolte, *Continuous gradient projection method in Hilbert spaces*, J. Optim. Theory Appl., **119**(2) (2003), pp. 235–259.
- [23] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland/Elsevier, New-York, (1973).
- [24] R.E. Bruck, *Asymptotic convergence of nonlinear contraction semigroups in Hilbert spaces*, J. Funct. Anal., **18** (1975), pp. 15–26.
- [25] A. Cabot, *The steepest descent dynamical system with control. Applications to constrained minimization*, ESAIM Control Optim. Calc. Var., **10** (2004), pp. 243–258.
- [26] P. L. Combettes, *Solving monotone inclusions via compositions of nonexpansive averaged operators*, Optimization, **53** (2004), pp. 475–504.
- [27] P. L. Combettes and S. A. Hirstoaga, *Approximating curves for nonexpansive and monotone operators*, Journal of Convex Analysis, **13**(3-4) (2006), pp. 633–646.
- [28] R. Cominetti, J. Peypouquet, and S. Sorin, *Strong asymptotic convergence of evolution equations governed by maximal monotone operators with Tikhonov regularization*, J. Differential Equations, **245** (2008), pp. 3753–3763.
- [29] A. Daniilidis, *Gradient dynamical systems, tame optimization and applications*, Lecture Notes, Spring School on Variational Analysis Paseky nad Jizerou, Czech Republic, April 20-24, 2009.
- [30] S. A. Hirstoaga, *Approximation et résolution de problèmes d'équilibre, de point fixe et d'inclusion monotone*, PhD thesis, Université Pierre et Marie Curie, Paris VI, (2006).
- [31] P.L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM Journal on Numerical Analysis, **16** (1979), pp. 964–979.
- [32] Y. E. Nesterov, *A method for solving the convex programming problem with convergence rate $O(1/k^2)$* , Dokl. Akad. Nauk SSSR, **269**(3) (1983), pp. 543–547.
- [33] Z. Opial, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. Amer. Math. Soc., **73** (1967), pp. 591–597.

[34] D. L. Zhu and P. Marcotte, Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities, *J. Optim.*, **6** (1996), pp. 714–726.

4.3 Numerical Illustrations

4.3.1 Introduction

In a Hilbert framework, we consider continuous and discrete dynamical systems which aim at solving structured monotone inclusions of the following type

$$(\mathcal{P}) \quad \partial\Phi(x) + Bx \ni 0,$$

where $\partial\Phi$ is the subdifferential of a convex lower semicontinuous function Φ , and B is a monotone cocoercive operator. We rely on Abbas and Attouch (2014) (Section 4.2), where several convergence results have been obtained for these dynamics and algorithms. We first list them, then give numerical illustrations in a model situation.

1. We have first studied the regularized Newton dynamic which was considered in Abbas, Attouch, Svaiter *JOTA* (2014), and developed in Abbas and Attouch (2014). The analysis focuses on the convergence properties (as $t \rightarrow +\infty$) of the orbits of the system (4.3.1)-(4.3.2)

$$v(t) \in \partial\Phi(x(t)), \tag{4.3.1}$$

$$\lambda \dot{x}(t) + \dot{v}(t) + v(t) + B(x(t)) = 0. \tag{4.3.2}$$

Indeed, by introducing the new unknown function $y(\cdot) = x(\cdot) + \mu v(\cdot)$, and setting $\mu = \frac{1}{\lambda}$, (4.3.1) and (4.3.2) can be written equivalently as

$$x(t) = \text{prox}_{\mu\Phi}(y(t)), \tag{4.3.3}$$

$$\dot{y}(t) + y(t) - \text{prox}_{\mu\Phi}(y(t)) + \mu B(\text{prox}_{\mu\Phi}(y(t))) = 0. \tag{4.3.4}$$

The above system makes use only of B , and of the proximal mapping associated to $\mu\Phi$, which are both Lipschitz continuous operators. Hence it is relevant of the Cauchy-Lipschitz theorem. Its *explicit* discretization gives (with a constant step size $h > 0$) the following backward-forward algorithm : $(x_k, y_k) \rightarrow (x_k, y_{k+1}) \rightarrow (x_{k+1}, y_{k+1})$

$$(\text{BF}) \quad \begin{cases} x_k = \text{prox}_{\mu\Phi}(y_k) \\ y_{k+1} = (1-h)y_k + h(x_k - \mu B(x_k)) \end{cases} \tag{4.3.5}$$

Under the assumption $0 < h < \delta := \frac{1}{2} + \inf\left\{1, \frac{\beta}{\mu}\right\}$, and $0 < \mu < 2\beta$, (BF) generates sequences that converge weakly to equilibria.

2. Then, we have considered a naturally related dynamical system, which is the semigroup of contractions generated by $-A$, $A = \partial\Phi + B$, whose orbits are the solution trajectories of the differential inclusion

$$\dot{x}(t) + \partial\Phi(x(t)) + B(x(t)) \ni 0. \tag{4.3.6}$$

The discretization of (4.3.6) with respect to the time variable t , in an implicit way with respect to the nonsmooth $\partial\Phi$, and explicit with respect to the smooth term B , and with constant step size $h > 0$, gives

$$x_{k+1} = (I + \mu\partial\Phi)^{-1}(x_k - \mu B(x_k)). \quad (4.3.7)$$

This is the classical forward–backward algorithm, whose convergence properties have been well established. The weak convergence of (x_k) to equilibria is obtained under the stepsize limitation : $0 < \mu < 2\beta$.

3. Finally, we consider the proximal-gradient dynamic which is associated to the reformulation of (\mathcal{P}) as a fixed point problem :

$$\dot{x}(t) + x(t) - \text{prox}_{\mu\Phi}(x(t) - \mu B(x(t))) = 0. \quad (4.3.8)$$

The explicit discretization of the regular dynamic (4.3.8) with respect to the time variable t , with constant step size $h > 0$, gives

$$x_{k+1} = (1 - h)x_k + h \text{prox}_{\mu\Phi}(x_k - \mu B(x_k)). \quad (4.3.9)$$

This is the relaxed forward–backward algorithm, whose convergence properties are well known. The weak convergence of (x_k) to an equilibria is obtained under the stepsize limitation :

$$0 < \mu < 2\beta \text{ and } 0 < h < \delta := \frac{1}{2} + \inf \left\{ 1, \frac{\beta}{\mu} \right\}.$$

One of the main advantages of the forward-backward algorithms is that the backward (implicit) step can handle nondifferentiable functions for which the proximal mapping can be easily computed. This is the case for many problems coming from sparse approximation and compressed sensing. As well, the method is efficient for solving hard-constrained problem for which the projection on the constraint can be easily computed.

4.3.2 Constrained least squares problems :

We consider ill-posed linear problems that are accompanied by a *priori* knowledge that the solutions belong to a closed and convex set. This particular formulation captures a wide class of problems. Examples : image reconstruction and signal restoration.

The data collection experiment is assumed to be linear and represented by the operator L

$$Lx = d + w \quad (4.3.10)$$

where $d \in \mathbb{R}^m$ is the collection of measurements, x is the unknown signal to be recovered, and w is the noise vector. Here x is assumed to be an element of a Hilbert space \mathcal{H} . The *priori* constraints on x belongs to a set C of admissible signals; C is assumed to be a closed and convex subset of \mathcal{H} .

In order to recover the signal x from the noisy measurement d , a common approach for this estimation problem is to solve the least squares minimization problem

$$\min_{x \in \mathcal{H}} \frac{1}{2} \|Lx - d\|^2. \quad (4.3.11)$$

A natural way to regularize this problem is to force the solutions to lie in a given closed set modeling a *priori* constraints. This leads to the following formulation

$$\min_{x \in C} \frac{1}{2} \|Lx - d\|^2 \quad (4.3.12)$$

When $\mathcal{H} = \mathbb{R}^n$, $L \neq 0$ is an $m \times n$ real matrix.

In that case, the above system specializes to the structured minimization problem (\mathcal{P}) with :

- $\Phi(x) = \delta_C(x)$ is the indicator function of a closed convex set $C \subset \mathbb{R}^n$, and $\text{prox}_{\mu\Phi}$ is independent of μ , and is equal to P_C , the projection operator on C .
- $\Psi(x) = \frac{1}{2} \|Lx - d\|^2$, is a convex C^1 function, $L \in \mathbb{R}^{m \times n} \setminus \{0\}$, $d \in \mathbb{R}^m$. Its gradient $\nabla\Psi : x \rightarrow L^T(Lx - d)$ is Lipschitz continuous with Lipschitz constant $\frac{1}{\beta} = \rho(L^T L)$, the spectral radius of the matrix $L^T L$, which is also its largest eigenvalue.

Let us now suppose that C is a hyperplane.

In general, hyperplanes play an important role in applications, particularly in signal restoration and tomography, whereas the projection onto hyperplanes can be calculated easily. In fact, if a hyperplane C is given by

$$C = \{x \in \mathbb{R}^n; \langle x, a \rangle = b\}, \quad (4.3.13)$$

for $a \in \mathbb{R}^n \setminus \{0\}$ and $b \in \mathbb{R}$, then, for every $x \in \mathbb{R}^n$

$$P_C(x) = x + \frac{(b - \langle x, a \rangle)}{\|a\|^2} a. \quad (4.3.14)$$

Let us now particularize the algorithms which have been described above to this situation : we take

$b \in \mathbb{R}$, $a \in \mathbb{R}^n$, $\|a\| = 1$, $L \in \mathbb{R}^{m \times n}$, $d \in \mathbb{R}^m$.

1. Let us consider the backward-forward algorithm (4.3.5) and apply it to the structured minimization problem (4.3.12)

$$x_k = y_k + (b - \langle y_k, a \rangle) a, \quad (4.3.15)$$

$$y_{k+1} = (1 - h) y_k + h (x_k - \mu L^T (Lx_k - d)). \quad (4.3.16)$$

The algorithm (4.3.15) - (4.3.16) generates sequences (x_k) converging to solutions of (\mathcal{P}) if the parameters h and μ satisfy $0 < h < \delta := \frac{1}{2} + \inf \left\{ 1, \frac{\beta}{\mu} \right\}$, and $0 < \mu < 2\beta$.

2. Then, we consider the classical forward-backward algorithm (4.3.7), and apply it to the structured minimization problem (4.3.12)

$$x_{k+1} = P_C (x_k - \mu L^T (Lx_k - d)), \quad (4.3.17)$$

$$x_{k+1} = (x_k - \mu L^T (Lx_k - d)) + (b - \langle x_k - \mu L^T (Lx_k - d), a \rangle) a \quad (4.3.18)$$

This is the classical forward-backward algorithm, whose convergence has been well established. The weak convergence of x_k to an equilibria is obtained under the stepsize limitation : $0 < \mu < 2\beta$.

3. Finally, we consider the forward-backward algorithm (4.3.9), and apply it to the structured minimization problem (4.3.12)

$$x_{k+1} = (1 - h) x_k + h P_C (x_k - \mu L^T (Lx_k - d)), \quad (4.3.19)$$

$$x_{k+1} = (1 - h) x_k + h ((x_k - \mu L^T (Lx_k - d)) + (b - \langle x_k - \mu L^T (Lx_k - d), a \rangle) a) \quad (4.3.20)$$

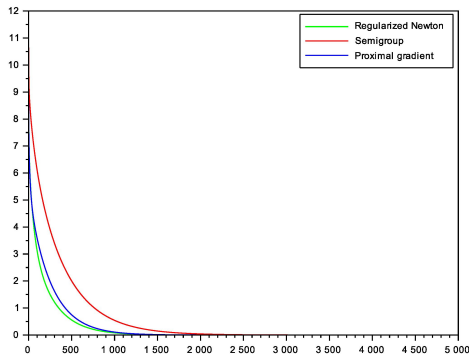
This is the relaxed forward-backward algorithm, whose convergence properties are well known. The weak convergence of x_k to an element of the solution set S is obtained under the stepsize limitation : $0 < \mu < 2\beta$, and $0 < h < \delta := \frac{1}{2} + \inf \left\{ 1, \frac{\beta}{\mu} \right\}$.

In the following simple illustration, we aim at comparing the performances of the three algorithms (4.3.15-4.3.16), (4.3.17-4.3.18), and (4.3.19-4.3.20).

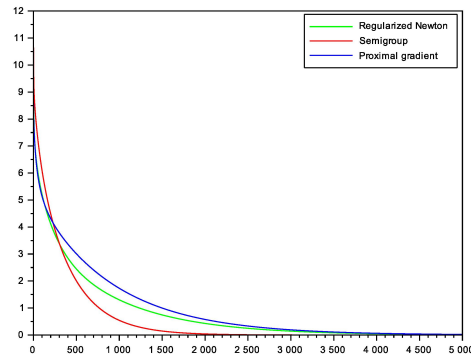
A simple illustration. The following example is intended to illustrate the behavior of the above algorithms. Consider the hyperplane $C = \{x \in \mathbb{R}^n; \langle x, a \rangle = b\}$, with $a = \frac{1}{\sqrt{n}} \cdot I_n = \frac{1}{\sqrt{n}} \cdot \text{ones}(n, 1)$, $b = 3$, and the system $Lx = d$.

