UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

DIFFERENTIAL INCLUSIONS AND OPTIMIZATION ALGORITHMS OF FORWARD-BACKWARD TYPE

TESIS PARA OPTAR AL GRADO DE
DOCTOR EN CIENCIAS DE LA INGENIERÍA, MENCIÓN MODELACIÓN MATEMÁTICA

ANDRÉS ANTONIO CONTRERAS TAVÁREZ

PROFESOR GUÍA:
JUAN PEYPOUQUET

MIEMBROS DE LA COMISIÓN:
ARIS DANIILIDIS
ABDERRAHIM HANTOUTE
JALAL FADILI
MARCO ANTONIO LÓPEZ CERDÁ

Este trabajo ha sido parcialmente financiado por CMM- Conicyt PIA AFB170001

SANTIAGO DE CHILE
2020

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE DOCTOR EN CIENCIAS DE
LA INGENIERIA, MENCIÓN MODELACIÓN
MATEMÁTICA
POR: ANDRÉS CONTRERAS TAVÁREZ
FECHA: ENERO DE 2020
PROF. GUÍA: JUAN PEYPOUQUET

DIFFERENTIAL INCLUSIONS AND OPTIMIZATION ALGORITHMS OF FORWARD-BACKWARD TYPE

This thesis has two main purposes: first, we shall compare continuous and discrete trajectories associated to a differential inclusion governed by the sum of a maximally monotone operator with a cocoercive operator and the discrete trajectories generated by means of forward backward algorithm, in a finite and infinite horizon. As a consequence of the general theory, we obtain new results on the strong convergence for the forward backward algorithm. In the particular case of the explicit iterations governed by an operator deriving from a potential, we present important results concerning to the strong convergence of gradient algorithm. The second purpose in this thesis is to show some results obtained about the acceleration of certain primal-dual algorithms for image processing, by using the asymptotic properties of the preconditioned forward backward algorithm. We deal specifically with the algorithms presented by Loris–Verhoeven [53] and Condat–Vu [27, 69], when the cocoercive operator in both algorithms is affine.

Esta tesis tiene dos propósitos principales: primero, vamos a comparar trayectorias continuas asociadas a una inclusón diferencial gobernada por la suma de un operador maximálmente monótono con un operador cocoercivo y las trayectorias discretas generadas a través del algoritmo forward backward, en un horizonte temporal finito e infinito. Como consecuencia de la teoría general, obtenemos nuevos resultados sobre la convergencia fuerte para el algoritmo forward backward. En el caso particular de iteraciones explícitas por un oerador derivado de un potencial, presentamos inportantes resultados concernientes a la convergencia fuerte del algoritmo del gradiente. El segundo propósito de esta tesis es mostrar algunos resultados obtenidos sobre la aceleración de ciertos algoritmos primal–dual para procesamiento de imágenes. Tratamos específicamente con los algoritmos presentados por Loris–Verhoeven [53] y Condat–Vu [27, 69], cuando el operador cocoercivo en ambos algoritmos es afín.

Dedicated to my parents

Acknowledgments

My first words of thanks are to my advisor Juan Peypouquet, an outstanding teacher, thesis director and human being, a brilliant researcher who drove my attention to the fascinating topic of differential inclusions and its relevance in the study of optimization algorithms and for his valuable recommendations for the writing of this manuscript. It would be an honor and pleasure to continue working in collaboration in the future.

I also thank to Aris Daniilidis, Abderrahim Hantoute, Jalal Fadili, Marco Antonio López Cerdá for attending the pressentation and having reviewed this manuscript. I would like to express my gratitude to proffesors Hong-Kun Xu and Laurent Condat for receiving me as visiting student in the Department of Mathematics of Hangzhou Dianzi University and GIPSAL-LAB of Grenobles-Alpes University respectively. Interacting with them was a remarkable experience.

The development of this thesis would not have been possible without the financial support of CONICYT-PCHA/Doctorado Nacional/2016-21160994 and CMM-Conicyt PIA AFB170001, who supported my attendance to several workshops, conferences in Chile and visit research abroad, which gave me the oportunity to present my work to a broader community and receive useful feedback. I also thank the Department of Mathematical Engineering at the UCH and their staff for their valuable help. I would like to thank Department of Mathematics of the Universidad Técnica Federico Santa María where I started my Ph.D studies in Chile, especially the professors Luis Briceño Arias, Eduardo Cerpa and Erwin Hernández.

Finally, I wish to thank all the people who contributed to the human enviroment in which I carried out my work: Jesús Vellojín and his wife Rosa Delgado, for being my brothers and friend and their invaluable support during my stay in Valparaíso; Cesare Molinari, mi hermano académico; Alejandra Alvarado, Diana Narváez, Claudia Soto, Ariel Perez, Stephanie Caro for being there; and Roberto Molina for his unconditional support and patience during these years.

Table of Contents

Chapter 1

Introduction

This thesis has two main purposes: first, we shall compare continuous and discrete trajectories associated to a differential inclusion governed by a multivalued operator, looking at their qualitative differences and similarities in terms of the convergence of these trajectories. If the operator is maximally monotone and the discrete trajectories are obtained by the proximal algorithm, there exists a theory which enables us, on the one hand, to quantify the aproximation of the solution of the differential inclusion by proximal iterations in a finite horizon and, on the other hand, it allows to analyze the relationship between the long term behavior of the continuous and discrete trajectories. This theory can be applied even when the operator is the sum of a maximally monotone operator with a cocoercive operator, but in practice, the computation of the proximal iterations governed by the sum of two operators is complicated. It is more convenient to obtain the discrete trajectories by means of the forward backward algorithm. However, the comparison theory in the latter case does not exist. One of our main goals is to build this theory in this manuscript. The second purpose in this thesis is to show some results obtained about the acceleration of certain primal-dual algorithms for image processing, by using the asymptotic properties of the preconditioned forward backward algorithm. We deal specifically with the algorithms presented by Loris–Verhoeven [53] and Condat–Vu [27, 69], when the cocoercive operator in both algorithms is affine.

This chapter is organized as follows: In Section 1.1, we present the comparison theory between the discrete trajectories generated by the forward backward algorithm and the continuous ones associated to the differential inclusion governed by sum of a maximally monotone operator with a cocoercive operator in a finite and infinite horizon. As a consequence of the general theory, we obtain new results on the strong convergence for the forward backward algorithm. In the particular case of the explicit iterations governed by an operator deriving from a potential, we present important results concerning to the strong convergence of gradient algorithm. We finish this section by prensenting some comments about the work in progress. In Section 1.2, we present the main results concerning to acceleration of some primal-dual algorithm. We finish this section with some comments about the work in progress.

For simplicity, we restrict ourselves to the Hilbert spaces setting in this chapter.

1.1 Forward-backward approximation of evolution equations in finite and infinite horizon

In this section we summarize the developed research into the papers [30] and [29], related to the relationship of the asymptotic behavior between continuous and discrete dinamycal systems governed by the sum of a maximally monotone operator with a cocoercive operator in Hilbert space setting. The proof of the main results in this section can be found in Chapters 3 and 4.

Throughout this chapter, H will be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ its norm, $A : H \to 2^H$ will be a *maximally monotone* operator, $B : H \to H$ will be a *cocoercive* operator with parameter $\theta > 0$. Then $A + B$ is maximally monotone, the operator

$$E_\lambda = I - \lambda B \tag{1.1}$$

is nonexpansive for all $\lambda \in [0, 2\theta]$ and the *forward backward splitting operator* $T_\lambda : H \to H$ defined by

$$T_\lambda = J_\lambda^A \circ E_\lambda, \tag{1.2}$$

is single-valued, everywhere defined and nonexpansive.

Let us consider the following *forward-backward* iterations defined by

$$x_k = T_{\lambda_k}(x_{k-1}) = J_{\lambda_k}^A(E_{\lambda_k}(x_{k-1})), \qquad k \in \mathbb{N}, \tag{1.3}$$

where $\{\lambda_k\}_{k\in\mathbb{N}}$ is a sequence of positive numbers, called step sizes and $x_0 \in H$.

These *forward-backward iterations* are fundamental in the numerical analysis of optimization problems, since they serve as building blocks for first order methods. The gradient method, originally introduced by Cauchy in [15], and its variant, the projected gradient method [38, 49], the proximal point algorithm introduced by Martinet [54] and further extended by Rockafellar [67] and Brézis-Lions [11], and the proximal-gradient algorithm [63, 51], with applications in image and signal processing, such as the *iterative shrinkage thresholding algorithm* [33, 24], are keynote particular cases. Moreover, some primal dual methods [16, 27, 69] can be reduced to these types of iterations.

We will not study the convergence of the iterations (1.3) in this section. Instead, our main purpose here is to analyze them as discrete approximations of an evolution equation governed by the sum $A + B$. To this end, it is useful to rewrite (1.3) in a more general way:

$$-\frac{x_k - x_{k-1}}{\lambda_k} + \varepsilon_k \in Ax_k + Bx_{k-1}, \quad k \in \mathbb{N}, \tag{1.4}$$

where ε_k accounts for possible perturbations or computational errors. In the notation of (1.3), this is

$$x_k = J_{\lambda_k}^A \left(E_{\lambda_k}(x_{k-1}) + \lambda_k \varepsilon_k \right). \qquad k \in \mathbb{N} \tag{1.5}$$

When $\varepsilon_k = 0$ for all $k \in \mathbb{N}$, the left-hand in (1.4) side can be interpreted as a discretization of the velocity for a trajectory $t \mapsto u(t)$, so (1.4) can be related to the differential inclusion

$$-\dot{u}(t) \in Au(t) + Bu(t), \tag{1.6}$$

for $t > 0$. In the following sections, we shall establish the nature of this relationship. On the one hand, we shall prove that the iterations described in (1.4) can be used, in at least two different ways, to construct a sequence of curves that approximate the solutions of (1.6) uniformly on each compact time interval. The existence of such solutions is obtained as a byproduct. On the other hand, we shall show that, given A and B, the trajectories satisfying (1.6) will have the same convergence properties, when $t \to \infty$, as the sequences satisfying (1.4), when $k \to \infty$, provided the step sizes are sufficiently small.

The following inequality allows to estimate the distance between two arbitrary iterates of two independent sequences generated by (1.5). It generalizes previous versions given by [44, Lemma 2.1] for m-accretive in Banach space setting and $B = 0$ (see also [65, Proposition 17] for A maximally monotone Hilbert space setting) and [30, Lemma 2.4] for $A = 0$ in the Hilbert space setting.

Theorem 1.1 *Let $\{x_k\}_{k\in\mathbb{N}}$, $\{\widehat{x}_l\}_{l\in\mathbb{N}}$ be two sequences generated by (1.5), with stepsizes $\{\lambda_k\}_{k\in\mathbb{N}}$ and $\{\widehat{\lambda}_l\}_{l\in\mathbb{N}}$, respectively and $0 < \lambda_k, \widehat{\lambda}_l \le 2\theta$ for all $k, l \in \mathbb{N}$. Then, for $u \in D(A)$*

$$\|x_k - \widehat{x}_l\| \le \|x_0 - u\| + \|\widehat{x}_0 - u\| + |\!|\!|(A+B)u|\!|\!| \sqrt{(\sigma_k - \widehat{\sigma}_l)^2 + \tau_k + \widehat{\tau}_l} + S_k + \widehat{S}_l, \qquad (1.7)$$

where $S_k = \sum_{\mathrm{i}=1}^{k} \lambda_{\mathrm{i}} \|\varepsilon_{\mathrm{i}}\|$, $|\!|\!|Au|\!|\!| = \inf_{v\in Au} \|v\|$, $\sigma_k = \sum_{\mathrm{i}=1}^{k} \lambda_{\mathrm{i}}$ and $\tau_k = \sum_{\mathrm{i}=1}^{k} \lambda_{\mathrm{i}}^2$ (similarly for $\widehat{S}_l$, $\widehat{\sigma}_l$ and $\widehat{\tau}_l$).