We implement our algorithms in SCILAB with L is randomly generated in $\mathbb{R}^m \times \mathbb{R}^n$ $n = m = 20$, $d = I_m$, and starting from randomly generated initial points in $[-2, 2]^{20}$.

We compare the algorithms “regularized Newton dynamic”, “the semigroup of contractions generated by $-A$ ”, and “Proximal gradient” with the different cases of the parameters μ , β , and h for 5000 iterations.

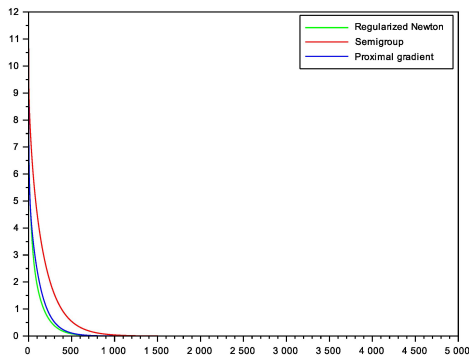


(a) $h > 1$

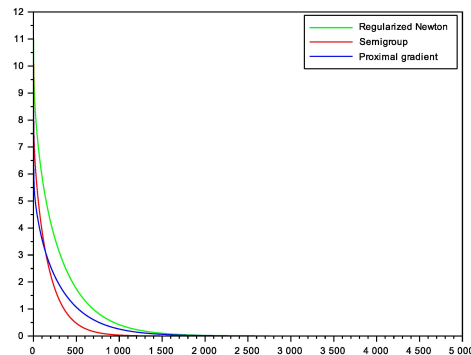


(b) $h < 1$

FIGURE 4.1 – Error values $\|x_k - \bar{x}\|$: case $0 < \mu < \beta$.

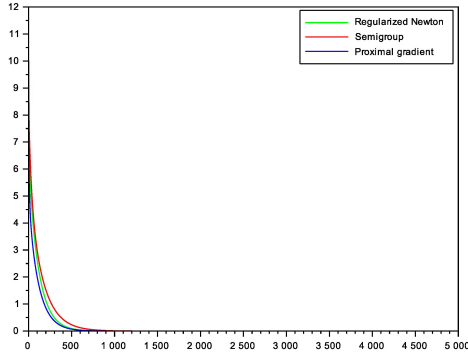


(a) $h > 1$

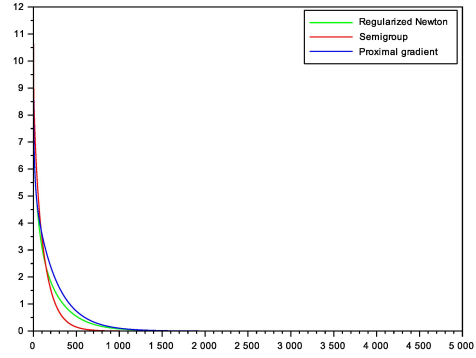


(b) $h < 1$

FIGURE 4.2 – Error values $\|x_k - \bar{x}\|$: case $\mu = \beta$.



(a) $h > 1$



(b) $h < 1$

FIGURE 4.3 – Error values $\|x_k - \bar{x}\|$: case $\beta < \mu < 2\beta$.

We note that, the three algorithms have similar performances, the regularized Newton dynamic and proximal gradient are very close, when $h > 1$ they are more effective than the semi-group of contractions algorithm. And more, the convergence of the trajectories toward the solution is faster than the case $h < 1$ where the algorithm of semi group of contraction generated by $-A$ is the most effective.

In addition when the parameters satisfy $\beta \leq \mu < 2\beta$ the convergence toward the solution is faster than the case $0 < \mu < \beta$.

4.3.3 Perspective

1. Application to the signal restoration and tomography by taking the minimization problem (4.3.12) with C is a hyperplane.
2. Application to image reconstruction and sparse approximation.

Chapitre 5

Méthode de viscosité et sélection asymptotique d'équilibre pour la dynamique de Newton régularisée

5.1 Description et résultats principaux :

Dans ce chapitre, nous nous intéressons à l'étude du comportement asymptotique des trajectoires de systèmes dynamiques de type Newton régularisé, dans lesquels on introduit un terme supplémentaire de viscosité évanescence de type Tikhonov. Étant donné une fonction convexe, semi-continue inférieurement, et propre $\Phi : H \rightarrow \mathbb{R} \cup \{+\infty\}$, que l'on cherche à minimiser, nous montrons que toute trajectoire de ce système "viscosifié" converge faiblement vers l'équilibre de norme minimale de $C = \operatorname{argmin}\Phi \neq \emptyset$.

Problème 5.1.1 *Soit H un espace de Hilbert réel*

$$\text{Trouver } x \in H \text{ tel que } \partial\Phi(x) \ni 0.$$

Attouch-Svaiter ont introduit dans SIAM (2011) un système dynamique de type Newton régularisé qui vise à trouver les zéros de l'opérateur maximal monotone $\partial\Phi$.

Le problème de Cauchy pour ce système de Newton régularisé s'écrit sous la forme suivante :

$$v(t) \in \partial\Phi(x(t)), \tag{5.1.1}$$

$$\lambda \dot{x}(t) + \dot{v}(t) + v(t) = 0. \tag{5.1.2}$$

$$x(0) = x_0, v(0) = v_0, v_0 \in \partial\Phi(x_0). \tag{5.1.3}$$

Dans ce qui suit, on supposera que λ , le terme de régularisation de type Levenberg - Marquardt, est une constante strictement positive.

Pour certaines considérations physiques, économiques et autres, une solution stationnaire particulière est plus intéressante que d'autres. Lorsqu'on a la convergence globale des trajectoires, une méthode classique pour atteindre un point d'équilibre particulier, consiste à introduire un terme évanescent dans le système, ne changeant pas l'ensemble des équilibres mais favorisant la convergence vers le type d'équilibre désiré. Cette méthode rentre dans le cadre général des solutions de viscosité, et donne lieu à un principe de minimisation hiérarchique asymptotique.

Un exemple classique est l'introduction d'un terme de régularisation de type Tikhonov $\epsilon(\cdot)x(\cdot)$ dans la dynamique pour atteindre asymptotiquement un équilibre de norme minimale. Pour cela nous considérons la méthode de Newton régularisée (à la Tikhonov) suivante :

$$v(t) \in \partial\Phi(x(t)), \quad (5.1.4)$$

$$\lambda\dot{x}(t) + \dot{v}(t) + v(t) + \epsilon(t)x(t) = 0, \quad (5.1.5)$$

où $\epsilon : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, tends vers zéro de façon modérée lorsque $t \rightarrow +\infty$.

En utilisant la représentation de Minty de l'opérateur maximal monotone $\partial\Phi$, le système (5.1.4)-(5.1.5) se reformule comme suit :

$$x(t) = \text{prox}_{\mu\Phi}(y(t)), \quad (5.1.6)$$

$$\dot{y}(t) + \mu\nabla\Phi_\mu(y(t)) + \mu\epsilon(t)\text{prox}_{\mu\Phi}(y(t)) = 0, \quad (5.1.7)$$

où $\mu = \frac{1}{\lambda}$, $y : [0, +\infty[\rightarrow H$; $y(\cdot) = x(\cdot) + \mu v(\cdot)$, $\text{prox}_{\mu\Phi} = J_\mu^{\partial\Phi} = (I + \mu\partial\Phi)^{-1}$ est l'opérateur proximal associé à $\mu\Phi$ qui est la résolvante de l'opérateur $\partial\Phi$, et $\nabla\Phi_\mu = \frac{1}{\mu}(I - \text{prox}_{\mu\Phi})$ est l'approximation de Yosida d'indice μ . Insistons sur le fait que $\text{prox}_{\mu\Phi} : H \rightarrow H$ et $\nabla\Phi_\mu : H \rightarrow H$ sont partout définis et lipschitziens. Le système (5.1.4)-(5.1.5) relève du théorème de Cauchy Lipschitz.

Notre résultat principal concernant l'existence et l'unicité d'une solution globale forte du système (5.1.4)-(5.1.5) se formule comme suit :

Théorème 5.1.2 *Soit $\Phi : H \rightarrow \mathbb{R} \cup \{+\infty\}$, une fonction convexe, semi-continue inférieurement, et propre, et $\lambda > 0$ une constante positive. Soit $\epsilon : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ une fonction nonnegative et localement intégrable. Soit $(x_0, v_0) \in H \times H$ et $v_0 \in \partial\Phi(x_0)$. Alors,*

(i) *il existe une solution unique globale forte $(x(\cdot), v(\cdot)) : [0, +\infty[\rightarrow H \times H$ du problème de Cauchy (5.1.4)-(5.1.5) avec $x(0) = x_0, v(0) = v_0 \in \partial\Phi(x_0)$.*

(ii) *la solution $(x(\cdot), v(\cdot))$ du (5.1.4)-(5.1.5) peut être représentée par, pour tout $t \in [0, +\infty[$*

$$x(t) = \text{prox}_{\mu\Phi}(y(t)),$$

$$v(t) = \nabla\Phi_\mu(y(t)),$$

où $y(\cdot) : [0, +\infty[\rightarrow H$ est la solution unique globale forte du problème de Cauchy (5.1.7) avec $y(0) = x_0 + \mu v_0$.

Puis, nous étudions le comportement asymptotique des trajectoires du système (5.1.4)-(5.1.5). Pour cela, nous commençons à étudier le comportement asymptotique des trajectoires du système (5.1.6)-(5.1.7).

Le point clé de notre approche repose sur le fait que $\text{prox}_{\mu\Phi}$ est le gradient de la fonction convexe différentiable $\psi : H \rightarrow \mathbb{R}$, définie pour tout $y \in H$ par

$$\text{prox}_{\mu\Phi} = \nabla\psi, \quad \psi(y) = \mu(\Phi^*)_{\frac{1}{\mu}} \left(\frac{1}{\mu}y \right)$$

où Φ^* est la conjuguée de Fenchel de Φ . Une formulation équivalente de ψ est donnée par : pour tout $y \in H$, $\psi(y) = \frac{1}{2}\|y\|^2 - \mu\Phi_\mu(y)$, où l'on rappelle que Φ_μ est l'enveloppe de Moreau de la fonction Φ .

Ainsi, nous pouvons reformuler le système (5.1.6)-(5.1.7) sous la forme suivante :

$$x(t) = \text{prox}_{\mu\Phi}(y(t)) \quad (5.1.8)$$

$$\dot{y}(t) + \nabla\Theta(y(t)) + \epsilon(t) \nabla\Psi(y(t)) = 0, \quad (5.1.9)$$

avec, pour tout $y \in H$

$$\Theta(y) := \mu\Phi_\mu(y); \quad (5.1.10)$$

$$\Psi(y) = \mu^2 (\Phi^*)_{\frac{1}{\mu}} \left(\frac{1}{\mu} y \right). \quad (5.1.11)$$

Notons que Θ et Ψ sont convexes continument différentiables. On observe que le système (5.1.8)-(5.1.9) est un cas particulier du système dynamique multi-échelle qui est étudié en détail dans Attouch et Czarnecki (2010) :

$$(\text{MAG})_\epsilon \quad \dot{y}(t) + \partial\Theta(y(t)) + \epsilon(t)\partial\Psi(y(t)) \ni 0.$$

Dans notre travail, nous nous intéressons au cas où $t \mapsto \epsilon(t)$ vérifie la propriété "de décroissance lente vers zéro"

$$\int_0^{+\infty} \epsilon(t) dt = +\infty \quad (5.1.12)$$

Cette condition exprime que $\epsilon(\cdot)$ ne tend pas trop vite vers zéro, ce qui permet au terme $\epsilon(\cdot) x(\cdot)$ d'être efficace asymptotiquement.

Dans le théorème suivant (5.1.3), sous la condition (5.1.12) et une hypothèse de croissance modérée de $\epsilon(\cdot)$, nous montrons que chaque trajectoire (x, v) de (5.1.4)-(5.1.5) converge faiblement vers un minimiseur de Φ , qui minimise ψ sur l'ensemble des minimiseurs de Φ . Puis, nous montrons que cet élément n'est rien d'autre que l'élément de norme minimal de l'ensemble des solutions $\arg \min \Phi$

$$x(t) \rightharpoonup \text{proj}_{\arg \min \Phi} 0 \quad \text{lorsque } t \rightarrow +\infty.$$

On retrouve ainsi le principe classique de sélection de Tikhonov : sous l'hypothèse de décroissance modérée de $\epsilon(\cdot)$ vers zéro, nous obtenons que toutes les trajectoires convergent vers le même point, qui est l'équilibre de norme minimale. Ceci fait contraste avec les situations sans terme de contrôle, ou qui sont avec un contrôle rapide, auquel cas les limites dépendent des données initiales et sont difficile à identifier.

Théorème 5.1.3 *Soit $\Phi : H \rightarrow \mathbb{R}^+ \cup \{+\infty\}$ une fonction convexe, semi-continue inférieurement, et propre, telle que $C = \arg \min \Phi = \Phi^{-1}(0) \neq \emptyset$.*

Supposons que,

- (\mathcal{H}_1) $\forall p \in R(N_C), \quad \int_0^{+\infty} \mu\Phi^*\left(\frac{1}{\mu}\epsilon(t)p\right) - \sigma_C(\epsilon(t)p) + \frac{1}{2}\|\epsilon(t)p\|^2 dt < +\infty;$
- (\mathcal{H}_2) $\epsilon(\cdot)$ est décroissante de classe C^1 , lipschitzienne sur $[0, +\infty[$, et telle que $\lim_{t \rightarrow +\infty} \epsilon(t) = 0$, $\int_0^{+\infty} \epsilon(t) dt = +\infty$, et pour $k \geq 0$, $-k\epsilon^2 \leq \dot{\epsilon}$.

Alors, pour chaque trajectoire $(x(\cdot), v(\cdot)) : [0, +\infty[\rightarrow H \times H$ solution de (5.1.4)-(5.1.5), avec $y(t) = x(t) + \mu v(t)$, on a

(i) Convergence faible $w - \lim_{t \rightarrow +\infty} y(t) = \text{proj}_{\arg \min \Phi} 0$.

Supposons par ailleurs que $\Phi(0) < +\infty$. Alors

(ii) Convergence faible $w - \lim_{t \rightarrow +\infty} x(t) = w - \lim_{t \rightarrow +\infty} y(t) = \text{proj}_{\arg \min \Phi} 0;$

(iii) Convergence forte $s - \lim_{t \rightarrow +\infty} v(t) = 0$, et par conséquent $s - \lim_{t \rightarrow +\infty} x(t) - y(t) = 0$.

Puis, nous passons à l'étude de la convergence forte des trajectoires vers le minimiseur de norme minimale. Dans le théorème suivant (5.1.4), nous examinons le cas où $\Phi : H \rightarrow \mathbb{R}$ est une fonction convexe différentiable, dont le gradient est Lipschitzien (par conséquence $\text{prox}_{\mu\Phi}$ est fortement monotone). Nous obtenons alors la convergence forte des trajectoires :

Théorème 5.1.4 *Sous les hypothèses du théorème 5.1.3, supposons en outre que Φ est convexe, différentiable, et que son gradient est L -Lipschitzien avec $L > 0$. Alors pour $\mu L < 1$, $x(\cdot)$ et $y(\cdot)$ convergent fortement vers l'élément de norme minimale dans $C = \arg \min \Phi = \Phi^{-1}(0) \neq \emptyset$.*

Enfin, nous considérons les propriétés de sélection asymptotiques liées à d'autres types de méthode de viscosité.

$$v(t) \in \partial\Phi(x(t)) \tag{5.1.13}$$

$$\lambda \dot{x}(t) + \dot{v}(t) + v(t) + \varepsilon(t) \partial g(x(t)) \ni 0, \tag{5.1.14}$$

où g est une "fonction de viscosité", supposée convexe.

En utilisant la représentation de Minty de $\partial\Phi$, nous reformulons de façon équivalente le système précédent sous la forme suivante

$$x(t) = \text{prox}_{\mu\Phi}(y(t)) \tag{5.1.15}$$

$$\dot{y}(t) + \mu \nabla \Phi_{\mu}(y(t)) + \mu \varepsilon(t) \partial g(\text{prox}_{\mu\Phi}(y(t))) \ni 0. \tag{5.1.16}$$

En prenant $g(\cdot) = \frac{1}{2\mu} \|\cdot\|^2 + \Phi(\cdot)$, et $\Theta(y) = \mu \Phi_{\mu}(y)$, on récupère exactement l'approximation de Tikhonov, et (5.1.16) se reformule comme suit :

$$\dot{y}(t) + \nabla \Theta(y(t)) + \varepsilon(t) y(t) = 0 \tag{5.1.17}$$

et le système (5.1.13)-(5.1.14) devient

$$v(t) \in \partial\Phi(x(t)) \tag{5.1.18}$$

$$\lambda \dot{x}(t) + \dot{v}(t) + (1 + \mu \varepsilon(t))v(t) + \varepsilon(t) x(t) = 0. \tag{5.1.19}$$

D'après le théorème de Cauchy-Lipschitz, (5.1.17) admet une solution globale forte $y(\cdot)$. Alors $(x(\cdot), v(\cdot))$, avec $x(\cdot) = \text{prox}_{\mu\Phi}(y(\cdot))$ et $y(\cdot) = \nabla \Phi_{\mu}(y(\cdot))$, est une solution globale forte de (5.1.18)-(5.1.19).

Dans le théorème suivant nous montrons les propriétés de convergence du système (5.1.18)-(5.1.19).

Théorème 5.1.5 *Soit $\Phi : H \rightarrow \mathbb{R}^+ \cup \{+\infty\}$ une fonction convexe, semi-continue inférieurement, et propre, telle que $C = \arg \min \Phi = \Phi^{-1}(0) \neq \emptyset$. Supposons que $\lim_{t \rightarrow +\infty} \varepsilon(t) = 0$, et $\int_0^{+\infty} \varepsilon(t) dt = +\infty$. Alors, pour chaque trajectoire $(x(\cdot), v(\cdot)) : [0, +\infty[\rightarrow H \times H$ du système (5.1.18)-(5.1.19), avec $y(t) = x(t) + \mu v(t)$ la propriété de la convergence forte suivante est vérifiée :*

$$s - \lim_{t \rightarrow +\infty} x(t) = s - \lim_{t \rightarrow +\infty} y(t) = \text{proj}_{\arg \min \Phi} 0.$$

5.2 Article en anglais

An asymptotic viscosity selection result for the regularized Newton dynamic¹

abstract

Let $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a closed convex proper function on a real Hilbert space \mathcal{H} , and $\partial\Phi : \mathcal{H} \rightrightarrows \mathcal{H}$ its subdifferential. For any control function $\varepsilon : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ which tends to zero as t goes to $+\infty$, and λ a positive parameter, we study the asymptotic behavior of the trajectories of the regularized Newton dynamical system

$$\begin{aligned}v(t) &\in \partial\Phi(x(t)) \\ \lambda\dot{x}(t) + \dot{v}(t) + v(t) + \varepsilon(t)x(t) &= 0.\end{aligned}$$

Assuming that $\varepsilon(t)$ tends to zero moderately as t goes to $+\infty$, we show that the term $\varepsilon(\cdot)x(\cdot)$ asymptotically acts as a Tikhonov regularization, which forces the trajectories to converge to a particular equilibrium. Precisely, when $C = \operatorname{argmin}\Phi \neq \emptyset$, and $\varepsilon(\cdot)$ is a “slow” control, i.e., $\int_0^{+\infty} \varepsilon(t) dt = +\infty$, then each trajectory of the system converges weakly, as t goes to $+\infty$, to the element of minimal norm of the closed convex set C . When Φ is a convex differentiable function whose gradient is Lipschitz continuous, we show that the strong convergence property is satisfied. Then we examine the effect of other types of regularizing methods.

5.2.1 Introduction

Throughout this paper, \mathcal{H} is a real Hilbert space with scalar product $\langle \cdot, \cdot \rangle$, and $\|x\|^2 = \langle x, x \rangle$ for any $x \in \mathcal{H}$. Given $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ a closed convex proper function, we will analyze some asymptotic viscosity selection properties for the regularized Newton dynamic governed by Φ .

Let us first recall some basic facts about this dynamical system. Given λ a positive constant, the Regularized Newton dynamic ((RN) for short) attached to solving the minimization problem

$$(\mathcal{P}) \quad \min_{x \in \mathcal{H}} \Phi(x)$$

is written as follows

$$v(t) \in \partial\Phi(x(t)) \tag{5.2.1}$$

$$\lambda\dot{x}(t) + \dot{v}(t) + v(t) = 0, \tag{5.2.2}$$

where the subdifferential of Φ at $x \in \operatorname{dom}\Phi$ is classically defined by

$$\partial\Phi(x) = \{p \in \mathcal{H} : \Phi(y) \geq \Phi(x) + \langle p, y - x \rangle \ \forall y \in \mathcal{H}\}.$$

When Φ is a smooth function, (RN) is equivalent to

$$\lambda\dot{x}(t) + \nabla^2\Phi(x(t)) + \nabla\Phi(x(t)) = 0$$

1. B. Abbas, An asymptotic viscosity selection result for the regularized Newton dynamic, arXiv :1504.07793v1 [math.OA] 29 Apr 2015.

where λ acts as a Levenberg-Marquard regularization parameter of the continuous Newton equation, whence the terminology. This dynamical system has been first introduced in [9], [8]. Its extension to the case of two potentials gives rise to a new class of forward-backward algorithms, see [1], [2], [7]. In [9], for a general closed convex and proper function Φ , it is shown that the Cauchy problem for the (x, v) system (5.2.1)-(5.2.2) admits a unique strong global solution. In addition, under the sole assumption that $C = \operatorname{argmin}\Phi \neq \emptyset$, for any orbit of (5.2.1)-(5.2.2), $x(t)$ converges weakly to an element of C , as t goes to $+\infty$.

In many applications, a particular stationary solution is more interesting than others due to physical, economic or design considerations. When we have the global convergence of trajectories, one could let the trajectory reach a particular target equilibrium by appropriately adjusting the initial conditions. Nevertheless, in many practical situations it is not possible to have an accurate control of the initial state. An alternative approach consists in introducing a term into the system which forces convergence to the desired stationary solution, independently of the initial state. Such a term should vanish at infinity in order to recover, at least asymptotically, an equilibrium point of (5.2.1)-(5.2.2).

The above discussion motivates the introduction of the following abstract evolution system :

$$v(t) \in \partial\Phi(x(t)) \tag{5.2.3a}$$

$$\lambda \dot{x}(t) + \dot{v}(t) + v(t) + \varepsilon(t)x(t) = 0, \tag{5.2.3b}$$

where $\varepsilon : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is an open-loop control function, that tends to zero as t goes to $+\infty$.

Let us briefly describe our approach. Following a similar device as in [9], setting $\mu = \frac{1}{\lambda}$, and introducing the new unknown function $y(\cdot) = x(\cdot) + \mu v(\cdot)$, we can equivalently rewrite (5.2.3a)-(5.2.3b) as

$$\begin{aligned} x(t) &= \operatorname{prox}_{\mu\Phi}(y(t)), \\ \dot{y}(t) + \mu \nabla\Phi_\mu(y(t)) + \mu\varepsilon(t) \operatorname{prox}_{\mu\Phi}(y(t)) &= 0 \end{aligned}$$

where $\operatorname{prox}_{\mu\Phi}$ is the proximal mapping associated to $\mu\Phi$. Recall that $\operatorname{prox}_{\mu\Phi} = (I + \mu\partial\Phi)^{-1}$ is the resolvent of index $\mu > 0$ of the maximal monotone operator $\partial\Phi$, and $\nabla\Phi_\mu = \frac{1}{\mu}(I - (I + \mu\partial\Phi)^{-1})$ is its Yosida approximation of index $\mu > 0$. As a key point of our analysis, we notice that $\operatorname{prox}_{\mu\Phi}$ is a gradient vector field, namely $\operatorname{prox}_{\mu\Phi} = \nabla\psi$, with

$$\psi(y) = \mu(\Phi^*)_{\frac{1}{\mu}}\left(\frac{1}{\mu}y\right), \tag{5.2.4}$$

where Φ^* is the Fenchel conjugate of Φ . Doing so, we can reformulate our dynamic in the form

$$x(t) = (I + \mu\partial\Phi)^{-1}(y(t)), \tag{5.2.5a}$$

$$\dot{y}(t) + \mu \nabla\Phi_\mu(y(t)) + \varepsilon(t) \nabla\mu\psi(y(t)) = 0. \tag{5.2.5b}$$

Equation (5.2.5b) is a particular case of the multiscale dynamic

$$\dot{y}(t) + \partial\Theta(y(t)) + \varepsilon(t) \partial\Psi(y(t)) \ni 0 \tag{5.2.6}$$

where Θ and Ψ are two convex potential functions. Following [4], Ψ will be referred to as the "viscosity function". A detailed study of the asymptotic behavior of the orbits of (5.2.6) can

be found in [6], [14], [15], [16], [18]. Following [6] and [14], we focus our attention on the case where the parametrization $t \mapsto \epsilon(t)$ satisfies the following "slow" decay property

$$\int_0^{+\infty} \epsilon(t) dt = +\infty.$$

This condition expresses that $\epsilon(\cdot)$ does not tend to zero too rapidly, which allows the term $\epsilon(\cdot)x(\cdot)$ to be effective asymptotically. In that case, we will show an asymptotic selection property. Precisely, in Theorem 5.2.11, under some additional moderate growth property on $\epsilon(\cdot)$, we will show that, for any trajectory (x, v) of (5.2.3a)-(5.2.3b), $x(\cdot)$ converges weakly to the minimizer of Φ which also minimizes ψ over all minima of Φ . Then we show that this element is nothing but the element of minimal norm of the solution set $\text{argmin}\Phi$, i.e.,

$$x(t) \rightharpoonup \text{proj}_{\text{argmin}\Phi} 0 \quad \text{as } t \rightarrow +\infty.$$

Thus we recover the classical Tikhonov viscosity selection principle, which consists in selecting the solution of minimal norm.