Approximation in finite horizon

Theorem 1.1 provides existence and regularity results for the evolution equation

$$\begin{cases} -\dot{u}(t) & \in & (A+B)u(t), \quad \text{for almost every } \ t > 0, \\ u(0) & = & u_0 \in \overline{D(A)}, \end{cases} \qquad (1.8)$$

by means of an approximation scheme. For each $t \ge 0$ and $m \ge 1$, set

$$u_m(t) = \left[T_{\frac{t}{m}}\right]^m u_0. \qquad (1.9)$$

In other words, $u_m(t)$ is the m-th term of the forward–backward sequence generated by (1.3) from u_0 using the constant step size $\lambda_k \equiv t/m$ for all $k \in \mathbb{N}$. We shall prove that $\{u_m\}$ converges uniformly on compact intervals to a Lipschitz-continuous function satisfying (1.8).

Proposition 1.2 *The sequence $\{u_m\}_{m\in\mathbb{N}}$ converges pointwise on $[0, +\infty)$ and uniformly on $[0, S]$ for each $S > 0$, to a function $u : [0, +\infty) \to H$ which is globally Lipschitz-continuous with constant $|\!|\!|(A+B)u_0|\!|\!|$ and satisfies (1.8).*

As a consequence of Theorem 1.1 and Proposition 1.2, we have

Corollary 1.3 . *Let $\{x_k\}_{k\in\mathbb{N}}$ be a sequence generated by* (1.3) *for all $k \in \mathbb{N}$ and consider $u : [0, S] \to H$ a solution of* (1.9)*. Then*

(i) The function $t \mapsto |||(A + B)u(t)|||$ is nonincreasing.

(ii) $\|x_k - u(t)\| \leq \|x_0 - u_0\| + \min\{|||(A + B)x_0|||, |||(A + B)u_0|||\}\sqrt{(\sigma_k - t)^2 + \tau_k}$.

Approximation in infinite horizon

In this section, we show that forward backward sequences presented in (1.3) have the same asymptotic behavior as the number of the iterations goes to infinity, as the solutions of the evolution equation (1.8). The key argument is the idea of asymptotic equality introduced by Passty [63], closely related to the notion of almost-orbit, introduced by Miyadera and Kobayasi [46] as we shall explain.

Evolution systems, almost-orbit and asymptotic equivalence. Let C a convex subset of H and let I be the identity operator on H. A *nonexpansive evolution system* on H is a family $\big(U(t, s)\big)_{0\leq s\leq t}$ such that

(i) $U(t, t)z = z$ for all $z \in H$ and $t \geq 0$.

(ii) $U(t, s)U(s, r)z = U(t, r)z$ for all $z \in H$ and all $t \geq s \geq r \geq 0$.

(iii) $\|U(t, s)x - U(t, s)y\| \leq \|x - y\|$ for all $x, y \in H$ and $t \geq s \geq 0$.

Example 1.4 Given $x \in \overline{D(A)}$ and $t \geq 0$, we write

$$\mathcal{S}_t x = u(t), \tag{1.10}$$

where u satisfies (1.8) with $x_0 = x$. Also, for $0 \leq s \leq t$, we write

$$U_{\mathcal{S}}(t, s) = \mathcal{S}(t - s). \tag{1.11}$$

In a similar fashion, if $n \in \mathbb{N}$ and $x \in H$, we denote

$$\mathcal{T}_n x = T_{\lambda_n} \circ \cdots \circ T_{\lambda_1} x, \tag{1.12}$$

where T_λ is given by (1.2). In other words, $\mathcal{T}_n x$ is the n-th term of the forward-backward sequence (1.3) starting from $x \in \overline{D(A)}$. Assume $\{\lambda_n\}_{n\in\mathbb{N}} \notin \ell^1$, and write $\nu(t) = \max\{n \in \mathbb{N} : \sigma_n \leq t\}$. For $0 \leq s \leq t$, we set

$$U_{\mathcal{T}}(t, s) = \prod_{\mathrm{i}=\nu(s)+1}^{\nu(t)} T_{\lambda_{\mathrm{i}}}, \tag{1.13}$$

where the product denotes composition of functions and the empty composition is the identity.

The families $\big(U_{\mathcal{S}}\big)$ and $\big(U_{\mathcal{T}}\big)$, defined in (1.11) and (1.13), respectively, are nonexpansive evolution systems. Actually, the same is true if $\mathcal{S}$ is replaced by any other semigroup of nonexpansive functions on X, and if each $T_{\lambda_{\mathrm{i}}}$ is replaced by any other nonexpansive function on H.

On the other hand, let U be a nonexpansive evolution system. An *almost-orbit* of U is a function $\phi : [0, +\infty) \to H$ which satisfies

$$\lim_{t\to+\infty} \sup_{h\geq 0} \|\phi(t+h) - U(t+h,t)\phi(t)\| = 0. \tag{1.14}$$

The notion of almost orbit was introduced by Miyadera-Kobayashi in [46]. This function is a kind of approximate solution to the differential inclusion $-\dot{x} \in Ax$. The authors used the Kobayashi's inequality in order to prove that the continuous path constructed by linear interpolations of some proximal iterations is an almost orbit of the semigroup generated by A. A converse result can be found in [47].

The following result from [2, Theorem 3.3] reveals the usefulness of the concept of almost-orbit.

Proposition 1.5 *Let U be a nonexpansive evolution system and let ϕ be an almost-orbit of U. If, for each $x \in H$ and $s \geq 0$, $U(t,s)x$ converges weakly (resp. strongly) as $t \to \infty$, then so does $\phi(t)$. The same holds if the word "converges" is replaced by "almost-converges" or "converges in average".*

We have the following result, which establishes a relationship between the trajectories generated by $U_{\mathcal{S}}$ and $U_{\mathcal{T}}$ and generalizes [63, Lemmas 4 & 6], [47, Proposition 2.3], [46, Proposition 7.4],[65, Proposition 8.6 i) & 8.7], [30, Theorem 3.1].

Theorem 1.6 *Let $\{\lambda_n\}_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1$, and fix $x \in H$. For each $t > 0$, define $\phi_{\mathcal{S}}(t) = \mathcal{S}_t x$ and $\phi_{\mathcal{T}}(t) = \mathcal{T}_{\nu(t)}x$ [1]. Then, $\phi_{\mathcal{S}}$ is an almost-orbit of $U_{\mathcal{T}}$, and $\phi_{\mathcal{T}}$ is an almost-orbit of $U_{\mathcal{S}}$.*

Combining Theorem 1.6 with Proposition 1.5, and using [74, Lemma 5.3], we obtain:

Theorem 1.7 *The following statements are equivalent:*

i) For every $z \in \overline{D(A)}$, $\mathcal{S}_t z$ converges strongly (weakly) as $t \to \infty$.

ii) For every initial point $x_0 \in H$, every sequence of step sizes $\{\lambda_n\}_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1$ and every sequence of errors $\{\varepsilon_n\}_{n\in\mathbb{N}}$ such that $\sum_{n=1}^{\infty} \|\varepsilon_n\| < +\infty$, the sequence $\{x_n\}_{n\in\mathbb{N}}$ generated by (1.5), converges strongly (weakly), as $n \to \infty$.

iii) There exist a sequence of step sizes $\{\lambda_n\}_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1$ such that, for every initial point $x_0 \in H$, the sequence $\{x_n\}_{n\in\mathbb{N}}$ generated by (1.3), converges strongly (weakly) as $n \to \infty$.

New convergence results for forward backward sequences on Hilbert spaces.

Recall A is maximally monotone and B is cocoercive. Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence representing computational errors and let $\{x_n\}_{n\in\mathbb{N}}$ satisfy (1.5). We assume that $\sum_{n=1}^{\infty} \|\varepsilon_n\| < \infty$. Finally, set $\mathcal{A} = A + B$ and $\Sigma = \mathcal{A}^{-1}(0)$, and assume that $\mathcal{A} \neq \emptyset$. We know that Σ is closed and convex, and the projection P_Σ is well defined, single-valued and continuous.

[1] This is a piecewise constant interpolation of the sequence $\mathcal{T}_n x$.

Theorem 1.8 *Let $\{\lambda_n\}_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1$. Assume one of the following conditions holds:*

i) There is $\alpha > 0$ such that for every $x \notin \Sigma$ and every $y \in \mathcal{A}(x)$,

$$\langle y, x - P_\Sigma(x)\rangle \geq \alpha \|x - P_\Sigma(x)\|^2$$

ii) J_1^A is compact and, for every every $y \in \mathcal{A}(x)$, $\langle y, x - P_\Sigma(x)\rangle \geq 0$; or

iii) The interior of Σ is not empty.

Then, $\{x_n\}_{n\in\mathbb{N}}$ converges strongly, as $n \to \infty$, to a point in Σ.

Asymptotic equivalence between continuous and discrete trajectories governed by a cocoercive operator.

Throughout this section we shall consider X a real Hilbert space and the *Euler sequence* $\{x_n\}_{n\in\mathbb{N}}$ described by (1.3) with $A = 0$. All the results from Theorem 1.1 to Theorem 1.6 remain valid in the particular case when $A = 0$. In what follows, we shall present some important consequences of this particular case. Now, for $n \in \mathbb{N}$ and $x \in H$, we denote

$$\mathcal{P}_n x = (I + \lambda_n B)^{-1} \circ \cdots \circ (I + \lambda_1 B)^{-1} x, \tag{1.15}$$

so that $\mathcal{P}_n x$ is the n-th term of the *proximal sequence* starting from x (see [67, 11, 65]).

As a consequence of Theorem 1.6 and Proposition 1.5, we obtain the following result on asymptotic equivalence between the continuous trajectories $\mathcal{S}_t x$ associated to the evolution equation governed by B ($A = 0$ in (1.8)), the corresponding Euler's sequence $\{x_n\}_{n\in\mathbb{N}}$ and the proximal sequence $\{\mathcal{P}_n x\}_{n\in\mathbb{N}}$ described in (1.15):

Theorem 1.9 *The following are equivalent:*

i) For every $x \in H$, $\mathcal{S}_t x$ converges weakly (resp. strongly), as $t \to +\infty$.

ii) For every $\{\lambda_n\}_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1$ and every $x \in H$, $\mathcal{P}_n x$ converges weakly (resp. strongly), as $n \to +\infty$.

iii) There is $\{\lambda_n\}_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1$ such that, for every $x \in H$, $\mathcal{P}_n x$ converges weakly (resp. strongly), as $n \to +\infty$.

iv) For every $\{\lambda_n\}_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1$ and every $x \in H$, the sequence $\{x_n\}_{n\in\mathbb{N}}$ described by (1.3) converges weakly (resp. strongly), as $n \to +\infty$.

v) There is $\{\lambda_n\}_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1$ such that, for every $x \in H$, the sequence $\{x_n\}_{n\in\mathbb{N}}$ described by (1.3) converges weakly (resp. strongly), as $n \to +\infty$.