This result can be viewed as an asymptotic selection property : by using such a slow control ϵ , one can force all the trajectories to converge to the same equilibrium, which here is the equilibrium of minimal norm. This makes a sharp contrast with the non controlled situation, or fast control, where the limits of the trajectories depend on the initial data, and are in general difficult to identify.

The paper is organized as follows : we first show the existence and uniqueness of a strong global solution to the Cauchy problem (5.2.3a)-(5.2.3b). Then, we study the asymptotic convergence as t goes to $+\infty$ of the trajectories of (5.2.3a)-(5.2.3b). In our main result, Theorem 5.2.11, under the key assumption that $\epsilon(\cdot)$ is a "slow control", i.e., $\int_0^{+\infty} \epsilon(t) dt = +\infty$, and has moderate growth, we show the weak convergence of the trajectories toward the optimal solution of problem (\mathcal{P}) of minimum norm. When Φ is a convex differentiable function whose gradient is Lipschitz continuous, we show that the convergence holds for the strong topology. Finally, we examine some variants of this principle of hierarchical minimization.

5.2.2 Existence and Uniqueness of Global Solutions

We consider the Cauchy problem for the differential inclusion system (5.2.3a)-(5.2.3b)

$$v(t) \in \partial\Phi(x(t)) \tag{5.2.7a}$$

$$\lambda \dot{x}(t) + \dot{v}(t) + v(t) + \epsilon(t)x(t) = 0 \tag{5.2.7b}$$

$$x(0) = x_0, v(0) = v_0 \tag{5.2.7c}$$

First, we are going to define a notion of strong solution to the above system. Then, we shall reformulate this system with the help of the Minty representation of $\partial\Phi$. Finally, we shall prove the existence and uniqueness of a strong solution to system (5.2.7a)–(5.2.7c), by applying the Cauchy–Lipschitz theorem to this equivalent formulation.

Definition of strong solutions

We say that the pair $(x(\cdot), v(\cdot))$ is a strong global solution of (5.2.7a)–(5.2.7c) iff the following properties are satisfied :

- i) $x(\cdot), v(\cdot) : [0, +\infty[\longrightarrow \mathcal{H}$ are absolutely continuous on each interval $[0, b]$, $0 < b < +\infty$;
- ii) $v(t) \in \partial\Phi(x(t))$ for all $t \in [0, +\infty[$;
- iii) $\lambda\dot{x}(t) + \dot{v}(t) + v(t) + \varepsilon(t)x(t) = 0$ for almost all $t \in [0, +\infty[$;
- iv) $x(0) = x_0, v(0) = v_0$.

Equivalent formulation as a classical differential equation

In order to solve system (5.2.7a)–(5.2.7c) we use Minty's device. Set

$$\mu = \frac{1}{\lambda}.$$

Let us rewrite inclusion (5.2.7a) by using the following equivalences : for any $t \in [0, +\infty[$

$$v(t) \in \partial\Phi(x(t)) \Leftrightarrow \quad (5.2.8)$$

$$x(t) + \mu v(t) \in x(t) + \mu\partial\Phi(x(t)) \quad (5.2.9)$$

$$x(t) = (I + \mu\partial\Phi)^{-1}(x(t) + \mu v(t)). \quad (5.2.10)$$

Let us introduce the new unknown function $y : [0, +\infty[\longrightarrow \mathcal{H}$ which is defined for $t \in [0, +\infty[$ by

$$y(t) := x(t) + \mu v(t), \quad (5.2.11)$$

and rewrite the system (5.2.7a)–(5.2.7c) with the help of (x, y) . From (5.2.10) and (5.2.11)

$$\begin{aligned} x(t) &= (I + \mu\partial\Phi)^{-1}(y(t)), \\ v(t) &= \frac{1}{\mu}(y(t) - (I + \mu\partial\Phi)^{-1}(y(t))). \end{aligned}$$

Equivalently,

$$x(t) = \text{prox}_{\mu\Phi}(y(t)); \quad (5.2.12)$$

$$v(t) = \nabla\Phi_{\mu}(y(t)), \quad (5.2.13)$$

where $\text{prox}_{\mu\Phi}$ is the proximal mapping associated to $\mu\Phi$. Recall that $\text{prox}_{\mu\Phi} = (I + \mu\partial\Phi)^{-1}$ is the resolvent of index $\mu > 0$ of the maximal monotone operator $\partial\Phi$, and $\nabla\Phi_{\mu}$ is its Yosida approximation of index $\mu > 0$.

Let us show how (5.2.7b) can be reformulated as a classical differential equation with respect to $y(\cdot)$. First, let us rewrite (5.2.7b) as

$$\dot{x}(t) + \mu\dot{v}(t) + \mu v(t) + \mu\varepsilon(t)x(t) = 0. \quad (5.2.14)$$

Differentiating (5.2.11), and using (5.2.14) we obtain

$$\dot{y}(t) = \dot{x}(t) + \mu\dot{v}(t) \quad (5.2.15)$$

$$= -\mu v(t) - \mu\varepsilon(t)x(t). \quad (5.2.16)$$

From (5.2.12), (5.2.13), and (5.2.16) we deduce that

$$\dot{y}(t) + \mu\nabla\Phi_{\mu}(y(t)) + \mu\varepsilon(t)\text{prox}_{\mu\Phi}(y(t)) = 0.$$

Finally, the (x, y) system can be written as

$$x(t) = \text{prox}_{\mu\Phi}(y(t)) \quad (5.2.17a)$$

$$\dot{y}(t) + \mu\nabla\Phi_\mu(y(t)) + \mu\epsilon(t)\text{prox}_{\mu\Phi}(y(t)) = 0. \quad (5.2.17b)$$

Conversely, if $y(\cdot)$ is a solution of (5.2.17b), then $(x(\cdot), v(\cdot))$ with $x(t) = \text{prox}_{\mu\Phi}(y(t))$, $v(t) = \nabla\Phi_\mu(y(t))$ is a solution of (5.2.7a)-(5.2.7c). Let us stress the fact that the operators $\text{prox}_{\mu\Phi} : \mathcal{H} \rightarrow \mathcal{H}$, $\nabla\Phi_\mu : \mathcal{H} \rightarrow \mathcal{H}$ are everywhere defined and Lipschitz continuous, which makes this system relevant to the Cauchy–Lipschitz theorem.

Global existence and uniqueness results

Let us state our main result of existence and uniqueness for the system (5.2.7a)–(5.2.7c).

Theorem 5.2.1 *Suppose that $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex lower semicontinuous proper function, and $\lambda > 0$ is a positive constant. Let $\epsilon : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a nonnegative locally integrable function, and $(x_0, v_0) \in \mathcal{H} \times \mathcal{H}$ be such that $v_0 \in \partial\Phi(x_0)$. Then the following properties hold :*

- i) there exists a unique strong global solution $(x(\cdot), v(\cdot)) : [0, +\infty[\rightarrow \mathcal{H} \times \mathcal{H}$ of the Cauchy problem (5.2.7a)–(5.2.7c) ;*
- ii) the solution pair $(x(\cdot), v(\cdot))$ of (5.2.7a)–(5.2.7c) can be represented as follows : for any $t \in [0, +\infty[$,*

$$x(t) = \text{prox}_{\mu\Phi}(y(t)) ; \quad (5.2.18)$$

$$v(t) = \nabla\Phi_\mu(y(t)) , \quad (5.2.19)$$

where $y(\cdot) : [0, +\infty[\rightarrow \mathcal{H}$ is the unique strong global solution of the Cauchy problem

$$\dot{y}(t) + \mu\nabla\Phi_\mu(y(t)) + \mu\epsilon(t)\text{prox}_{\mu\Phi}(y(t)) = 0, \quad (5.2.20a)$$

$$y(0) = x_0 + \mu v_0. \quad (5.2.20b)$$

Proof. Let us first prove the existence and uniqueness of a strong global solution of the Cauchy problem (5.2.20a)–(5.2.20b). The Cauchy problem (5.2.20a)–(5.2.20b) can be equivalently written in abstract form, as the following non-autonomous differential system

$$\begin{cases} \dot{y}(t) = F(t, y(t)) ; \\ y(0) = x_0 + \mu v_0, \end{cases} \quad (5.2.21)$$

with

$$F(t, y) = G(t, y) + K(t, y), \quad (5.2.22)$$

$$G(t, y) = -\mu\nabla\Phi_\mu(y), \quad (5.2.23)$$

$$K(t, y) = -\mu\epsilon(t)\text{prox}_{\mu\Phi}(y). \quad (5.2.24)$$

In order to apply the Cauchy–Lipschitz theorem to (5.2.21), let us first examine the Lipschitz continuity properties of $F(t, \cdot)$.

(a) Take arbitrary $y_i \in \mathcal{H}$, $i = 1, 2$. The Yosida approximation $\nabla\Phi_\mu$ is $\frac{1}{\mu}$ -Lipschitz continuous (see [12]), and hence, for any $t \geq 0$, $G(t, \cdot) : \mathcal{H} \rightarrow \mathcal{H}$ is nonexpansive, i.e.,

$$\|G(t, y_2) - G(t, y_1)\| \leq \|y_2 - y_1\|. \quad (5.2.25)$$

By the nonexpansive property of the resolvent operators we have

$$\|K(t, y_2) - K(t, y_1)\| \leq \mu\epsilon(t) \|y_2 - y_1\|. \quad (5.2.26)$$

Hence,

$$\|F(t, y_2) - F(t, y_1)\| \leq (1 + \mu\epsilon(t)) \|y_2 - y_1\|. \quad (5.2.27)$$

Since $\epsilon(\cdot)$ is nonnegative and locally integrable, (5.2.27) shows that the Lipschitz constant $L_F(t) = (1 + \mu\epsilon(t))$ of $F(t, \cdot)$ satisfies

$$L_F(\cdot) \in L^1([0, b]) \text{ for any } 0 < b < +\infty. \quad (5.2.28)$$

(b) Let us show that

$$\forall y \in \mathcal{H}, \forall b > 0, F(\cdot, y) \in L^1([0, b]; \mathcal{H}). \quad (5.2.29)$$

Returning to the definition (5.2.22) of F , we deduce that

$$\|F(t, y)\| \leq \mu \|\nabla\Phi_\mu(y)\| + \mu\epsilon(t) \|\text{prox}_{\mu\Phi}(y)\|.$$

By assumption, $\epsilon : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a nonnegative locally integrable function, which gives (5.2.29). From (5.2.27) and (5.2.29), by Cauchy-Lipschitz theorem (see [17], [21] for the nonautonomous version used here), we deduce the existence and uniqueness of a strong global solution of the Cauchy problem (5.2.21), and hence of (5.2.20a)-(5.2.20b).

(2) Let us return to the initial problem (5.2.7a)-(5.2.7c). Given $y(\cdot) : [0, +\infty[\rightarrow \mathcal{H}$ which is the unique strong solution of Cauchy problem (5.2.20a)-(5.2.20b), let us define $x(\cdot), v(\cdot) : [0, +\infty[\rightarrow \mathcal{H}$ by

$$x(t) = \text{prox}_{\mu\Phi}(y(t)), \quad v(t) = \nabla\Phi_\mu(y(t)). \quad (5.2.30)$$

(a) Let us show that $x(\cdot), v(\cdot)$ are absolutely continuous on each bounded interval, and satisfy (5.2.7a)-(5.2.7c). Let us give arbitrary $y_1 \in \mathcal{H}, y_2 \in \mathcal{H}$. By the nonexpansive property of the resolvents, we have

$$\|\text{prox}_{\mu\Phi}(y_2) - \text{prox}_{\mu\Phi}(y_1)\| \leq \|y_2 - y_1\|. \quad (5.2.31)$$

Assuming that $s, t \in [0, b]$, by taking $y_1 = y(s), y_2 = y(t)$ in (5.2.31), and owing to the definition of the absolute continuity property, we deduce that $x(t) = \text{prox}_{\mu\Phi}(y(t))$ is absolutely continuous on $[0, b]$ for any $b > 0$. As a linear combination of two absolutely continuous functions, the same property holds true for $v(t) = \lambda(y(t) - x(t))$.