The potential setting. Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a convex and differentiable function, with ∇f Lipschitz continuous with constant $L > 0$ and set $B = \nabla f$. Consider the *steepest descent* evolution equation

$$\begin{cases} -\dot{u}(t) &= \nabla f(u(t)), \quad t > 0 \\ u(0) &= x_0, \end{cases} \tag{1.16}$$

and its explicit discretization given by:

$$x_n = x_{n-1} - \lambda_n \nabla f(x_{n-1}), \quad n \in \mathbb{N} \tag{1.17}$$

with initial point $x_0 \in H$ and step sizes $\{\lambda_n\}_{n\in\mathbb{N}}$ satisfying $\sup\limits_{n\in\mathbb{N}} \lambda_n < 2/L$.

The step size assumption $\{\lambda_n\}_{n\in\mathbb{N}} \in \ell^2$, considered in Theorem 1.6 and Theorem 1.9 is removed later for operators deriving from a potential: Let $\mathcal{S}_t x$ be the solution of the evolution equation (1.16), with initial condition $x \in H$. The following result complements Theorem 1.9, establishing the equivalence between the convergence of trajectories/sequences generated by the steepest descent dynamics, the proximal point algorithm and the gradient method:

Theorem 1.10 *The following are equivalent:*

i) *For every $x \in H$, $\mathcal{S}_t x$ converges weakly (resp. strongly), as $t \to +\infty$.*

ii) *For every $\{\lambda_n\}_{n\in\mathbb{N}} \notin \ell^1$ and every $x \in H$, $\mathcal{P}_n x$ converges weakly (resp. strongly), as $n \to +\infty$.*

iii) *There is $\{\lambda_n\}_{n\in\mathbb{N}} \notin \ell^1$ such that, for every $x \in H$, $\mathcal{P}_n x$ converges weakly (resp. strongly), as $n \to +\infty$.*

iv) *For every $\{\lambda_n\}_{n\in\mathbb{N}} \notin \ell^1$ with $\sup_{n\in\mathbb{N}} \lambda_n < 2/L$, and every $x \in H$, the sequence $\{x_n\}_{n\in\mathbb{N}}$ described by (1.17) with initial point x, converges weakly (resp. strongly), as $n \to +\infty$.*

v) *There is $\{\lambda_n\}_{n\in\mathbb{N}} \notin \ell^1$ with $\sup_{n\in\mathbb{N}} \lambda_n < 2/L$ such that, for every $x \in H$, the sequence $\{x_n\}_{n\in\mathbb{N}}$ described by (1.17) with initial point x, converges weakly (resp. strongly), as $n \to +\infty$.*

Strong convergence of the gradient algorithm

Baillon's counterexample revisited

For $f : H \to \mathbb{R} \cup \{+\infty\}$ proper, lower-semicontinuous and convex, Bruck [12] proved that if f has minimizers and $u : [0, +\infty) \to H$ satisfies

$$\begin{cases} -\dot{u}(t) & \in & \partial f(u(t)), \text{ a.e.} \quad t > 0, \\ u(0) & \in & \overline{\partial f} \end{cases} \tag{1.18}$$

then $u(t)$ converges weakly, as $t \to +\infty$, to a minimizer of f. A few years later, Baillon [6] constructed a proper lower-semicontinuous convex function $f : \ell^2 \to \mathbb{R} \cup \{+\infty\}$ for which (1.18) has solutions that do not converge strongly. Baillon's function is not continuous, and its domain is not all of ℓ^2.

In [54], Martinet introduced the proximal point algorithm, and showed, for a constant sequence of step sizes $\lambda_n \equiv \lambda$, that $\{z_n\}$ converges weakly, as $n \to +\infty$, to a minimizer of f. This result was extended later by Rockafellar in [67] to the case where $\{\lambda_n\}_{n\in\mathbb{N}}$ is bounded from below by a positive number and to the case where $\{\lambda_n\}_{n\in\mathbb{N}} \notin \ell^1$ [11]. In [67], the author posed the question whether or not the convergence of proximal algorithm is always strong. Using Baillon's counterexample and the relationship between the asymptotic behavior of continuous and discretes trajectories presented by Passty in [63, Lemma 1], Güler [39] answered the question posed by Rockafellar, showing that sequences generated by the proximal point algorithm do not always converge strongly.

Using the theory of asymptotic equivalence, we present a family of *smooth* convex functions for which the steepest descent dynamics, the proximal point algorithm and the gradient method all produce trajectories/sequences that do not converge strongly. Related results have been found in [4] by a different (constructive) argument.

Theorem 1.11 *Let $\{\Lambda_n\}_{n\in\mathbb{N}} \notin \ell^1$ be a bounded sequence. There exist a convex function $f : H \to \mathbb{R}$, with Lipschitz continuous gradient, such that*

i) There is $u : [0, +\infty) \to H$ satisfying (1.18)*, that converges weakly but not strongly as $t \to +\infty$ to a minimizer of f.*

ii) There is a proximal sequence $\{z_n\}_{n\in\mathbb{N}}$, generated using step sizes $\{\Lambda_n\}_{n\in\mathbb{N}}$, that converges weakly but not strongly as $n \to +\infty$ to a minimizer of f.

iii) There is an Euler sequence $\{x_n\}_{n\in\mathbb{N}}$, generated using step sizes $\{\Lambda_n\}_{n\in\mathbb{N}}$, that converges weakly but not strongly as $n \to +\infty$ to a minimizer of f.

Work in progress 1: Motivated by the information obtained for the autonomous case in the section above and using the techniques presented in [1] related to asymptotic equivalence between non-autonomous systems governed by m-accretive operators, we are interested in extending the asymptotic equivalence results between the continuous and discrete evolution systems governed by the sum of a non-autonomous m-accretve operator with a cocoercive operator. Sor far, we have obtained a nonautonomous version of (1.7), which allows to estimate the distance between two iterates corresponding to independent sequences governed by two families of nonautonomous operators $\{A(t) + B\}_{t\geq 0}$ and $\{\widehat{A}(t) + B\}_{t\geq 0}$, where $A(t)$ and $\widehat{A}(t)$ are m-accretive operators for all $t \geq 0$ and B a cocoercive operator. This inequality could reveal us the relationship between the corresponding continuous and discrete systems in a finite and infinite horizon. However, the inequality has some terms expressed implicitly, which makes its implementation difficult. We plan to use combinatorial techniques to obtain estimates of the implicit terms in (1.7). More details can be found in Chapter 5.

1.2 Relaxed Forward–Backward Splitting and Primal–Dual Algorithms

The forward–backward iteration: Let H be a real Hilbert space and $\langle \cdot, \cdot \rangle$ its inner product, $A : H \to 2^H$ a maximally monotone operator and $B : H \to H$ a θ–cocoercive operator, for some real $\theta > 0$. The forward–backward algorithm, proposed by Mercier [55] and further developed by many authors [51, 37, 68, 18, 23, 25], allows to approach the solutions of the monotone inclusion

$$0 \in Az + Bz. \tag{1.19}$$

The *relaxed forward–backward* algorithm, which is described as follows: for $\{\rho_n\}_{n\in\mathbb{N}}$ a positive sequence of relaxation parameters, the iterations are given by:

Relaxed Forward–Backward iteration for (1.19): for $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} z_{n+\frac{1}{2}} = J_\gamma^A(z_n - \gamma B z_n) \\ z_{n+1} = z_n + \rho_n(z_{n+\frac{1}{2}} - z_n) \end{array} \right. . \tag{1.20}$$

While there is little interest in doing underrelaxation with ρ_n less than 1, it is expected that convergence is faster if doing overrelaxation with ρ_n larger than 1; this is what is most often observed in practice. The standard convergence for the forward–backward iteration is given by the following result from [27, Lemma 4.4] (see also [8, Theorem 26.14]).

Theorem 1.12 ***(Forward–Backward algorithm*** (1.20)*)* *Suppose that* $0 < \gamma < 2\theta$. *Let* $z_0 \in H$ *and set* $\delta = 2 - \gamma/(2\theta)$. *Suppose that* $\{\rho_n\}_{n\in\mathbb{N}}$ *is a sequence in* $[0, \delta]$ *such that* $\sum_{n\in\mathbb{N}} \rho_n(\delta - \rho_n) = +\infty$. *Then the sequence* $\{z_n\}_{n\in\mathbb{N}}$ *defined by the iteration* (1.20) *converges weakly to a solution of* (1.19).

Preconditioned Forward–Backward Algorithm: let P be a bounded, self-adjoint, strongly positive, linear operator on H. Clearly, solving (1.19) is equivalent to solving

$$0 \in P^{-1}Az + P^{-1}Bz. \tag{1.21}$$

Let H_P be the Hilbert space obtained by endowing H with the inner product $\langle x, x' \rangle_P = \langle x, Px' \rangle$, $(x, x' \in H)$. According with [8, Proposition 20.24], $P^{-1}A$ is maximally monotone in H_P. However, the cocoercivity of $P^{-1}B$ in H_P has to be checked on a case-by-case basis.

The *preconditioned* forward–backward iteration to solve (1.21) is

Preconditioned Forward–Backward iteration for (1.21) for $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} z_{n+\frac{1}{2}} = J_1^{P^{-1}A}(z_n - P^{-1}B z_n) \\ z_{n+1} = z_n + \rho_n(z_{n+\frac{1}{2}} - z_n) \end{array} \right. . \tag{1.22}$$

In the case when $P^{-1}B$ is a cocoercive operator, the corresponding convergence result follows:

Theorem 1.13 ***(Preconditioned Forward–Backward algorithm)*** *Suppose that* $P^{-1}B$ *is* χ*-cocoercive in* H_P, *with* $\chi > \frac{1}{2}$. *Set* $\delta = 2 - 1/(2\chi)$ *and let* $z_0 \in H$. *Suppose that* $\{\rho_n\}_{n\in\mathbb{N}}$ *is a sequence in* $[0, \delta]$ *such that* $\sum_{n\in\mathbb{N}} \rho_n(\delta - \rho_n) = +\infty$. *Then the sequence* $\{z_n\}_{n\in\mathbb{N}}$ *defined by the iteration* (1.22) *converges weakly to a solution of* (1.21).

If $B = 0$, the forward–backward iteration reduces to the proximal point algorithm [67, 8]. In that case, weak convergence to a zero of A is obtained with any $\gamma > 0$ and $\delta = 2$, in the notations of Theorem 1.13. Now, let P be a bounded, self-adjoint, strongly positive, linear operator on H. Let $z_0 \in H$ and let $\{\rho_n\}_{n\in\mathbb{N}}$ be a sequence of relaxation parameters. The relaxed and preconditioned proximal point algorithm is the iteration:

Preconditioned proximal point iteration for (1.21) **with B=0:** for $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} z_{n+\frac{1}{2}} = J_1^{P^{-1}A} z_n \\ z_{n+1} = z_n + \rho_n(z_{n+\frac{1}{2}} - z_n) \end{array} \right. . \tag{1.23}$$

The convergence of the preconditioned proximal point algorithm can be stated from Theorem 1.13 as follow:

Theorem 1.14 ***(Preconditioned proximal point algorithm*** (1.23)***)*** *Suppose that* $\{\rho_n\}_{n\in\mathbb{N}}$ *is a sequence in* $[0, 2]$ *such that* $\sum_{n\in\mathbb{N}} \rho_n(2 - \rho_n) = +\infty$. *Let* $z_0 \in H$. *Then the sequence* $\{z_n\}_{n\in\mathbb{N}}$ *defined by the iteration* (1.23), *converges weakly to a solution of* (1.21) *with* $B = 0$.