Moreover, for any $t \in [0, +\infty]$

$$v(t) \in \partial\Phi(x(t)), \quad y(t) = x(t) + \mu v(t).$$

Differentiation of the above equation shows that, for almost every $t > 0$,

$$\dot{x}(t) + \mu\dot{v}(t) = \dot{y}(t).$$

On the other hand, owing to $v(t) = \nabla\Phi_\mu(y(t)), x(t) = \text{prox}_{\mu\Phi}(y(t))$, (5.2.20a) can be equivalently written as

$$\dot{y}(t) + \mu v(t) + \mu\epsilon(t)x(t) = 0.$$

Combining the two above equations, we obtain

$$\dot{x}(t) + \mu\dot{v}(t) + \mu v(t) + \mu\epsilon(t)x(t) = 0.$$

From $\mu = \frac{1}{\lambda}$, we conclude that $(x(\cdot), v(\cdot))$ is a solution of system (5.2.7a)-(5.2.7b). Regarding the initial condition, we observe that

$$y(0) = x_0 + \mu v_0 \quad (5.2.32)$$

$$= x(0) + \mu v(0), \quad (5.2.33)$$

with $v_0 \in \partial\Phi(x_0)$ and $v(0) \in \partial\Phi(x(0))$. Hence

$$x(0) = x_0 = (I + \mu\partial\Phi)^{-1}(x_0 + \mu v_0).$$

After simplification, we obtain $v(0) = v_0$.

(b) Let us now prove the uniqueness. Suppose that

$$x(\cdot), v(\cdot) : [0, +\infty[\longrightarrow \mathcal{H}$$

is a solution pair of (5.2.7a)-(5.2.7c). Defining $\mu = \frac{1}{\lambda}$ and

$$y(t) = x(t) + \mu v(t) \quad (5.2.34)$$

we conclude that $y(\cdot)$ is absolutely continuous, $y_0 = x_0 + \mu v_0$, and for any $t \in [0, +\infty[$

$$x(t) = (I + \mu\partial\Phi)^{-1}(y(t)), \quad v(t) = \nabla\Phi_\mu(y(t)). \quad (5.2.35)$$

Since the functions involved in the definition (5.2.34) of $y(\cdot)$, namely $x(\cdot)$ and $v(\cdot)$, are differentiable for almost all $t \in [0, +\infty[$, we have for almost all $t \in [0, +\infty[$

$$\begin{aligned} \dot{y}(t) &= \dot{x}(t) + \mu\dot{v}(t) \\ &= -\mu(\dot{v}(t) + v(t) + \epsilon(t)x(t)) + \mu\dot{v}(t). \end{aligned}$$

Since $v(t) = \nabla\Phi_\mu(y(t))$, we finally obtain

$$\dot{y}(t) + \mu\nabla\Phi_\mu(y(t)) + \mu\epsilon(t)\text{prox}_{\mu\Phi}(y(t)) = 0.$$

Moreover

$$y_0 = x_0 + \mu v_0.$$

Arguing as before, by the Cauchy–Lipschitz theorem, the solution $y(\cdot)$ of the above system is uniquely determined, and locally absolutely continuous. Thus, by (5.2.35), $x(\cdot)$ and $v(\cdot)$ are uniquely determined. \square

5.2.3 Asymptotic analysis and convergence properties

In this section, we study the asymptotic behavior, as $t \rightarrow +\infty$, of the trajectories of system (5.2.7a)-(5.2.7b). Let us recall our standing assumption, namely the parametrization $\epsilon(\cdot)$ is supposed to be nonnegative, and locally integrable. In view of the asymptotic analysis, we also suppose that $\epsilon(t) \rightarrow 0$ as $t \rightarrow \infty$, and satisfies the "slow" decay property

$$\int_0^{+\infty} \epsilon(t) dt = +\infty.$$

By Theorem 5.2.1, for any given Cauchy data $v_0 \in \partial\Phi(x_0)$, the above properties guarantee the existence and uniqueness of a global solution of system (5.2.7a)-(5.2.7b)-(5.2.7c). From now on in this section, $(x(\cdot), v(\cdot)) : [0, +\infty[\rightarrow \mathcal{H} \times \mathcal{H}$ is the solution of (5.2.7a)-(5.2.7b)-(5.2.7c). We first study the asymptotic behavior, as $t \rightarrow +\infty$, of the trajectories of the associated system (5.2.20a)

$$\dot{y}(t) + \mu \nabla \Phi_\mu(y(t)) + \mu \epsilon(t) \operatorname{prox}_{\mu\Phi}(y(t)) = 0$$

whose existence is guaranteed by Theorem 5.2.1. The central point of our analysis is to reformulate this system as a multi-scale gradient system, which will allow us to use the known results concerning the asymptotic behavior, and the hierarchical selection property for such systems.

Preliminary results

Let us state some definitions and classical properties that will be useful (see [5], [11], [12], [20] for an extended presentation of these notions) :

Definition 5.2.2 *Let f and g be functions from \mathcal{H} to $\mathbb{R} \cup \{+\infty\}$. The infimal convolution (or epi-sum) of f and g is the function $f \square g : \mathcal{H} \rightarrow [-\infty, +\infty]$ which is defined by*

$$f \square g(x) = \inf_{\xi \in \mathcal{H}} (f(\xi) + g(x - \xi)).$$

Definition 5.2.3 *Let $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$, $\gamma \in \mathbb{R}_{++}$. The Moreau envelope of f of parameter γ is defined by*

$$f_\gamma = f \square \left(\frac{1}{2\gamma} \|\cdot\|^2 \right).$$

Definition 5.2.4 *Let $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex lower semicontinuous proper function, and let $x \in \mathcal{H}$. Then $\operatorname{prox}_f x$ is the unique point in \mathcal{H} that satisfies*

$$f_1(x) = \min_{\xi \in \mathcal{H}} \left(f(\xi) + \frac{1}{2} \|x - \xi\|^2 \right) = f(\operatorname{prox}_f x) + \frac{1}{2} \|x - \operatorname{prox}_f x\|^2.$$

The operator $\operatorname{prox}_f : \mathcal{H} \rightarrow \mathcal{H}$ is called the proximity operator, or proximal mapping of f .

Definition 5.2.5 *Let $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$. The conjugate (or Legendre-Fenchel transform, or Fenchel conjugate) of f is*

$$f^* : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\} : u \mapsto \sup_{x \in \mathcal{H}} (\langle x, u \rangle - f(x)).$$

Remark 5.2.6 *Let $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper then*

$$f^*(0) = - \inf_{\mathcal{H}} f.$$

f is lower semi continuous and convex if and only if

$$f = f^{**}.$$

Remark 5.2.7 Let f and g be functions from \mathcal{H} to $\mathbb{R} \cup \{+\infty\}$. Then

$$(f \square g)^* = f^* + g^*.$$

Conversely, if one of the functions (f or g) is continuous at a point of the domain of the other, then

$$(f + g)^* = f^* \square g^*.$$

Remark 5.2.8 a) Let $\varphi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper, and $\gamma \in \mathbb{R}_{++}$. Set $f = \varphi + \frac{1}{2\gamma} \|\cdot\|^2$. Then $\forall u \in \mathcal{H}$

$$f^*(u) = \frac{\gamma}{2} \|u\|^2 - \varphi_\gamma(\gamma u).$$

b) Let C be a nonempty subset of \mathcal{H} , and let $f = \delta_C + \|\cdot\|^2/2$, where δ_C is the indicator function of the set C ($\delta_C(x) = 0$ for $x \in C$, $+\infty$ outwards). Then

$$f^* = (\|\cdot\|^2 - d_C^2)/2,$$

where d_C is the distance function to the set C . For the proof, set $\varphi = \delta_C$ and $\gamma = 1$ in Remark 5.2.8.

In the next lemma, we show that the proximal mapping can be written as the gradient of a convex differentiable function. This result will play a crucial role in our analysis.

Lemma 5.2.9 Let $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper convex lower semicontinuous function, and let $\mu > 0$. Then, the proximal mapping $\text{prox}_{\mu\Phi} : \mathcal{H} \rightarrow \mathcal{H}$ can be written as the gradient

$$\text{prox}_{\mu\Phi} = \nabla \psi$$

of the convex continuously differentiable function $\psi : \mathcal{H} \rightarrow \mathbb{R}$ which is defined, for any $y \in \mathcal{H}$, by

$$\psi(y) = \mu (\Phi^*)_{\frac{1}{\mu}} \left(\frac{1}{\mu} y \right)$$

where Φ^* is the Fenchel conjugate of Φ .

Proof. For any $y \in \mathcal{H}$, set

$$x = \text{prox}_{\mu\Phi}(y).$$

By definition of the proximal mapping, we have the following equivalent formulations

$$\begin{aligned} x &= (I + \mu\partial\Phi)^{-1}(y) \\ y &\in (I + \mu\partial\Phi)(x) \\ y - x &\in \mu\partial\Phi(x) \\ \frac{1}{\mu}(y - x) &\in \partial\Phi(x). \end{aligned}$$

From $(\partial\Phi)^{-1} = \partial\Phi^*$ and the above equality, we successively obtain

$$\begin{aligned}
x &\in \partial\Phi^* \left(\frac{1}{\mu} (y - x) \right); \\
\frac{1}{\mu} y &\in \frac{1}{\mu} (y - x) + \frac{1}{\mu} \partial\Phi^* \left(\frac{1}{\mu} (y - x) \right) = \left(I + \frac{1}{\mu} \partial\Phi^* \right) \left(\frac{1}{\mu} (y - x) \right); \\
\frac{1}{\mu} (y - x) &= \left(I + \frac{1}{\mu} \partial\Phi^* \right)^{-1} \left(\frac{1}{\mu} y \right); \\
x &= \mu \left(\frac{1}{\mu} y - \left(I + \frac{1}{\mu} \partial\Phi^* \right)^{-1} \left(\frac{1}{\mu} y \right) \right); \\
x &= \mu \left(I - \text{prox}_{\frac{1}{\mu}\Phi^*} \right) \left(\frac{1}{\mu} y \right).
\end{aligned}$$

By the definition of the Yosida approximation of index $\frac{1}{\mu} > 0$ of the maximal monotone operator $\partial\Phi^*$

$$x = \nabla \left[(\Phi^*)_{\frac{1}{\mu}} \right] \left(\frac{1}{\mu} y \right).$$

By the classical derivation chain rule, we deduce that the proximal mapping $\text{prox}_{\mu\Phi} y : \mathcal{H} \rightarrow \mathcal{H}$ is the gradient of the convex continuously differentiable function $\psi : \mathcal{H} \rightarrow \mathbb{R}$ which is defined, for any $y \in \mathcal{H}$, by

$$\psi(y) = \mu (\Phi^*)_{\frac{1}{\mu}} \left(\frac{1}{\mu} y \right).$$

□

Let us further analyze the function ψ , and give equivalent formulations which come with different proofs of the above lemma. By Definition 5.2.3 of the Moreau envelope of Φ of parameter μ

$$\Phi_\mu(y) = \inf_{x \in \mathcal{H}} \left\{ \Phi(x) + \frac{1}{2\mu} \|y - x\|^2 \right\}, \quad (5.2.36)$$

and since the Yosida approximation of the subdifferential of Φ is the Fréchet derivative of Moreau envelope,

$$\begin{aligned}
\nabla\Phi_\mu(y) &= \frac{1}{\mu} (y - J_\mu^{\partial\Phi} y) \\
\mu\nabla\Phi_\mu(y) &= y - J_\mu^{\partial\Phi} y.
\end{aligned}$$

By definition, we have $\text{prox}_{\mu\Phi}(y) = J_\mu^{\partial\Phi} y = (I + \mu\partial\Phi)^{-1}(y)$. Hence

$$\mu\nabla\Phi_\mu(y) = y - \text{prox}_{\mu\Phi}(y).$$

$$\begin{aligned}
\text{prox}_{\mu\Phi}(y) &= y - \mu\nabla\Phi_\mu(y) \\
&= \nabla \left(\frac{1}{2} \|\cdot\|^2 - \mu\Phi_\mu \right) (y).
\end{aligned}$$

Let us make the link with the previous formulation of the prox as a gradient, and show that, for any $y \in \mathcal{H}$

$$\frac{1}{2} \|y\|^2 - \mu \Phi_\mu(y) = \mu \Phi_{\frac{1}{\mu}}^* \left(\frac{1}{\mu} y \right);$$

By (5.2.36), we have

$$\begin{aligned} \frac{1}{2} \|y\|^2 - \mu \Phi_\mu(y) &= \frac{1}{2} \|y\|^2 - \mu \inf_{x \in \mathcal{H}} \left\{ \Phi(x) + \frac{1}{2\mu} \|y - x\|^2 \right\} \\ &= \frac{1}{2} \|y\|^2 - \mu \inf_{x \in \mathcal{H}} \left\{ \Phi(x) + \frac{1}{2\mu} \|y\|^2 - \frac{1}{\mu} \langle y, x \rangle + \frac{1}{2\mu} \|x\|^2 \right\} \\ &= \sup_{x \in \mathcal{H}} \left\{ \langle y, x \rangle - \mu \left(\Phi(x) + \frac{1}{2\mu} \|x\|^2 \right) \right\} \\ &= \mu \sup_{x \in \mathcal{H}} \left\{ \left\langle \frac{1}{\mu} y, x \right\rangle - \left(\Phi(x) + \frac{1}{2\mu} \|x\|^2 \right) \right\}. \end{aligned}$$

By using Remark 5.2.7 concerning the conjugate of a sum, we obtain

$$\begin{aligned} \frac{1}{2} \|y\|^2 - \mu \Phi_\mu(y) &= \mu \left(\Phi + \frac{1}{2\mu} \|\cdot\|^2 \right)^* \left(\frac{1}{\mu} y \right) \\ &= \mu \left(\Phi^* \square \frac{\mu}{2} \|\cdot\|^2 \right) \left(\frac{1}{\mu} y \right) \\ &= \mu (\Phi^*)_{\frac{1}{\mu}} \left(\frac{1}{\mu} y \right). \end{aligned}$$

So we obtain the same function ψ as given in Lemma 5.2.9. Note that by Remark 5.2.8, the equivalent (dual) formulation of ψ given by $\psi(y) = \frac{1}{2} \|y\|^2 - \mu \Phi_\mu(y)$, which is written as a d.c. function, is actually a convex function.