The case where B is affine

Let us suppose that, in addition to being θ–cocoercive, B is an affine operator. That means, $B : z \in H \mapsto Qz + c$, for some bounded, self-adjoint, positive, nonzero, linear operator Q on H and some element $c \in H$. In that case, the forward–backward iteration (1.20) can be interpreted as a preconditioned proximal point iteration (1.23) when $P = \frac{1}{\gamma} I - Q$, applied to find a zero of $A + B$. Since P must be strongly positive, we must have $0 < \gamma < \theta$, so that the admissible range for γ is halved. But in return, we get the larger range $[0, 2]$ for relaxation. As a concequence of Theorem 1.14, we obtain:

Theorem 1.15 ***(Forward–Backward algorithm, affine case).*** *Suppose that* $0 < \gamma < \theta$ *and that* $\{\rho_n\}_{n\in\mathbb{N}}$ *is a sequence in* $[0, 2]$ *such that* $\sum_{n\in\mathbb{N}} \rho_n(2 - \rho_n) = +\infty$. *Then the sequence* $\{z_n\}_{n\in\mathbb{N}}$ *defined by the iteration* (1.20) *converges weakly to a solution of* (1.19).

Remark 1.16 The affine case is considerably important because it will enable us to study the convergence of some important primal–dual algorithms in order to solve problems in image recovery, where the corresponding objective function has an additive estructure with the smooth term, which is generally a quadratic function. Moreover, Theorem 1.15 will enable us to obtain a wider range of relaxation parameters and, therefore, an accelerated version of these primal–dual algorithms. These algorithms will be presented as follow. More details can be found in the Chapter 5.

Applications to convex optimization

In what follow, we denote by $\Gamma_0(H)$ the set of convex, proper, lower semicontinuous functions from H to $\mathbb{R} \cup \{+\infty\}$ [8]. Let $f, h \in \Gamma_0(H)$ and suppose that h is differentiable function with β–Lipschitz continuous gradient ∇h, for some real $\beta > 0$. Let us consider the convex optimization problem

$$\underset{x\in H}{\text{minimize}} \; f(x) + h(x), \tag{1.24}$$

whose solution set is supposed nonempty. The well known Fermat's rule [8, Theorem 27.2] states that the problem (1.24) is equivalent to (1.19) with $A = \partial f$, which is maximally monotone, and $B = \nabla h$, which is θ-cocoercive, with $\theta = 1/\beta$ [8, Corollary 18.17]. Hence, it is natural to use the forward–backward iteration (1.20) to solve (1.24). Now, we shall focus on the case where h is quadratic:

$$h : x \mapsto \tfrac{1}{2}\langle x, Qx\rangle + \langle x, c\rangle, \tag{1.25}$$

for some bounded, self-adjoint, positive, nonzero, linear operator Q on H and some element $c \in H$. A very common example is a least-squares penalty, in particular to solve inverse problems, that is,

$$h : x \mapsto \tfrac{1}{2}\|Kx - y\|^2, \tag{1.26}$$

for some bounded linear operator K from H to a real Hilbert space Y and some element $y \in Y$. Clearly, (1.26) is an instance of (1.25) with $Q = K^*K$, where K^* is the adjoint of K, and $c = K^*y$. In this case, for every $x \in H$, we have

$$\nabla h(x) = Qx + c, \tag{1.27}$$

with $\beta = \|Q\|$. Setting $P = \frac{1}{\gamma}I - Q$, we can remark that the update in (1.22) can be written as

$$x_{n+\frac{1}{2}} = \underset{x \in H}{\arg\min}\, f(x) + h(x) + \tfrac{1}{2}\|x - x_n\|_P^2,$$

where we introduce the norm $\|\cdot\|_P : x \mapsto \sqrt{\langle x, Px\rangle}$. So, $x_{n+\frac{1}{2}}$ can be viewed as being obtained by applying the proximity operator of $f + h$ with the preconditioned norm $\|\cdot\|_P$. Hence, as a direct consequence of Theorem 1.15, we have:

Theorem 1.17 ***(Proximal–gradient algorithm, quadratic case)*** *Let $x_0 \in H$ and suppose $0 < \gamma < 1/\beta$. Suppose that $\{\rho_n\}_{n\in\mathbb{N}} \subset [0, 2]$ satisfying $\sum_{n\in\mathbb{N}} \rho_n(2 - \rho_n) = +\infty$. Then the sequence $\{x_n\}_{n\in\mathbb{N}}$ defined by the iteration (1.20) converges weakly to a solution of (1.24).*

The Loris–Verhoeven iteration

For H and U be two real Hilbert spaces, let $g \in \Gamma_0(U)$, $h : H \to \mathbb{R}$ be a convex and differentiable function with β–Lipschitz continuous gradient ∇h, for some real $\beta > 0$ and let $L : H \to U$ be a bounded linear operator. Often, the template problem (1.24) of minimizing the sum of two functions is too simple and we would like, instead, to

$$\underset{x \in H}{\text{minimize}}\ g(Lx) + h(x), \tag{1.28}$$

where the solution set is supposed nonempty.

The dual convex optimization problem associated to the primal problem (1.28):

$$\underset{u \in U}{\text{minimize}}\ g^*(u) + h^*(-L^*u). \tag{1.29}$$

Let $\tau > 0$ and $\sigma > 0$, let $x_0 \in H$ and $u_0 \in U$, and let $\{\rho_n\}_{n\in\mathbb{N}}$ be a sequence of relaxation parameters. The primal–dual forward–backward iteration, which we call the Loris–Verhoeven iteration is:

Loris–Verhoeven iteration for (1.28) and (1.29): for $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} u_{n+\frac{1}{2}} = \mathrm{prox}_{\sigma g^*}\Big(u_n + \sigma L\big(x_n - \tau\nabla h(x_n) - \tau L^* u_n\big)\Big) \\ x_{n+1} = x_n - \rho_n \tau\big(\nabla h(x_n) + L^* u_{n+\frac{1}{2}}\big) \\ u_{n+1} = u_n + \rho_n(u_{n+\frac{1}{2}} - u_n) \end{array} \right. . \tag{1.30}$$

This algorithm was first proposed by Loris and Verhoeven, in the case where h is a least-squares term [53]. It was then rediscovered several times and named *Primal–Dual Fixed-Point algorithm based on the Proximity Operator* (PDFP2O) [19] or *Proximal Alternating Predictor–Corrector* (PAPC) algorithm [34]. The above interpretation of the algorithm as a primal–dual forward–backward iteration has been presented in [22]. As an application of Theorem 1.13, we obtain the following convergence result for (1.30):

Theorem 1.18 ***(Loris–Verhoeven iteration for*** (1.30)***)*** *Suppose* $0 < \tau < 2/\beta$ *and* $\sigma\tau\|L\|^2 < 1$. *Set* $\delta = 2 - \tau\beta/2$. *Suppose that* $\{\rho_n\}_{n\in\mathbb{N}}$ *is a sequence in* $[0, \delta]$ *such that* $\sum_{n\in\mathbb{N}} \rho_n(\delta - \rho_n) = +\infty$. *Then the sequences* $\{x_n\}_{n\in\mathbb{N}}$ *and* $\{u_n\}_{n\in\mathbb{N}}$ *defined by the iteration* (1.30) *converge weakly to a solution of* (1.28) *and to a solution of* (1.29)*, respectively.*

The following result has been shown, which makes it possible to have $\sigma\tau\|L\|^2 = 1$ [19, Theorem 3.4 and Theorem 3.5]:

Theorem 1.19 ***(Loris–Verhoeven iteration for*** (1.30)***)*** *Suppose that H and U are of finite dimension. Suppose that* $0 < \tau < 2/\beta$, $\sigma\tau\|L\|^2 \leq 1$, *and* $\rho_n = 1$, $\forall n \in \mathbb{N}$. *Then the sequences* $\{x_n\}_{n\in\mathbb{N}}$ *and* $\{u_n\}_{n\in\mathbb{N}}$ *defined by the iteration* (1.30) *converge to a solution of* (1.28) *and to a solution of* (1.29)*, respectively.*

Again, let us focus on the case where h is quadratic; that is, $h : x \mapsto \frac{1}{2}\langle x, Qx\rangle + \langle x, c\rangle$, for some bounded, self-adjoint, positive, nonzero, linear operator Q on H and $c \in H$. We have $\beta = \|Q\|$. As an application of Theorem 1.15, we have:

Theorem 1.20 ***(Loris-Verhoeven algorithm*** (1.30)***, quadratic case)*** *Suppose that* $0 < \tau < \frac{1}{\beta}$, $\sigma\tau\|L\|^2 < 1$ *and* $\{\rho_n\}_{n\in\mathbb{N}}$ *is a sequence in* $[0, 2]$ *such that* $\sum_{n\in\mathbb{N}} \rho_n(2 - \rho_n) = +\infty$. *Then the sequences* $\{x_n\}_{n\in\mathbb{N}}$ *and* $\{u_n\}_{n\in\mathbb{N}}$ *defined by the iteration* (1.30) *converge weakly to a solution of* (1.28) *and to a solution of* (1.29)*, respectively.*

The Condat–Vũ iteration

Let us consider the primal optimization problem:

$$\underset{x\in H}{\mathrm{minimize}}\; f(x) + g(Lx) + h(x), \tag{1.31}$$

and its correspondig dual problem

$$\underset{u\in U}{\mathrm{minimize}}\, (f + h)^*(-L^* u) + g^*(u). \tag{1.32}$$

Thus, let $\tau > 0$ and $\sigma > 0$, let $x_0 \in H$ and $u_0 \in U$, and let $\{\rho_n\}_{n\in\mathbb{N}}$ be a sequence of relaxation parameters. The primal–dual forward–backward iteration, which we call the Condat–Vũ iteration is:

Condat–Vũ iteration form I for (1.31) **and** (1.32): for $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} x_{n+\frac{1}{2}} = \operatorname{prox}_{\tau f}\big(x_n - \tau\nabla h(x_n) - \tau L^* u_n\big) \\ u_{n+\frac{1}{2}} = \operatorname{prox}_{\sigma g^*}\big(u_n + \sigma L(2x_{n+\frac{1}{2}} - x_n)\big) \\ x_{n+1} = x_n + \rho_n(x_{n+\frac{1}{2}} - x_n) \\ u_{n+1} = u_n + \rho_n(u_{n+\frac{1}{2}} - u_n) \end{array} \right. . \tag{1.33}$$

This algorithm was proposed independently by the first author [27] and by B. C. Vũ [69]. An alternative is to update u before x, instead of the converse. The corresponding primal–dual forward–backward iteration in this case is given by:

Condat–Vũ iteration form II for (1.31) **and** (1.32): for $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} u_{n+\frac{1}{2}} = \operatorname{prox}_{\sigma g^*}\big(u_n + \sigma L x_n\big) \\ x_{n+\frac{1}{2}} = \operatorname{prox}_{\tau f}\big(x_n - \tau\nabla h(x_n) - \tau L^*(2u_{n+\frac{1}{2}} - u_n)\big) \\ u_{n+1} = u_n + \rho_n(u_{n+\frac{1}{2}} - u_n) \\ x_{n+1} = x_n + \rho_n(x_{n+\frac{1}{2}} - x_n) \end{array} \right. . \tag{1.34}$$

As an application of Theorem 1.13, we obtain the following result [27, Theorem 3.1]:

Theorem 1.21 ***(Condat–Vũ algorithm*** (1.33) ***and*** (1.34)***)*** *Suppose that* $\tau > 0$ *and* $\sigma > 0$ *satisfy* $\tau\big(\sigma\|L\|^2 + \beta/2\big) < 1$ *and consider* $\delta = 2 - (\beta/2)\left(\frac{1}{\tau} - \sigma\|L\|^2\right)^{-1} > 1$. *Suppose that* $\{\rho_n\}_{n\in\mathbb{N}}$ *is a sequence in* $[0, \delta]$ *such that* $\sum_{n\in\mathbb{N}} \rho_n(\delta - \rho_n) = +\infty$. *Then the sequences* $\{x_n\}_{n\in\mathbb{N}}$ *and* $\{u_n\}_{n\in\mathbb{N}}$ *defined by either the iteration* (1.33) *or the iteration* (1.34) *converge weakly to a solution of* (1.31) *and to a solution of* (1.32), *respectively.*

For the Condat–Vũ algorithm, let us focus on the case where h is a quadratic function; that is, $h : x \mapsto \frac{1}{2}\langle x, Qx\rangle + \langle x, c\rangle$, for some bounded, self-adjoint, positive, nonzero, linear operator Q on H and some element $c \in H$. We have $\beta = \|Q\|$. The convergence result for quadratic case can be obtained by Theorem 1.20:

Theorem 1.22 ***(Condat–Vũ algorithm*** (1.33) ***and*** (1.34)***, quadratic case)*** *Suppose that* $\tau\|Q + \sigma L^* L\| < 1$. *Let* $\{\rho_n\}_{n\in\mathbb{N}}$ *be a sequence in* $[0, 2]$ *such that* $\sum_{n\in\mathbb{N}} \rho_n(2 - \rho_n) = +\infty$. *Then the sequences* $\{x_n\}_{n\in\mathbb{N}}$ *and* $\{u_n\}_{n\in\mathbb{N}}$ *defined by either the iteration* (1.33) *or the iteration* (1.34) *converge weakly to a solution of* (1.31) *and to a solution of* (1.32), *respectively.*

Numerical experiment. Finally, some numerical experiments related to the implementation of these algorithms in image recovery will be presented. We make a comparisson between the obtained result from the numerical experiment of Loris–Verhoeven and Condat–Vũ algorithms.

Work in progress 2:

In this work, we are studying whether it is possible to transfer the convergence properties obtained for the Frank-Wolfe algorithm in [72] to the gradient algorithm, under the same hypothesis on the objetive function assumed in [72], instead of the usual Lipschitz condition. So far, we have obtained some estimations that could reveals information about the convergence rate of the gradient algorithm. More details can be found in Chapter 7.

Part I

Asymptotic equivalence between continuous and discrete dynamical systems

Foreword

The main goal of this part is to obtain a relationship of the asymptotic behavior between continuous and discrete dinamycal systems, with special emphasis on differential inclusions governed by the sum of a m-accretive operator with cocoercive operator and its corresponding discretization obtained by the forward backward algorithm. The strategy is to use the idea of asymptotic equality introduced by Passty [63], closely related to the notion of almost-orbit, introduced by Miyadera and Kobayasi [46]. However, developing this new theory is a complex task, so we have divided it into the following chapters:

Chapter 2, where we introduce the notation and preliminaires in the study of the asymptotic behavior of continuous and discretes dynamical governed by a m-accretive operator in Banach spaces.

Chapter 3, where we study discrete approximations of evolution equations governed by cocoercive operators by means of Euler iterations, both in a finite and an infinite time horizon. On the one hand, we give precise estimations for the distance between iterates of independently generated Euler sequences, and use them to obtain bounds for the distance between the state, given by the continuous-time trajectory, and the discrete approximation obtained by the Euler iterations. On the other hand, we establish the asymptotic equivalence between the continuous- and discrete-time systems, under a sharp hypothesis on the step sizes, which can be removed for operators deriving from a potential. As a consequence, we are able to construct a family of smooth functions for which the trajectories/sequences generated by basic first order methods converge weakly but not strongly, extending the counterexample of Baillon [6]. Finally, we include a few guidelines to address the problem in smooth Banach spaces.

Using the information obtained for the explicit case, the Chapter 4 will be dedicated to study the relationship of the asymptotic behavior between continuous and discrete trajectories governed by a sum of a m-accretive operator with a cocoercive operator, when the discrete trajectory is generated by means of forward backward iterations. Following the same estructure as the explicit case, we first give a precise estimations for the distance between two iterates corresponding to sequences independently generated by the forward backward algorithm, and use them to obtain bounds for the distance between the continuous–time trajectory and the discrete approximation in the Banach space. Thus, we establish the asymptotic equivalence between the continuous- and discrete-time systems, under summability condition on the step sizes. The asymptotic equivalence result will enable us to obtain new result on strong convergence of the forward backward algorithm in Banach spaces setting.

We finish this part by presenting the Chapter 5 with current work on the study of nonautonomous case, where our main goal is to extend the results provided in [1] to the sum of a nonautonomous maximally monotone operator with a coercive operator.

Chapter 2

Preliminaires

This chapter is dedicated to present the basic tools in the study of the asymptotic behavior of continuous and discretes dinamycal systems governed by a m-accretive operator in Banach and Hilbert spaces setting, which will be useful throughout this thesis. We start this chapter fixing notation and basic geometric properties in Banach space setting, which will be used to obtain basic properties for accretives and m-accretive operators in Banach spaces. On the other hand, we comment on existence results, uniqueness and qualitative properties of the solution for first order differential inclusion governed by an operator m-accretive in Banach spaces setting, with special emphasis on those results in the Hilbert spaces setting. We also present some results on the asymptotic behavior of the continuous and discrete trajectories (the last ones generated by implicit or explicit discretization of the continuous system) and finally, we present some results about the asymptotic equivalence between continuous and discretes trajectories associated to a first order evolution equation and how use it to get asymptotic properties of some optimization algorithms.

2.1 Basic results and notation

Let us fix some notation which will be used throughout the thesis. Let $(X, \|\cdot\|)$ a real Banach space and X^* its topological dual space. The *duality product* $\langle \cdot, \cdot \rangle_{X^*,X} : X^* \times X \to \mathbb{R}$ is defined by $\langle f, x \rangle_{X^*,X} = f(x)$ for all $x \in X$ and $f \in X^*$. Let us consider the dual space X^* endowed with the norm $\|f\|_* = \sup\limits_{\|x\| \leq 1} \langle f, x \rangle_{X^*,X}$ and the duality mapping $\mathcal{J} : X \to 2^{X^*}$ defined by

$$\mathcal{J}(x) = \{f \in X^* : \|f\|_* = \|x\| \text{ and } \langle f, x \rangle_{X^*,X} = \|x\|^2\}.$$

The following result shows an important geometric property in Banach space from [43, Lemma 1.1], which will be useful in order to get some properties of accretive operators in Banach spaces:

Proposition 2.1 *Let $x, y \in X$. $\|x\| \leq \|x + \lambda y\|$ for all $\lambda > 0$ if, and only if, there is $g \in \mathcal{J}(x)$ such that $\langle g, y\rangle_{X^*,X} \geq 0$.*

Proof. We may assume that $x \neq 0$, because in the case $x = 0$ the result holds inmediatly. Suppose that $\langle g, y\rangle_{X^*,X} \geq 0$ for some $g \in X^*$. Then for every $\lambda > 0$, we obtain that $\|x\|^2 = \langle g, x\rangle_{X^*,X} \leq \langle g, x + \lambda y\rangle_{X^*,X} \leq \|x + \lambda y\| \, \|x\|$. Conversely, let us take $h_\lambda \in \mathcal{J}(x + \lambda y)$ and set $f_\lambda = h_\lambda/\|h_\lambda\|$. Then

$$\|x\| \leq \|x + \lambda y\| = \langle f_\lambda, x + \lambda y\rangle_{X^*,X} = \langle f_\lambda, x\rangle_{X^*,X} + \lambda\langle f_\lambda, y\rangle_{X^*,X} \leq \|x\| + \lambda\langle f_\lambda, y\rangle_{X^*,X}.$$

Thus $\liminf_{\lambda\to 0}\langle f_\lambda, x\rangle_{X^*,X} \geq \|x\|$ and $\langle f_\lambda, y\rangle_{X^*,X} \geq 0$. Since the closed unit ball $\overline{B}_{X^*}(0,1)$ in X^* is compact for the weak* topology, there is a sequence $\{r_n\}$ of positive numbers such that $\lim_{n\to\infty} r_n = 0$ and $\{f_{r_n}\}$ converges to some $f \in \overline{B}_{X^*}(0,1)$ in the weak* toplogy as $n \to \infty$. Notice that f satisfies $\langle f, x\rangle_{X^*,X} \geq \|x\|$ and $\langle f, y\rangle_{X^*,X} \geq 0$. Then $\|f\| = 1$ and $\langle f, x\rangle_{X^*,X} = \|x\|$. Finally, taking $g = \|x\| f$ we obtain $g \in \mathcal{J}(x)$ and $\langle g, y\rangle_{X^*,X} \geq 0$. $\square$

For the purpose of this work, some Banach space will be interesting for its geometric and topological properties, namely *strictly convex*, *uniformly smooth* and *uniformly convex* Banach spaces [50, 7].

A Banach space X is *strictly convex* if the unit ball $B(0,1) \subset X$ is a strictly convex set. It means that for all $x, y \in X$ such that $\|x\| = \|y\| = 1$, $x \neq y$ and $\lambda \in (0,1)$, then $\|\lambda x + (1-\lambda)y\| < 1$. According with [7, Theorem 1.2], if the dual X^* of X is strictly convex, the duality mapping $\mathcal{J}$ is single–valued and if X is also reflexive, then $\mathcal{J}$ is continuous (strong-weak). On the other hand, the Banach space X is *uniformly convex* if for each $\varepsilon > 0$, there exist $\delta > 0$ such that for all $x, y \in X$ with $\|x\| \leq 1$, $\|y\| \leq 1$ and $\|x + y\| \geq 2 - \delta$, then $\|x - y\| < \varepsilon$. Every uniformly convex Banach space is strictly convex and by [50, Proposition 1.e.3], every uniformly convex (and thus also every uniformly smooth) Banach space X is reflexive.

Monotone and accretive operators in Banach spaces

Consider a set valued mapping $A : X \to 2^X$ and let us denote its *domain* and *graph* by $D(A) = \{x \in X : Ax \neq \emptyset\}$ and $G(A) = \{[u, v] : v \in Au\}$ respectively. In order to simplify the notation, we shall identify the operator A with its graph by writing $[u, v] \in A$ for $v \in Au$. A mapping $A : X \to 2^X$ is said *monotone operator* if for all $[u_1, v_1], [u_2, v_2] \in A$, there exist $f \in \mathcal{J}(u_1 - u_2)$ such that

$$\langle f, v_1 - v_2\rangle_{X^*,X} \geq 0.$$

A monotone operator is said to be *maximal* if its graph is not properly contained in the graph of the any other monotone operator. Let I the identity mapping on X. For $\lambda > 0$, the resolvent of A is defined as the mapping $J_\lambda^A = (I + \lambda A)^{-1}$.