Asymptotic hierarchical minimization

Let us study the asymptotic behavior of the trajectories of system (5.2.7a)-(5.2.7b). We consider the equivalent system (5.2.20a), which, by Lemma 5.2.9, can be formulated as follows :

$$x(t) = \text{prox}_{\mu\Phi}(y(t)) \tag{5.2.37}$$

$$\dot{y}(t) + \nabla\Theta(y(t)) + \epsilon(t) \nabla\Psi(y(t)) = 0, \tag{5.2.38}$$

where, for any $y \in \mathcal{H}$

$$\Theta(y) := \mu \Phi_\mu(y); \tag{5.2.39}$$

$$\Psi(y) = \mu^2 (\Phi^*)_{\frac{1}{\mu}} \left(\frac{1}{\mu} y \right). \tag{5.2.40}$$

Note that Θ and Ψ are two convex continuously differentiable functions. We are within the framework of the multiscale gradient system $(\text{MAG})_\epsilon$, with a positive control $t \mapsto \epsilon(t)$ that converges to 0 as $t \rightarrow \infty$,

$$(\text{MAG})_\epsilon \quad \dot{y}(t) + \partial\Theta(y(t)) + \epsilon(t) \partial\Psi(y(t)) \ni 0,$$

that has been considered by Attouch-Czarnecki in [6]. Let us recall this general abstract result, that we formulate with notations adapted to our setting. Since Θ enters $(\text{MAG})_\varepsilon$ only by its subdifferential, it is not a restrictive assumption to assume this potential to be nonnegative, with its infimal value equal to zero (subtracting the infimal value does not affect the subdifferential).

Theorem 5.2.10 (*Attouch-Czarnecki, [6]*) *Let*

- $\Theta : \mathcal{H} \rightarrow \mathbb{R}^+ \cup \{+\infty\}$ *be a closed convex proper function, such that* $C = \operatorname{argmin}\Theta = \Theta^{-1}(0) \neq \emptyset$.
- $\Psi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ *be a closed convex proper function, such that* $S = \operatorname{argmin}\{\Psi|_{\operatorname{argmin}\Theta}\} \neq \emptyset$.

Let us assume that,

- $(\mathcal{H}_1)_\varepsilon \quad \forall p \in R(N_C), \int_0^{+\infty} \Theta^*(\varepsilon(t)p) - \sigma_C(\varepsilon(t)p) dt < +\infty$.
- $(\mathcal{H}_2)_\varepsilon \quad \varepsilon(\cdot)$ *is a non increasing function of class* C^1 , *such that* $\lim_{t \rightarrow +\infty} \varepsilon(t) = 0$, $\int_0^{+\infty} \varepsilon(t) dt = +\infty$, *and for some* $k \geq 0$, $-k\varepsilon^2 \leq \dot{\varepsilon}$.

Let $y(\cdot)$ *be a strong solution of* $(\text{MAG})_\varepsilon$. *Then :*

- (i) *weak convergence* $\quad \exists y_\infty \in S = \operatorname{argmin}\{\Psi|_{\operatorname{argmin}\Theta}\}, \quad w - \lim_{t \rightarrow +\infty} y(t) = y_\infty;$
- (ii) *minimizing properties* $\quad \lim_{t \rightarrow +\infty} \Theta(y(t)) = 0;$
 $\quad \lim_{t \rightarrow +\infty} \Psi(y(t)) = \min \Psi|_{\operatorname{argmin}\Theta};$
- (iii) $\quad \forall z \in S \quad \lim_{t \rightarrow +\infty} \|y(t) - z\| \text{ exists ;}$
- (iv) *estimations* $\quad \lim_{t \rightarrow +\infty} \frac{1}{\varepsilon(t)} \Theta(y(t)) = 0;$
 $\quad \int_0^{+\infty} \Theta(y(t)) dt < +\infty;$
 $\quad \limsup_{\tau \rightarrow +\infty} \int_0^\tau \varepsilon(t) \left(\Psi(y(t)) - \min \Psi|_{\operatorname{argmin}\Theta} \right) dt < +\infty.$

By specializing this result to our setting, we will obtain the weak convergence of $y(\cdot)$ to a particular minimizer of Φ , which is the solution of a hierarchical minimization property. The convergence of $x(\cdot)$ is less immediate, and will follow from an energetical argument.

Analysis of the condition $(\mathcal{H}_1)_\varepsilon$:

The condition

$$(\mathcal{H}_1)_\varepsilon \quad \forall p \in R(N_C), \quad \int_0^{+\infty} \Theta^*(\varepsilon(t)p) - \sigma_C(\varepsilon(t)p) dt < +\infty,$$

plays a crucial role in our asymptotic analysis. Before proceeding in the discussion of this hypothesis, we recall some classical notions from convex analysis, that will be useful.

— σ_C is the support function of C ,

$$\sigma_C(x^*) = \sup_{c \in C} \langle x^*, c \rangle.$$

— $N_C(x)$ is the normal cone to C at x ,

$$N_C(x) = \{x^* \in \mathcal{H} : \langle x^*, c - x \rangle \leq 0 \text{ for all } c \in C\} \text{ if } x \in C, \text{ and } \emptyset \text{ otherwise.}$$

— $R(N_C)$ is the range of N_C , i.e. $p \in R(N_C)$ if and only if $p \in N_C(x)$ for some $x \in C$.

— Note that $\delta_C^* = \sigma_C$ where δ_C is the indicator function of C ,

$$\delta_C := \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise.} \end{cases}$$

Observe that in $(\mathcal{H}_1)_\epsilon$, all the terms in the integral are nonnegative. Indeed, since Θ is bounded from above by the indicator function of the set C , i.e. $\Theta \leq \delta_C$ (recall that $\Theta = 0$ on C), the reverse inequality holds for their Fenchel conjugates, whence

$$\Theta^*(\epsilon(t)p) - \sigma_C(\epsilon(t)p) \geq 0 \quad \forall p \in \mathcal{H}.$$

Thus, Hypothesis $(\mathcal{H}_1)_\epsilon$ means that, for all $p \in R(N_C)$ the nonnegative function

$$t \mapsto [\Theta^*(\epsilon(t)p) - \sigma_C(\epsilon(t)p)]$$

is integrable on $(0, +\infty)$. For more clarity, let us discuss the following special case : Suppose that

$$\Theta(x) \geq \frac{r}{2} \text{dis}^2(x, C),$$

for some $r > 0$. Then $\Theta^*(x) \leq \frac{1}{2r} \|x\|^2 + \sigma_C(x)$ and

$$\Theta^*(z) - \sigma_C(z) \leq \frac{1}{2r} \|z\|^2.$$

Hence, in this situation $(\mathcal{H}_1)_\epsilon$ is satisfied if the following condition on $\epsilon(\cdot)$ is satisfied :

$$\int_0^{+\infty} \epsilon^2(t) < +\infty.$$

In this situation, the moderate growth condition on $\epsilon(\cdot)$, can be formulated as

$$\epsilon(\cdot) \in L^2(0, +\infty) \setminus L^1(0, +\infty).$$

Let us return to the general situation, and summarize our results in the following theorem, which is our main statement.

Theorem 5.2.11 *Let $\Phi : \mathcal{H} \rightarrow \mathbb{R}^+ \cup \{+\infty\}$ be a closed convex function, such that $C = \text{argmin}\Phi = \Phi^{-1}(0) \neq \emptyset$. Let us assume that,*

- $(\mathcal{H}_1)_\varepsilon \forall p \in R(N_C), \int_0^{+\infty} \mu \Phi^*\left(\frac{1}{\mu}\varepsilon(t)p\right) - \sigma_C(\varepsilon(t)p) + \frac{1}{2}\|\varepsilon(t)p\|^2 dt < +\infty;$
- $(\mathcal{H}_2)_\varepsilon \varepsilon(\cdot)$ is a nonincreasing function of class C^1 , Lipschitz continuous on $[0, +\infty[$, and such that $\lim_{t \rightarrow +\infty} \varepsilon(t) = 0$, $\int_0^{+\infty} \varepsilon(t) dt = +\infty$, and for some $k \geq 0$, $-k\varepsilon^2 \leq \dot{\varepsilon}$.

Then, for any trajectory $(x(\cdot), v(\cdot)) : [0, +\infty[\rightarrow \mathcal{H} \times \mathcal{H}$ solution of (5.2.7a)-(5.2.7b), with $y(t) = x(t) + \mu v(t)$

$$(i) \text{ weak convergence} \quad w - \lim_{t \rightarrow +\infty} y(t) = \text{proj}_{\text{argmin}\Phi} 0.$$

Let us further assume that $\Phi(0) < +\infty$. Then

- (ii) weak convergence $w - \lim_{t \rightarrow +\infty} x(t) = w - \lim_{t \rightarrow +\infty} y(t) = \text{proj}_{\text{argmin}\Phi} 0;$
- (iii) strong convergence $s - \lim_{t \rightarrow +\infty} v(t) = 0$, and hence $s - \lim_{t \rightarrow +\infty} x(t) - y(t) = 0$.

Proof. Let us apply Theorem 5.2.10 with $\Theta(y) = \mu\Phi_\mu(y)$, and $\Psi(y) = \mu^2(\Phi^*)_{\frac{1}{\mu}}\left(\frac{1}{\mu}y\right)$, which are two convex continuously differentiable functions. By the general properties of the Moreau envelope, the function Θ is still nonnegative, and

$$\text{argmin}\Theta = \text{argmin}\Phi_\mu = \text{argmin}\Phi = C.$$

Moreover, on C we have $\Theta = 0$. Let us particularize the conditions $(\mathcal{H}_1)_\varepsilon$ and $(\mathcal{H}_2)_\varepsilon$ to our setting. Let us first compute Θ^* , the conjugate of Θ . For any $z \in \mathcal{H}$

$$\begin{aligned} \Theta^*(z) &= \sup_y \{ \langle z, y \rangle - \mu\Phi_\mu(y) \} \\ &= \mu \sup_y \left\{ \left\langle \frac{1}{\mu}z, y \right\rangle - \Phi_\mu(y) \right\} \\ &= \mu (\Phi_\mu)^* \left(\frac{1}{\mu}z \right) \\ &= \mu \left(\Phi^* \left(\frac{1}{\mu}z \right) + \frac{\mu}{2} \left\| \frac{1}{\mu}z \right\|^2 \right) \\ &= \mu \Phi^* \left(\frac{1}{\mu}z \right) + \frac{1}{2} \|z\|^2. \end{aligned}$$

Hence conditions $(\mathcal{H}_1)_\varepsilon$ and $(\mathcal{H}_2)_\varepsilon$ of Theorem 5.2.10 are satisfied. As a consequence, we obtain the convergence of $y(\cdot)$ to a solution y_∞ of the constrained minimization problem

$$\min \{ \Psi(x) : x \in C \}. \tag{5.2.41}$$

Let us write the first-order optimality condition satisfied by y_∞ . We have

$$\nabla \Psi(y_\infty) + N_C(y_\infty) \ni 0. \tag{5.2.42}$$

Since $\nabla \Psi = \mu \text{prox}_{\mu\Phi}$, equivalently

$$\mu \text{prox}_{\mu\Phi}(y_\infty) + N_C(y_\infty) \ni 0. \tag{5.2.43}$$

Noticing that $y_\infty \in C$, and that $z = \text{prox}_{\mu\Phi} z$ for $z \in C = \text{argmin}\Phi$, we obtain

$$\mu y_\infty + N_C(y_\infty) \ni 0. \quad (5.2.44)$$

By definition of N_C , equivalently, the following property is satisfied

$$\langle 0 - y_\infty, c - y_\infty \rangle \leq 0 \quad \forall c \in C.$$

Since $y_\infty \in C$, this is the condition of the obtuse angle that characterizes the projection of the origin on C . Thus

$$y_\infty = \text{proj}_C(0). \quad (5.2.45)$$

We have obtained that $y(\cdot)$ converges weakly to the element of minimal norm of the solution set C , that's item (i). In order to pass from the convergence of $y(\cdot)$ to the convergence of $x(\cdot)$ we use the relation (5.2.12)

$$x(t) = \text{prox}_{\mu\Phi}(y(t))$$

that links the two variables.