An operator A is *accretive* if for all $\lambda > 0$, and $[u_1, v_1], [u_2, v_2] \in A$ one has

$$\|u_1 - u_2\| \leq \|u_1 - u_2 + \lambda(v_1 - v_2)\| \tag{2.1}$$

This implies that J_λ^A is a single-valued nonexpansive mapping. If, in addition, the range of $I + \lambda A$ is equals to X for all $\lambda > 0$, the operator A is said to be *m-accretive*. Notice that if the duality mapping $\mathcal{J}$ is single valued and A is *m-accretive*, then for each $x \in D(A)$, the set Ax is closed and convex. The following theorem summarizes some result in [56] and [43]:

Theorem 2.2 *Let $A : X \to 2^X$. Then:*

1. *A is monotone if, and only if, it is accretive;*
2. *If A is m-accretive, then it is maximal monotone.*
3. *If X is a Hilbert space and A is a maximally monotone operator, then it is m-accretive.*

PROOF. 1. This result holds from Proposition 2.1.

2. Let us consider $[x, x^*] \in X \times X$ such that $\langle f, x^* - y^* \rangle_{X^*,X} \geq 0$ for all $[y, y^*] \in A$ and some $f \in \mathcal{J}(x - y)$. Since $I + A$ is surjective, then there exist $[v, v^*] \in A$ such that $v + v^* = x + x^*$. Thus $\|x - v\|^2 \leq 0$ and then $x = v$. Finally we have $x^* = v + v^* - x = v*$ and $[x, x^*] \in A$.
3. Suppose that X is a Hilbert space with inner product $\langle \cdot, \cdot \rangle$. For this result will be suffices to prove that $I + A$ is surjective. Given $z_0 \in X$, we shall find $x_0 \in X$ such that $\langle y - (z_0 - x_0), x - x_0 \rangle \geq 0$ for $[x, y] \in G(A)$ and then, the maximality of A will imply that $z_0 - x_0 \in Ax_0$. In fact, consider the family of weakly compact sets $\{C_{x,y} : [x, y] \in A\}$ defined by $C_{x,y} = \{x_0 \in X : \langle y + x_0 - z_0, x - x_0 \rangle \geq 0\}$. It suffices to show that this family has the finite intersection property. Take $[x_\mathrm{i}, y_\mathrm{i}] \in G(A)$ for $\mathrm{i} = 1, 2, \cdots, n$. Let $\Delta = \{(\lambda_1, \lambda_2, \cdots, \lambda_n) : \lambda_\mathrm{i} \geq 0;\ \sum_{\mathrm{i}=1}^n \lambda_\mathrm{i} = 1\}$ be the n-dimensional simplex and consider the function $f : \Delta \times \Delta \to \mathbb{R}$ given by

$$f(\lambda, \mu) = \sum_{\mathrm{i}=1}^{n} \mu_\mathrm{i} \langle y_\mathrm{i} + x(\lambda) - z_0, x(\lambda) - x_\mathrm{i} \rangle,$$

with $x(\lambda) = \sum_{\mathrm{i}=1}^n \lambda_\mathrm{i} x_\mathrm{i}$. Notice that $f(\cdot, \mu)$ is a convex and continuous function while $f(\lambda, \cdot)$ is a linear function. The minimax theorem implies the existence of $\lambda_0 \in \Delta$ such that

$$\max_{\mu \in \Delta} f(\lambda_0, \mu) = \max_{\mu \in \Delta} \min_{\lambda \in \Delta} f(\lambda, \mu) \leq \max_{\mu \in \Delta} f(\mu, \mu).$$

Now, from monotonicity of A, we have

$$f(\mu, \mu) = \sum_{\mathrm{i}=1}^{n} \mu_\mathrm{i} \mu_j \langle y_\mathrm{i}, x_j - x_\mathrm{i} \rangle = \frac{1}{2} \sum_{\mathrm{i}=1}^{n} \langle y_\mathrm{i} - y_j, x_\mathrm{i} - x_j \rangle \leq 0$$

and thus $f(\lambda_0, \mu) \leq 0$ for all $\mu \in \Delta$. Now, for μ the canonical vectors we obtain that $\langle y_\mathrm{i} + x(\lambda_0) - z_0, x(\lambda_0) - x_\mathrm{i} \rangle \leq 0$ for all $\mathrm{i} = 1, \cdots, n$, that is to say $x(\lambda_0) \in C_{x_1,y_1} \cap \cdots \cap C_{x_n,y_n}$. $\square$

In general Banach spaces, the converse of second point of Theorem 2.2 does not hold. See [40] for a counterexample. From now on, $\Gamma_0(X)$ will denote the set of convex, lower-semicontinuous and proper functions on X.

Example 2.3 In Banach space setting, for example, if $A = I - T$, with T is a nonexpansive operator on X, then A is a maximally monotone operator and the solution set $\mathcal{S}$ is the set of fixed points of T. In Hilbert space, for $f \in \Gamma_0(X)$, ∂f is a maximal monotone operator and $\mathcal{S}$ is the set of minimizers of f.

For each $x \in D(A)$, we define $|||Ax||| = \inf\{\|y\| : y \in Ax\}$. The *minimal section* of A is the operator A^0 defined on $D(A)$ by $A^0x = \{y \in Ax : \|y\| = |||Ax|||\}$. Notice that A^0 accretive but not necessarily m-accretive. If both X and X^* are strictly convex and reflexive, then $A^0x = \text{Proj}_{Ax}0$ is a single–valued operator.

Remark 2.4 The operator A^0 is not always well-defined. This occurs when X is reflexive (see [7, Proposition 1.4]).

The *The Yosida approximation* of a m-accretive operator A is the single valued maximal montonone $A_\lambda = I - J_\lambda^A/\lambda$. Let us denote the *solution set* of A by $\mathcal{S} = A^{-1}(0)$. The following results summarize the main properties of A_λ and J_λ^A. The proof can be found in [7] (see [10] for Hilbert space setting):

Proposition 2.5 *Let A be m-accretive on X and $\lambda > 0$. We have*

1. $\left\|J_\lambda^A x - J_\lambda^A y\right\| \leq \|x - y\|$.
2. A_λ *is monotone and* $\frac{2}{\lambda}$*-Lipschitz.*
3. $A_\lambda x \in AJ_\lambda^A x$.
4. $\|A_\lambda x\| \leq |||Ax|||$.
5. $\lim\limits_{\lambda \to 0} J_\lambda^A x = x$.
6. *A is closed: $x_n \to x$,$y_n \to y$ and $[x_n, y_n] \in A$ together imply that $y \in Ax$.*

Special classes of monotone operators

Definition 2.6 *A mapping $B : X \to X$ is a cocoercive operator with parameter $\theta > 0$ if for all $[x, u], [y, v] \in B$, there exist $f \in \mathcal{J}(x - y)$ such that*

$$\langle f, u - v\rangle_{X^*,X} \geq \theta\|u - v\|^2.$$

These kind of operators are usually called *inverse strongly accretive* in Banach space setting. If X is a Hilbert space, the cocoercivity es given by

$$\langle Bx - By, x - y\rangle \geq \theta\|Bx - By\|^2.$$

Example 2.7 . Suppose that X is a Hilbert space, $D \subset X$ and $T : D \to X$ is a nonexpansive operator. Then $I - T$ is (1/2)-cocoercive operator [8, Corollary 4.11]. On the other hand, when $f\Gamma_0(X)$ is Gateaux differentiable, with ∇f Lipschitz continuous with constant $L > 0$, from Baillon-Haddad theorem [5] (see also [64, Theorem 3.13]) we have ∇f is cocoercive with constant $1/L$.

Definition 2.8 *Let* $\alpha > 0$. *A mapping* $A : X \to 2^X$ *is a* α-strongly monotone operator *if for all* $[x, x^*], [y, y^*] \in A$ *there is* $f \in \mathcal{J}(x - y)$ *such that*

$$\langle f, x^* - y^* \rangle_{X^*,X} \geq \alpha \|x - y\|^2.$$

It is easy to get from the definition that A is single-valued. In particular, the solution set $\mathcal{S}$ is, at most, a singleton. Observe that if A is α-strongly monotone, from Proposition 1.1, the mapping $J^A_{1/\alpha}$ is a strict contraction. Therefore it has a fixed point $\widehat{x}$ and only one, moreover $Fix(J^A_{1/\alpha}) = \mathcal{S}$. A classical example of α-strongly monotone operator is the subdifferential of proper, lower-semicontinuous α-strongly convex function (see [63, Proposition 3.23]).

2.2 Continuous and discrete dynamical systems governed by m-accretive operator

This section is dedicated to study some first order dynamical systems governed by a m-accretive operator. These systems have been studied extensively because they offer useful tools to solve and have several different applications, namely in optimization and fixed-point problems. We start with some existence and uniqueness result for the first order differential inclusion in Hilbert and Banach space setting. Later, we consider some discretizations of the differential inclusion that allows to get, on one part, solutions of the differential inclusion and, on the other part, some useful global estimations of the distance between two independent discretizations. We finish this section by showing some results concerning to the asymptotic behavior of the corresponding continuous and discrete dynamical systems. In what follow, A is an m-accretive operator in Banach space X.

2.2.1 The differential inclusion $\dot{x} \in -Ax$

Let $A : X \to 2^X$ be a m-accretive operator and $x_0 \in \overline{D(A)}$. Consider the following differential inclusion:

$$\left\{ \begin{array}{rcl} \dot{x}(t) & \in & -Ax(t) \;\; \text{a.e on } (0, \infty), \\ x(0) & = & x_0. \end{array} \right. \tag{2.2}$$

In what follow, we shall present some the existences and uniqueness results for the inclusion (2.2) in Hilbert and Banach space setting.

Classical existence result in Hilbert space setting: The problem of finding a trajectory satisfying (2.2) was first posed and studied in [48] and [31]. The classical proof can be found in [10].

Theorem 2.9 *There exist a unique absolutely continuous function* $x : [0, \infty) \to H$ *satisfying* (2.2). *Moreover, the solution satisfy:*

1. $\|\dot{x}(t)\| \leq \|A^0 x_0\|$ *almost everywhere on* $(0, \infty)$.

2. *$x(t) \in D(A)$ for all $t \geq 0$ and $\|A^0 x(t)\|$ decreases.*
3. *$A^0 x(t)$ is contiuous from the right and $x(t)$ admits a right derivative for all $t \geq 0$; namely $\dot{x}(t) = -A^0 x(t)$ (lazy behavior).*

The idea is to consider the differential inclusion (2.2) with $A = A_\lambda$ and using its properties described in Lemma 2.5 in order to prove the existence of solution x_λ. Then, one verifies that x_λ converges to some $x \in H$ satisfying (2.2) for the orignal A.

Semigroup of contractions: Let $S(t)x_0 = x(t)$ be the solution of (2.2), starting from x_0 and $t \geq 0$. S forms a strongly continuous semigroup of contractions on $D(A)$, namely:

1. $S(0) = I$;
2. $S(s)S(t) = S(s+t)$,
3. $\|S(t)x - S(t)y\| \leq \|x - y\|$;
4. $\lim\limits_{t \to 0} \|S(t)x - x\| = 0$.

Notice that the set of the fixed points of the semigroup coincides with the solution set $\mathcal{S}$ and that $\mathcal{S} \neq \emptyset$ if and only if, $\{S(t)x : t \geq 0\}$ is bounded for each $x \in D(A)$.

Let C be a closed and convex subset of X and S a semigroup of nonlinear contractions on C. Then S has a generator, which means that there exist a m-accretive operator A with domain $D(A)$ dense in C, such that for every $x \in D(A)$, $x(t) = S(t)x$ is the unique absolutely continuous solution of (2.2). Moreover (see [7]), one has

$$\lim_{t \to 0} \frac{x - S(t)x}{t} = A^0 x. \tag{2.3}$$

Conversely, every strongly continuous semigroup of contractions defines a maximal monotone operator via the limit formula (2.3).