In a finite dimensional setting, we can conclude the strong convergence of $x(\cdot)$ thanks to the continuity of the proximal mapping, and using again the fact that the set C of minimizers of Φ is invariant by the proximal mapping $\text{prox}_{\mu\Phi}$, i.e.,

$$\text{prox}_{\mu\Phi}(y) = y \quad \text{for all } y \in C = \text{argmin}\Phi.$$

In an infinite dimensional setting, we are going to use the particular structure of our dynamical system, and an energetical argument to show that

$$x(t) - y(t) \rightarrow 0 \text{ strongly as } t \rightarrow +\infty. \quad (5.2.46)$$

This will result from the finite energy property

$$\int_0^\infty \|\dot{y}(t)\|^2 dt < +\infty. \quad (5.2.47)$$

To obtain (5.2.47), take the scalar product of equation (5.2.38) with $\dot{y}(t)$. We obtain

$$\|\dot{y}(t)\|^2 + \frac{d}{dt}(\Theta(y(t))) + \epsilon(t) \frac{d}{dt}(\Psi(y(t))) = 0. \quad (5.2.48)$$

After integration by parts we obtain, for any $T > 0$

$$\int_0^T \|\dot{y}(t)\|^2 dt + \Theta(y(T)) - \Theta(y(0)) + \epsilon(T) \Psi(y(T)) - \epsilon(0) \Psi(y(0)) - \int_0^T \dot{\epsilon}(t) \Psi(y(t)) dt = 0. \quad (5.2.49)$$

By assumption $\Phi(0) < +\infty$. By Remark 5.2.6 $\Phi^{**} = \Phi$, we equivalently have $\inf \Phi^* > -\infty$, and hence

$$\inf \Psi = \inf \mu^2 (\Phi^*)_{\frac{1}{\mu}} = \mu^2 \inf \Phi^* > -\infty.$$

Set $m := \inf \Psi$. From (5.2.49), and $\dot{\epsilon}(t) \leq 0$ (recall that $\epsilon(\cdot)$ is a nonincreasing function) we deduce that

$$\int_0^T \|\dot{y}(t)\|^2 dt \leq \Theta(y(0)) + \epsilon(0) \Psi(y(0)) + |m| \epsilon(T) + m \int_0^T \dot{\epsilon}(t) dt. \quad (5.2.50)$$

Since the above majorization is valid for any $T > 0$, and ϵ is bounded (it decreases to zero), we obtain (5.2.47).

We now observe that $y(\cdot)$ is Lipschitz continuous on $[0, +\infty[$. This follows from equation (5.2.38), and the following argument. Since $y(\cdot)$ is converging weakly, it is bounded. By the Lipschitz continuity of the operators $\nabla\Theta = \mu\nabla\Phi_\mu$ and $\nabla\Psi = \mu\text{prox}_{\mu\Phi}$ (which are therefore bounded on bounded sets), and by equation (5.2.38), we deduce that $\dot{y}(\cdot)$ is bounded, and hence $y(\cdot)$ is Lipschitz continuous on $[0, +\infty[$.

Using again equation (5.2.38), and the Lipschitz continuity properties of $\nabla\Theta$ and $\nabla\Psi$, y , and ϵ (for this last property note that for some $c > 0$, $-c \leq -k\epsilon^2 \leq \dot{\epsilon} \leq 0$), we deduce that $t \mapsto \dot{y}(t)$ is Lipschitz continuous on $[0, +\infty[$. Hence $\dot{y}(\cdot)$ belongs to $L^2([0, +\infty[; \mathcal{H})$, and is Lipschitz continuous. By a classical result this implies

$$\lim_{t \rightarrow +\infty} \dot{y}(t) = 0.$$

Returning to (5.2.38), and noticing that $\epsilon(t)\nabla\Psi(y(t)) \rightarrow 0$, we obtain

$$\lim_{t \rightarrow +\infty} \nabla\Theta(y(t)) = 0.$$

Since $\nabla\Theta(y(t)) = y(t) - \text{prox}_{\mu\Phi}(y(t)) = y(t) - x(t)$, we finally obtain that $x(t) - y(t)$ converges strongly to zero as $t \rightarrow +\infty$, which clearly implies that $x(\cdot)$ and $y(\cdot)$ converge weakly to the same limit, which is the solution of a hierarchical minimization problem. \square

Example : Let us return to our model situation where

$$\Phi(x) \geq \frac{r}{2} \text{dis}^2(x, C),$$

for some $r > 0$. Then $\Phi^*(x) \leq \frac{1}{2r} \|x\|^2 + \sigma_C(x)$ and

$$\Phi^*(z) - \sigma_C(z) \leq \frac{1}{2r} \|z\|^2.$$

After elementary computation, one can verify that, in this situation, $(\mathcal{H}_1)_\epsilon$ is satisfied if the following condition on $\epsilon(\cdot)$ is satisfied :

$$\int_0^{+\infty} \epsilon^2(t) < +\infty.$$

Thus, in this situation, the moderate growth condition on $\epsilon(\cdot)$, can be formulated as

$$\epsilon(\cdot) \in L^2(0, +\infty) \setminus L^1(0, +\infty).$$

Strong convergence

Let us now examine the strong convergence properties of the trajectories. Let us first consider the variable $y(\cdot)$. Following [6, Theorem 2.2], and equation (5.2.38), the strong convergence of $y(\cdot)$ will result from the strong monotonicity property of $\nabla\Psi = \mu\text{prox}_{\mu\Phi}$. We recall that $\nabla\Psi$ is said to be strongly monotone if there exists some $\alpha > 0$ such that for any $x \in \text{dom}\nabla\Psi$, and $y \in \text{dom}\nabla\Psi$

$$\langle \nabla\Psi(x) - \nabla\Psi(y), x - y \rangle \geq \alpha \|x - y\|^2.$$

This property turns out to be equivalent to a regularity property for Φ , as stated in the following Lemma.

Lemma 5.2.12 *Let Φ be a convex differentiable function whose gradient is L -Lipschitz continuous for some $L > 0$. Then, for any $\mu > 0$ such that $\mu L < 1$, the proximal mapping $\text{prox}_{\mu\Phi}$ is strongly monotone.*

Proof. Take y_i , $i = 1, 2$. By definition of $\text{prox}_{\mu\Phi}(y_i)$, $\text{prox}_{\mu\Phi}(y_i) + \mu \nabla \Phi(\text{prox}_{\mu\Phi}(y_i)) = y_i$. Taking the difference of the two equations, and multiplying scalarly by $y_2 - y_1$, we obtain

$$\langle \text{prox}_{\mu\Phi}(y_2) - \text{prox}_{\mu\Phi}(y_1), y_2 - y_1 \rangle + \mu \langle \nabla \Phi(\text{prox}_{\mu\Phi}(y_2)) - \nabla \Phi(\text{prox}_{\mu\Phi}(y_1)), y_2 - y_1 \rangle = \|y_2 - y_1\|^2.$$

Then use the Cauchy-Schwarz inequality, the L -Lipschitz continuity of $\nabla \Phi$, and the fact that the proximal mapping is nonexpansive to obtain

$$\langle \text{prox}_{\mu\Phi}(y_2) - \text{prox}_{\mu\Phi}(y_1), y_2 - y_1 \rangle \geq (1 - \mu L) \|y_2 - y_1\|^2.$$

Conversely, one can easily establish that the strong monotonicity of the proximal mapping implies that Φ is a convex differentiable function whose gradient is Lipschitz continuous. \square

We can now complete Theorem 5.2.11 as follows.

Theorem 5.2.13 *Let us make the assumptions of Theorem 5.2.11, and assume moreover that Φ is a convex differentiable function whose gradient is L -Lipschitz continuous for some $L > 0$. Then for $\mu L < 1$, we have the strong convergence property of $x(\cdot)$ and $y(\cdot)$ to the element of minimal norm of $C = \text{argmin} \Phi \neq \emptyset$.*

Proof. By Theorem 5.2.11 item (iii), $x(t) - y(t)$ converges strongly to zero as $t \rightarrow +\infty$. Hence we just need to prove that $y(\cdot)$ converges strongly. Since $\mu L < 1$, by Lemma 5.2.12, the operator $\nabla \Psi = \mu \text{prox}_{\mu\Phi}$ is strongly monotone. Thus we are in the situation examined in [6, Theorem 2.2], which gives the strong convergence property. Another equivalent approach consists in noticing that by Theorem 5.2.10, we have $w - \lim_{t \rightarrow +\infty} y(t) = \text{proj}_{\text{argmin} \Phi} 0$ and $\Psi(y(t)) \rightarrow \Psi(\text{proj}_{\text{argmin} \Phi} 0)$. From this we easily deduce that the strong convexity of Ψ implies the strong convergence of $y(\cdot)$. We recover our result by noticing that the strong convexity of Ψ is equivalent to the strong monotonicity of its gradient, i.e., of the proximal mapping. \square

5.2.4 Other viscosity selection principles

Let us now examine the more general situation

$$v(t) \in \partial \Phi(x(t)) \tag{5.2.51a}$$

$$\lambda \dot{x}(t) + \dot{v}(t) + v(t) + \varepsilon(t) \partial g(x(t)) \ni 0, \tag{5.2.51b}$$

where g is a convex viscosity function.

Using the Minty transform, this system can be equivalently written as

$$x(t) = \text{prox}_{\mu\Phi}(y(t)) \tag{5.2.52a}$$

$$\dot{y}(t) + \mu \nabla \Phi_{\mu}(y(t)) + \mu \varepsilon(t) \partial g(\text{prox}_{\mu\Phi}(y(t))) \ni 0. \tag{5.2.52b}$$

In order to recover exactly the Tikhonov approximation, we look for some g such that, for all $y \in \mathcal{H}$

$$\partial(\mu g)(\text{prox}_{\mu\Phi}(y)) = y.$$

Equivalently

$$(I + \mu\partial\Phi)^{-1} = (\partial(\mu g))^{-1}.$$

We obtain

$$I + \mu\partial\Phi = \partial(\mu g),$$

that is, for all $y \in \mathcal{H}$

$$\mu g(y) = \frac{1}{2}\|y\|^2 + \mu\Phi(y).$$

Thus, by taking $g(y) = \frac{1}{2\mu}\|y\|^2 + \Phi(y)$, and $\Theta(y) = \mu\Phi_\mu(y)$, equation (5.2.52b) can be equivalently written as

$$\dot{y}(t) + \nabla\Theta(y(t)) + \epsilon(t)y(t) = 0. \quad (5.2.53)$$

Equation (5.2.53) is a particular case of the (SDC) system (steepest descent with control)

$$(SDC) \quad \dot{y}(t) + \partial\Theta(y(t)) + \epsilon(t)y(t) = 0.$$

Concerning the case $\int_0^{+\infty} \epsilon(t) dt = +\infty$, the first general convergence result was in [19] (based on previous work by [13]), and also requires $\epsilon(\cdot)$ to be nonincreasing, and converges to zero for $t \rightarrow +\infty$. Under these conditions, each trajectory of (5.2.53) converges strongly to the point of minimal norm in $C = \arg \min \Theta = \arg \min \Phi$. In [15], it is proved that the convergence result still holds without assuming $\epsilon(\cdot)$ to be nonincreasing.

When $g(\xi) = \frac{1}{2\mu}\|\xi\|^2 + \Phi(\xi)$, the dynamical system (5.2.51a)-(5.2.51b) becomes

$$v(t) \in \partial\Phi(x(t)) \quad (5.2.54a)$$

$$\lambda\dot{x}(t) + \dot{v}(t) + v(t) + \epsilon(t)\left(\frac{1}{\mu}x(t) + v(t)\right) = 0. \quad (5.2.54b)$$

Equivalently

$$v(t) \in \partial\Phi(x(t)) \quad (5.2.55a)$$

$$\dot{x}(t) + \mu\dot{v}(t) + \mu(1 + \epsilon(t))v(t) + \epsilon(t)x(t) = 0. \quad (5.2.55b)$$

As in the preceding section, by application of the Cauchy-Lipschitz theorem (recall that Θ is differentiable, and its gradient is Lipschitz continuous), we can show that (5.2.53) admits a strong global solution $y(\cdot)$. Then, $(x(\cdot), v(\cdot))$ with $x(\cdot) = \text{prox}_{\mu\Phi}(y(\cdot))$ and $v(\cdot) = \nabla\Phi_\mu(y(\cdot))$ is a strong solution of (5.2.55a)-(5.2.55b).