Existence result in Banach spaces setting: The method developed in Hilbert space in order to get solution for (2.2) can be extended to Banach spaces X, when X and X^* are uniformly convex (see [43]). In general Banach spaces, existence and uniqueness can be also derived by the method in [32], via approximation of the trajectory by the discretization, as will be explained below.

2.2.2 Proximal point algorithm (PROX)

Let $\{\lambda_n\}$ be the sequence of positive numbers, which we call *step sizes* and $A : X \to 2^X$ a m-accretive operator. We shall say $\{x_n\}_{n \in \mathbb{N}}$ is a *proximal sequence* if it satisfies

$$\begin{cases} -\dfrac{x_n - x_{n-1}}{\lambda_n} \in Ax_n, & n \geq 1 \\ x_0 \in \overline{D(A)}. \end{cases} \tag{2.4}$$

Notice that $x_n = J^A_{\lambda_n}(x_{n-1})$ for all $n \in \mathbb{N}$. Since A is m-accretive, then $I + \lambda_n A$ is surjective, therefore the sequence $\{x_n\}_{n\in\mathbb{N}}$ given by (2.4) is well defined. The notion of proximal sequences and the term "proximal" were introduced in [57] for $A = \partial f$ in Hilbert space, where finding x_n corresponds to minimizing the *Moreau-Yosida approximation* of f:

$$f_{(\lambda_n, x_{n-1})}(x) := f(x) + \frac{1}{\lambda_n}\|x - x_{n-1}\|^2.$$

Therefore, from the Fermat's rule we have that each iteration in this case can be expressed by $x_n = J^{\partial f}_{\lambda_n}(x_{n-1})$.

2.2.3 A global estimation.

The following inequality was proposed by Kobayashi in [44], provides an estimation for the distance between two independent proximal sequences generated by (2.4). Original inequality was proposed by the author also accounts possible errors in the determination of the proximal sequence. Before to present the main result, let us consider the following technical result from [44, Lemma 1.1] for Banach space setting and [65, Lemma 18] for Hilbert space setting. From now on, $A : X \to 2^X$ is m-accretive, $\sigma_n = \sum_{k=1}^n \lambda_k$ and $\tau_n = \sum_{k=1}^n \lambda_k^2$.

Lemma 2.10 *Given $[u_1, v_1], [u_2, v_2] \in A$ and $\lambda, \widehat{\lambda} > 0$, we have*

$$(\lambda + \widehat{\lambda})\|u_1 - u_2\| \le \lambda\|u_2 + \widehat{\lambda}v_2 - u_1\| + \widehat{\lambda}\|u_1 + \lambda_1 v_1 - u_2\|.$$

The main result of this section comes from [44, Lemma 2.1], which was presented in Banach space setting and [65, Proposition 17] for Hilbert space setting:

Proposition 2.11 *Consider the sequences $\{x_n\}$ and $\{\widehat{x}_m\}$ generated by (2.4) with stepsizes $\{\lambda_n\}$ and $\{\widehat{\lambda}_m\}$ respectively. Then, for $u \in \overline{D(A)}$, we have:*

$$\|x_n - \widehat{x}_m\| \le \|x_0 - u\| + \|\widehat{x}_0 - u\| + \|\|Au\|\|\sqrt{(\sigma_n - \widehat{\sigma}_m)^2 + \tau_n + \widehat{\tau}_m}, \tag{2.5}$$

Proof. Let shall simplify the notation setting $c_{n,m} = \sqrt{(\sigma_n - \widehat{\sigma}_m)^2 + \tau_n + \widehat{\tau}_m}$. The proof will use induction on the pair (n, m). First, we shall establish inequality (2.5) for the pair $(n, 0)$ with $n \ge 0$. Since A is an accretive operator and from (2.1), for each $u \in \overline{D(A)}$ we have

$$\|x_n - u\| \le \|x_0 - u\| + \|\|Au\|\|\sigma_n.$$

Thus,

$$\|x_n - \widehat{x}_0\| \le \|x_n - u\| + \|\widehat{x}_0 - u\| \le \|x_0 - u\| + \|\widehat{x}_0 - u\| + \|\|Au\|\|\sigma_n \le \|x_0 - u\| + \|\widehat{x}_0 - u\| + \|\|Au\|\|c_{n,0},$$

because $\sigma_n \le \sqrt{\sigma_n^2 + \tau_n} = c_{n,0}$. In a similar fashion we obtain the inequality for $(0, m)$. Suppose that (2.5) holds for the pairs $(n-1, m)$ and $(n, m-1)$. From Lemma 2.10, we have

$$(\lambda_n + \widehat{\lambda}_m)\|x_n - \widehat{x}_m\| \le \lambda_n\|\widehat{x}_m + \widehat{\lambda}_m\widehat{y}_m - x_n\| + \widehat{\lambda}_m\|x_n + \lambda_n y_n - \widehat{x}_m\|.$$

Let us setting $\alpha_{n,m} = \frac{\widehat{\lambda}_m}{\lambda_n+\widehat{\lambda}_m}$ and $\beta_{n,m} = 1-\alpha_{n,m} = \frac{\lambda_n}{\lambda_n+\widehat{\lambda}_m}$. Then:

$$\begin{aligned}
\|x_n - \widehat{x}_m\| &\leq \alpha_{n,m}\|x_{n-1} - \widehat{x}_m\| + \beta_{n,m}\|\widehat{x}_{m-1} - x_n\| \\
&\leq (\alpha_{n,m} + \beta_{n,m})\left[\|x_0 - u\| + \|\widehat{x}_0 - u\|\right] + \left[\alpha_{n,m}c_{n-1,m} + \beta_{n,m}c_{n,m-1}\right] |||Au||| \\
&= \|x_0 - u\| + \|\widehat{x}_0 - u\| + \left[\alpha_{n,m}c_{n-1,m} + \beta_{n,m}c_{n,m-1}\right] |||Au|||. \qquad (2.6)
\end{aligned}$$

On the other hand, from Cauchy–Shwartz inequality, we have

$$\begin{aligned}
\alpha_{n,m}c_{n-1,m} + \beta_{n,m}c_{n,m-1} &= \alpha_{n,m}^{1/2}(\alpha_{n,m}^{1/2}c_{n-1,m}) + \beta_{n,m}^{1/2}(\beta_{n,m-1}^{1/2}c_{n,m-1}) \\
&\leq (\alpha_{n,m} + \beta_{n,m})^{1/2}(\alpha_{n,m}c_{n-1,m}^2 + \beta_{n,m}c_{n,m-1}^2)^{1/2} \\
&= (\alpha_{n,m}c_{n-1,m}^2 + \beta_{n,m}c_{n,m-1}^2)^{1/2}.
\end{aligned}$$

Notice that $c_{n-1,m}^2 = c_{n,m}^2 - 2\lambda_n(\sigma_n - \widehat{\sigma}_m)$ and $c_{n,m-1}^2 = c_{n,m}^2 + 2\widehat{\lambda}_m(\sigma_n - \widehat{\sigma_m})$. Hence

$$\begin{aligned}
(\alpha_{n,m}c_{n-1,m} + \beta_{n,m}c_{n,m-1})^2 &\leq \alpha_{n,m}c_{n-1,m}^2 + \beta_{n,m}c_{n,m-1}^2 \\
&= c_{n,m}^2(\alpha_{n,m} + \beta_{n,m}) - 2(\alpha_{n,m}\lambda_n - \beta_{n,m}\widehat{\lambda}_m)(\sigma_n - \widehat{\sigma}_m) \\
&= c_{n,m}^2. \qquad (2.7)
\end{aligned}$$

From (2.6) and (2.7) we obtain (2.5). □

Existence and uniqueness of solutions: Set $t \in [0,T]$, $m \in \mathbb{N}$ and run the proximal algorithm (2.4) with constant stepsizes $\lambda_n \equiv t/m$. Let us denote the m-th iteration by $x_m(t) = \left[J^A_{\frac{t}{m}}\right]^m(x_0)$. The following result comes from [32, Theorem 2.1 and Theorem 2.2]:

Proposition 2.12 *The sequence $\{x_m(t)\}$ defined above converges uniformly on every compact interval $[0,T]$ for each $T > 0$ to a function $x(t)$, which is Lipschitz continuous and satisfies $\dot{x}(t) \in -Ax(t)$ almost everywhere on $[0,\infty]$.*

Proof. In this case we provide an easier proof using the Kobayashi (2.5). Fix $N, M \in \mathbb{N}$ and $s, t \in [0,T]$ with $T > 0$. Set $\lambda_n \equiv t/N$ for all n and $\widehat{\lambda}_m \equiv s/M$ for all m. Initialize $x_0(t)$ and $\widehat{x}_0(s)$ both at $u = x_0$. For $n = N$ and $m = M$ we have

$$\|x_N(t) - x_M(s)\| \leq |||Ax_0|||\sqrt{(t-s)^2 + \frac{T}{N} + \frac{T}{M}},$$

where we conclude that the sequence $\{x_n\}$ converges uniformly on $[0,T]$ to some function x, which is uniformly Lipschitz–continuous with constant $|||Ax_0|||$.

In order to prove that the function x satisfy (2.2), it suffices to verify that it is an integral solution in the Benilan's sense [9, Proposition 2.5], it means that for all $[u,v] \in A$ and $t > s \geq 0$, we have

$$\|x(t) - u\|^2 - \|x(s) - u\|^2 \leq 2\int_s^t \langle \mathcal{J}(u - x(\tau)), v\rangle_{X^*,X} \mathrm{d}\tau. \qquad (2.8)$$

Notice that if $\{x_n\}$ is a sequence generated by (2.4) with step sizes $\{\lambda_n\}$, $[u,v]\in A$.

$$\begin{aligned}
2\langle \mathcal{J}(x_n-u), x_n-x_{n-1}\rangle_{X^*,X} &= 2\langle \mathcal{J}(x_n-u), x_n-u\rangle_{X^*,X} - 2\langle \mathcal{J}(x_n-u), x_{n-1}-u\rangle_{X^*,X}\\
&\geq 2\|x_n-u\|^2 - \|x_{n-1}-u\|^2 - \|x_n-u\|^2\\
&\geq \|x_n-u\|^2 - \|x_{n-1}-u\|^2.
\end{aligned}$$

Thus

$$\|x_n-u\|^2 - \|x_{n-1}-u\|^2 \leq 2\lambda_n\langle \mathcal{J}(x_n-u), -v\rangle_{X^*,X}. \tag{2.9}$$

Now, summing up for $n=1,\cdots,N$ in (2.9), we obtain

$$\|x_N-u\|^2 - \|x_0-u\|^2 \leq 2\sum_{n=1}^{N}\lambda_n\langle \mathcal{J}(x_n-u), -v\rangle_{X^*,X}. \tag{2.10}$$

Setting $x_0 = x(s)$ and passing to the limit appropiately we obtain (2.8). Finally, $x(t)\in D(A)$ by maximality. □

Remark 2.13 Propositions 2.11 and 2.12 allow to get the proximal sequence approaches the continuous–time trajectory in a bounded temporal horizon. It means that for each $T>0$ and $s,t\in[0,T]$ we have:

(i) $\|x_n - x(t)\| \leq \|x_0-u\| + \|x(0)-u\| + \|\|Au\|\|\sqrt{(\sigma_n-t)^2+T\sigma_n}$.