Let us summarize our result in the following theorem.

Theorem 5.2.14 *Let $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a closed, convex, proper function, such that $C = \arg \min \Phi \neq \emptyset$. Let us assume that*

$$(i) \quad \lim_{t \rightarrow +\infty} \epsilon(t) = 0,$$

$$(ii) \quad \int_0^{+\infty} \epsilon(t) dt = +\infty.$$

Then, for any trajectory $(x(\cdot), v(\cdot)) : [0, +\infty[\rightarrow \mathcal{H} \times \mathcal{H}$ solution of (5.2.55a)- (5.2.55b), with $y(t) = x(t) + \mu v(t)$ the following strong convergence property holds :

$$s - \lim_{t \rightarrow +\infty} x(t) = s - \lim_{t \rightarrow +\infty} y(t) = \underset{\arg \min \Phi}{\text{Proj}} 0.$$

Proof. We are in the situation examined in [15, Theorem 2], which gives the strong convergence of each trajectory $y(\cdot)$ of (5.2.53) towards the point of minimal norm in $C = \operatorname{argmin} \Phi$. In order to pass from the convergence of $y(\cdot)$ to the convergence of $x(\cdot)$ we use the relation (5.2.52a)

$$x(t) = \operatorname{prox}_{\mu\Phi}(y(t))$$

that links the two variables. From the continuity property of the proximal mapping for the strong topology (indeed, it is a nonexpansive mapping), we deduce that each trajectory $x(\cdot)$ of (5.2.55a)- (5.2.55b) converges strongly to $\operatorname{prox}_{\mu\Phi} \left(\operatorname{Proj}_{\operatorname{argmin} \Phi} 0 \right) = \operatorname{Proj}_{\operatorname{argmin} \Phi} 0$. To obtain this last equality, we use the fact that the set C of minimizers of Φ is invariant by the proximal mapping $\operatorname{prox}_{\mu\Phi}$, i.e.,

$$\operatorname{prox}_{\mu\Phi}(y) = y \quad \text{for all } y \in C = \operatorname{argmin} \Phi.$$

□

5.2.5 Perspective

Let us list some interesting questions to be examined in the future :

1. Examine the discrete, algorithmical version, and the corresponding asymptotic selection property for the forward-backward algorithm.
2. Study the case where $\lambda(t)$ depends on t in an open-loop form, as in [9].
3. Study the case where the Levenberg-Marquart regularization term is given in a closed-loop form, $\lambda(t) = \alpha(\|\dot{x}(t)\|)$ as in [8].
4. Examine these questions for the related dynamical systems which have been considered in [1].

Bibliographie

- [1] B. Abbas, H. Attouch, Dynamical systems and forward-backward algorithms associated with the sum of a convex subdifferential and a monotone cocoercive operator, *Optimization*, (2014) <http://dx.doi.org/10.1080/02331934.2014.971412>.
- [2] B. Abbas, H. Attouch, B. F. Svaiter, Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces, *J. Optim. Theory Appl.*, **161** (2014), No. 2, pp. 331-360.
- [3] F. Alvarez, A. Cabot, Asymptotic selection of viscosity equilibria of semilinear evolution equations by the introduction of a slowly vanishing term, *Discrete and Continuous Dynamical Systems*, **15** (2006), pp. 921-938.
- [4] H. Attouch, Viscosity solutions of minimization problems, *SIAM J. Optimization*, **6** (1996), pp. 769-806.
- [5] H. Attouch, G. Buttazzo, G. Michaille, Variational analysis in Sobolev and BV spaces. Applications to PDE's and optimization, *MPS/SIAM Series on Optimization*, **6**, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, Second edition, 2014, 793 pages.
- [6] H. Attouch, M.-O. Czarnecki, Asymptotic behavior of coupled dynamical systems with multiscale aspects, *Journal of Differential Equations*, **248** (2010), pp. 1315-1344.
- [7] H. Attouch, J. Peypouquet, P. Redont, Backward-forward algorithms for structured monotone inclusions in Hilbert spaces, to appear.
- [8] H. Attouch, P. Redont, B. F. Svaiter, Global convergence of a closed-loop regularized Newton method for solving monotone inclusions in Hilbert spaces, *J. Optim. Theory Appl.*, 157 (2013), No. 3, pp. 624-650.
- [9] H. Attouch, B.F. Svaiter, A continuous dynamical Newton-like approach to solving monotone inclusions, *SIAM J. Control Optim.* **49** (2011), pp. 574-598.
- [10] J. B. Baillon, R. Cominetti, A convergence result for nonautonomous subgradient evolution equations and its application to the steepest descent exponential penalty trajectory in linear programming, *Journal of Functional Analysis*, **187** (2001), pp. 263-273.
- [11] H.H. Bauschke, P.L. Combettes, Convex analysis and monotone operator theory in Hilbert spaces. Springer, New York (2011).
- [12] H. Brézis, Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert, North-Holland/Elsevier, New-York, 1973.
- [13] F.E. Browder, Nonlinear Operators and Nonlinear Equations of Evolution in Banach Spaces, Proc. Sympos. Pure Math., vol. 18 (part 2), Amer. Math. Soc., Providence, RI, 1976.

- [14] A. Cabot, The steepest descent dynamical system with control. Applications to constrained minimization, *ESAIM : Control, Optimization and Calculus of Variations*, **10** (2004), pp. 243-258.
- [15] R. Cominetti, J. Peypouquet, S. Sorin, Strong asymptotic convergence of evolution equations governed by maximal monotone operators with Tikhonov regularization, *J. Differential Equations*, **245** (2008), pp. 3753–3763.
- [16] H. Furuya, K. Miyashiba, N. Kenmochi, Asymptotic behavior of solutions to a class of nonlinear evolution equations, *Journal of Differential Equations*, **62** (1986), pp. 73–94.
- [17] A. Haraux, Systèmes dynamiques dissipatifs et applications. RMA 17, Masson, Paris, (1991).
- [18] S. Hirstoaga, Approximation et résolution de problèmes d'équilibre, de point fixe et d'inclusion monotone, PhD thesis, UPMC Paris 6, (2006).
- [19] S. Reich, Nonlinear evolution equations and nonlinear ergodic theorems, *Nonlinear Anal.* 1 (1976) 319–330.
- [20] R.T Rockafellar, R. J-B. Wets, Variational Analysis, Grundlehren der mathematischen Wissenschaften **317**, Springer-Verlag, (1998).
- [21] E.D. Sontag, Mathematical Control Theory, second edition. Springer-Verlag, New-York (1998).

Chapitre 6

Conclusion et perspectives

Cette thèse a été consacrée à la recherche des zéros d'un opérateur maximal monotone structuré, à l'aide de systèmes dynamiques dissipatifs continus et discrets. Les solutions sont obtenues comme limites des trajectoires lorsque le temps t tend vers l'infini. Notre travail a porté principalement sur l'étude de nouvelles dynamiques obtenues par régularisation de la méthode de Newton.

Dans un cadre hilbertien, on s'est intéressé à la recherche des zéros d'un opérateur maximal monotone structuré $M = A + B$, où A est un opérateur maximal monotone général et B est un opérateur monotone lipschitzien. Nous avons introduit des dynamiques continues et discrètes de type Newton régularisé faisant intervenir d'une façon séparée les résolvantes de l'opérateur A (implicites), et des évaluations de B (explicites). A l'aide de la représentation de Minty de l'opérateur A comme une variété Lipschitzienne, nous avons reformulé ces dynamiques sous une forme relevant du théorème de Cauchy-Lipschitz, et nous avons montré l'existence et l'unicité d'une solution globale forte de notre problème d'évolution. On s'est intéressé au cas particulier où A est le sous différentiel d'une fonction convexe, semi-continue inférieurement, et propre, et B est le gradient d'une fonction convexe, différentiable, B étant supposé seulement localement lipschitzien (hypothèse un peu plus faible que dans le cas général, où B est supposé monotone globalement lipschitzien). Lorsque le terme de régularisation ne tend pas trop vite vers zéro et en s'appuyant sur une analyse asymptotique de type Lyapunov, nous avons étudié le comportement asymptotique des trajectoires, et montré leur convergence. Par ailleurs, nous avons montré la dépendance lipschitzienne des trajectoires par rapport au terme de régularisation, lorsque ce dernier agit en boucle ouverte.

Comme travail pour le futur :

- On voudrait prolonger cette étude par l'étude de la régularisation en boucle fermée de la méthode de Newton cf. [4], et étudier les propriétés de stabilité dans ce cas.
- Etablir le lien entre la dynamique de Newton régularisée étudiée avec la dynamique du deuxième ordre en temps, avec un terme accélération supplémentaire cf. [3]-[5].
- Application en traitement du signal et d'image, optimisation parcimonieuse (sparse optimization), et équations aux dérivées partielles.
- Etudier la stabilité des trajectoires sur l'intervalle $[0, \infty]$.
- Développer la méthode de Newton régularisée dans un cadre plus général non convexe non lisse (prox-régulier...). Comparer avec l'approche semi-smooth développée dans [2].

Le chapitre 4 a présenté différentes classes de systèmes dynamiques, et les algorithmes forward-backward associés, visant à résoudre les inclusions monotones gouvernées par un opérateur

maximal monotone structuré $M = \partial\Phi + B$, où $\partial\Phi$ désigne le sous-différentiel d'une fonction convexe, semi-continue inférieurement, et propre Φ , et B est un opérateur monotone cocoercif. Les systèmes dynamiques considérés sont : la dynamique continue Newton régularisé, le semi-groupe de contractions généré par $-(\partial\varphi + B)$, et le système dynamique proximal-gradient. En s'appuyant sur une analyse asymptotique de type Lyapunov, nous avons étudié le comportement asymptotique des trajectoires de ces systèmes. La discrétisation temporelle de ces dynamiques fournit des algorithmes forward-backward (certains nouveaux).

Une illustration numériques a permis de comparer les comportements de ces trois algorithmes dans un cas modèle simple.

Les perspectives concernant cette partie de notre travail sont les suivantes :

- Analyser le taux de convergence de la dynamique de Newton régularisé dans le cas d'inclusions gouvernées par un opérateur fortement monotone structuré cf. [7].
- Etudier dans leurs versions inertielles les trois dynamiques continues et discrètes considérées précédemment.
- Application à la restauration du signal et la tomographie.
- Etudier les algorithmes forward-backward précédents dans le cadre de la minimisation de la somme d'une fonction convexe non-lisse avec une fonction non-convexe lisse, en liaison avec [6].

Le chapitre 5 a proposé une méthode de Newton régularisée (à la Tikhonov) pour atteindre asymptotiquement un équilibre de norme minimale, dans le cas d'une fonction convexe, semi-continue inférieurement, et propre.

Comme travail pour le futur, on envisage :

- Examiner le cas discret, la version algorithmique, et étudier la sélection asymptotique d'équilibre pour l'algorithme backward-forward.
- Etudier le cas où le terme de régularisation à la Levenberg-Marquardt dépend du temps et agit en boucle ouverte.
- Etudier le cas où le terme de régularisation à la Levenberg-Marquardt agit en boucle fermée.
- Examiner ces mêmes questions pour les dynamiques considérées dans cf. [1].

Bibliographie

- [1] B. Abbas, H. Attouch, Dynamical systems and forward-backward algorithms associated with the sum of a convex subdifferential and a monotone cocoercive operator, *Optimization*, (2014) <http://dx.doi.org/10.1080/02331934.2014.971412>.
- [2] S.Adly, R. Cibulka, H.V. Ngai, Newton's method for solving inclusions using set-valued approximations, *SIAM Journal on Optimization*, 25 (2015), no. 1, 159–184.
- [3] Alvarez, F., Attouch, H., Bolte, J., Redont, P. : A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics. *J. Math. Pures Appl.* 81, 747–779 (2002)
- [4] Attouch, H., Redont, P., Svaiter, B.F. : Global convergence of a closed-loop regularized Newton method for solving monotone inclusions in Hilbert spaces. *J. Optim. Theory Appl.* 157(3), 624–650 (2013)
- [5] Attouch, H., Maingé, P.-E., Redont, P. : A second-order differential system with Hessian-driven damping ; application to non-elastic shock laws. *Differ. Equ. Appl.* 4, 27–65 (2012)
- [6] R.I. Boţ, E.R. Csetnek, A forward-backward dynamical approach to the minimization of the sum of a nonsmooth convex with a smooth nonconvex function, *arXiv :1507.01416* , (2015)
- [7] R.I. Boţ, E.R. Csetnek, Convergence rates for forward-backward dynamical systems associated with strongly monotone inclusions, *arXiv :1504.01863*, (2015)