(ii) For trajectories x and z we have

$$\|z(s)-x(t)\| \leq \|z(0)-u\| + \|x(0)-u\| + \|\|Au\|\|\ |s-t|.$$

(iii) In particular, the solution x of (2.2) is unique and

$$\|x(s)-x(t)\| \leq \|\|Ax(0)\|\|\ |s-t|.$$

2.2.4 Euler's discretization

Now, assume that A maps $D(A)$ into itself. Let $\{\lambda_n\}$ be a sequence of numbers in $(0,1]$. Define an *Euler sequence* $\{z_n\}$ recursively by

$$\left\{\begin{array}{rcl} -\dfrac{z_n-z_{n-1}}{\lambda_n} &\in& Az_{n-1},\quad n\geq 1\\ x_0 &\in& D(A). \end{array}\right. \tag{2.11}$$

Notice that the terms of the sequence can be computed explicitely, which take advantage for the implementation. Let us denote $w_{n+1} = z_n - z_{n+1}/\lambda_{n+1}$ the *velocity* of the system (2.11).

When $A = I-T$, for T a nonexpansive operator and $\lambda_n \equiv 1$, then $z_n = T^n z_0$. This particular case has been studied extensively by several authors in the search for fixed points of T. Notice also that in this framework, there is a Kobayashi-type inequality proposed by Vigeral [70], namely:

$$\|z_n - \widehat{z}_m\| \leq \|z_0-u\| + \|\widehat{z}_0-u\| + \|u-Tu\|\sqrt{(\sigma_n-\widehat{\sigma}_m)^2+\tau_n+\widehat{\tau}_m},$$

where u is any point in X.

The following result comes from [70] and establish a relationship between continuous trajectories of (2.2) and continuous generated by (2.11) in Banach space setting (see also [65] for Hilbert space setting):

Proposition 2.14 ***(Chernoff's Estimate).*** *If T is nonexpansive and v is a trajectorie such that*

$$z'(t) = (-1/\lambda)(I - T)z(t)$$

with $z(0) = z_0$, then

$$\|z(t) - T^n z_0\| \leq \|z'(0)\| \sqrt{\lambda t + (n\lambda - t)^2}.$$

Since $A = (-1/\lambda)(I - T)$ is a monotone operator, one can use Kobayashi's inequality to get (see [65])

$$\|z(t) - T^n(z_0)\| \leq \|Az_0\| \frac{t}{\sqrt{n}}.$$

This inequality is knowed as Chernoff's estimate with $\lambda = \dfrac{t}{n}$.

2.3 Asymptotic behavior

In this section we shall comment some results concerning to the asymptotic behavior of the system described in (2.2), (2.4) and (2.11). We start with the study of the asymptotic behavior of some discretes dynamical sytems in Hilbert space setting, namely Proximal and Euler scheme described in (2.4) and (2.11) respectively. Later, we continue with the study of the asymptotic behavior of the continuous trajectories of (2.2) and finish this sections with the study of longterm relationship between continuous trajectory of (2.2) and discrete trajectory given in (2.4).

The following result of [61] (see also [64, Lemma 5.2]) is a very useful tool for proving weak convergence of a sequence without any information about the limit. Here, It will be presented in Hilbert space:

Lemma 2.15 *(Opial's Lemma). Let $\{x_n\}$ be a sequence in a Hilbert space X and le $F \subset X$. Assume $\|x_n - f\|$ has a limit as $n \to \infty$ for each $f \in F$ and that every weak cluster point of $\{x_n\}$ lies in F. Then $\{x_n\}$ converges weakly to some $x^* \in F$.*

2.3.1 Continuous dynamical systems

In this section we shall study the main results concerning to the asymptotic behavior of continuous trajectories associated to the differential inclusion:

$$\begin{cases} \dot{x}(t) & \in & -Ax(t) \text{ a.e on } (0,\infty), \\ x(0) & = & x_0. \end{cases} \tag{2.12}$$

In general case, maximally monotone operators not always generates continuous trajectories weakly convergent. For example, consider the clockwise $\pi/2$-rotation operator $T_{\frac{\pi}{2}} : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T_{\frac{\pi}{2}}(x,y) = (-y,x)$. $T_{\frac{\pi}{2}}$ is a maximally monotone operator but the trajectories associated to the first order differential inclusion governed by this operator are not weakly convergent (see [65, Section 6]). However we have the following result from [12] (see also [65]):

Theorem 2.16 *Let $A = \partial f$, with $f \in \Gamma_0(X)$ such that $\mathcal{S} = argmin(f) \neq \emptyset$ and let $x : [0, \infty) \to X$ be the function given by Theorem 1.10. Then $x(t)$ converges weakly as $t \to \infty$ to some $\overline{x} \in argmin(f)$.*

Proof. From [65, Proposition 25], the function $t \longmapsto f(x(t))$ is nonincreasing and from subgradient inequality we have $f(u) \geq f(x(t)) + \langle \dot{x}(t), x(t) - u \rangle$ for all $u \in X$. In particular, for $u \in argmin(f)$ we get

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|x(t) - u\|^2 = \langle \dot{x}(t), x(t) - u \rangle \leq f(x(t)) - f^* \leq 0.$$

Thus $t \longmapsto \|x(t) - u\|$ is decreasing. On the other hand, let us consider $\{t_n\}$ a sequence of positive numbers such that $t_n \to \infty$ as $n \to \infty$ and suppose that $\{x(t_n)\}$ conveges weakly to some $\overline{x} \in X$. Since f is weak lower-semicontinuous, we have

$$f(\overline{x}) \leq \liminf_{n\to\infty} f(x(t_n)) = \inf\{f(u) : u \in X\},$$

and then $\overline{x} \in argmin(f)$. From Lemma 2.15 we have that $\{x(t)\}$ converges weakly to $\overline{x} \in argmin(f)$. □

Some comments about strong convergence: Let us consider $A = \partial f$ with $f \in \Gamma_0(X)$ having a minimizers and X a Hilbert space. The continuous trajectorie $x(t)$ need not converge strongly as $t \to \infty$. This is shown by Baillon's celebrated counterexample in [6, Proposition 1], where the author defines a function $\varphi \in \Gamma_0(\ell^2)$ having minimizers and proves that the continuous trajectories converge weakly but not strongly. However, it is possible to get strong convergen when the operator A is α-strongly monotone ([65, Proposition 59]) and $\mathring{\mathcal{S}} \neq \emptyset$ ([65, Proposition 60]). These results on strong convergence hold in Banach space setting, when X and X^* are uniformly convex (see [59]).

2.3.2 The proximal point algorithm (PROX)

In what follow, $\{x_n\}_{n\in\mathbb{N}}$ will be considered proximal sequence defined by (2.4) and X will be a Hilbert space.

Weak convergence

The study of the asymptotic behavior of the (2.4) was proposed by Rockafellar in [67], when the stepsizes are bounded away from zero. Later, Brezis and Lion in [11] obtained convergence results on (2.4) under more general hypothesis, as we will show below (see also [65]):

Theorem 2.17 *Let A be a maximally monotone operator, $\mathcal{S} \neq \emptyset$ and $\{x_n\}$ generated by (2.4) with step sizes $\{\lambda_n\}$. If $\{\lambda_n\} \notin \ell^2$ then $\{x_n\}$ converges weakly to some $x^* \in \mathcal{S}$.*

Strong convergence

This subsection will be dedicated to study the strong convergence of (2.4). The original results comes from [11], and a simplified version can be found in [65] for maximally monotone operators in Hilbert space. The first assert can be found in [64, Proposition 6.7]:

Proposition 2.18 . *Let A a maximally monotone operator, $\mathcal{S} \neq \emptyset$ and $\{x_n\}$ the sequence generated by* (2.4)*, with step size $\{\lambda_n\}$.*

1. *If A is the subdifferential of an even, convex and lower-semicontinuous function f, then $\{x_n\}$ converges strongly as $n \to \infty$.*

2. *If A is α–strongly monotone for some $\alpha > 0$, the $\{x_n\}$ converges strongly to the unique $x^* \in \mathcal{S}$ as $n \to \infty$.*

3. *If $\mathring{\mathcal{S}} \neq \emptyset$, then $\{x_n\}$ converges strongly as $n \to \infty$.*

Remark 2.19 In the case when X and X^* are uniformly convex, A is α-strongly monotone operator and $\{\lambda_n\} \notin \ell^1$, then $\{x_n\}$ converges strongly as $n \to \infty$ (see [59]).

Euler's discretization

The Euler's sequence $\{z_n\}_{n\in\mathbb{N}}$ defined by (2.11) can be computed explicitely, which take advantage for the implementation. However, the convergence of this algorithm depends of an adequate selection of the stepsizes, as shown in the following example:

Example 2.20 Let $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$. In this case $\nabla f(x) = 2x$ and straigt-forward computation show that each iteration for the gradient algorithm with constant step sizes $\lambda > 0$ is given by $z_{n+1} = z_n - \lambda \nabla f(z_n) = z_n - 2\lambda z_n = (1-2\lambda)^n z_0$, for $z_0 \in \mathbb{R}$ fixed. Let us set $z_0 = 1$. If $\lambda = 1$, we have $\{z_n\}$ diverges. Now, for $\lambda > 1$, the sequence $\{z_n\}$ is divergent but, when we set $\lambda < 1$, we obtain $\lim\limits_{n\to\infty} z_n = 0$.

The weak convergence result of Euler's sequence $\{z_n\}$ given by (2.11) was presented [13] in Banach space, but its version in Hilbert space can be found in [65]. In both cases, we need strong aditional conditions on the velocity sequence $\{w_n\}$ associated to the system and the stepsizes $\{\lambda_n\}$. The next following result from [65] in Hilbert space setting:

Proposition 2.21 *Let A a maximally montone operator with $\mathcal{S} \neq \emptyset$, $\{\lambda_n\} \in \ell^2 \setminus \ell^1$ and $\{z_n\}$ defined by* (2.11)*. If the velocity $w_{n+1} = z_n - z_{n+1}/\lambda_{n+1}$ is bounded, then $\{z_n\}$ converges weakly to a zero of A.*

On the strong convergence for Euler's sequences in Hilbert space, we have the following result from [65]:

Proposition 2.22 *Let A be a α-strongly operator, $\mathcal{S} \neq \emptyset$, $\{\lambda_n\} \in \ell^2 \setminus \ell^1$. If the velocity sequence $\{w_n\}$ is bounded, then the Euler's sequences $\{z_n\}$ given by* (2.11) *converges strongly.*

A remarkable particular case is the particular case when $A = \nabla f$. If f is strongly convex, even or $\mathring{\mathcal{S}} \neq \emptyset$, it is possible to get strong convergence, as shown in [64, Proposition 6.21].

2.4 Asymptotic equivalence between continuous and discrete dynamical systems

Definition 2.23 *Let C a convex subset of a Banach space X and let I the identity operator on X. A nonexpansive evolution system on X is a family $\big(U(t,s)\big)_{0\le s\le t}$ such that*

(i) $U(t,t)z = z$ *for all* $z \in X$ *and* $t \geq 0$.

(ii) $U(t,s)U(s,r)z = U(t,r)z$ *for all* $z \in X$ *and all* $t \geq s \geq r \geq 0$.

(iii) $\|U(t,s)x - U(t,s)y\| \leq \|x-y\|$ *for all* $x, y \in X$ *and* $t \geq s \geq 0$.

Example 2.24 Consider A a maximally monotone operator on the Hilbert space X. According with Proposition 2.5, J^A_λ is nonexpansive for all $\lambda > 0$. For $0 \leq s \leq t$, set

$$U(t,s) = \prod_{\mathrm{i}=\nu(s)+1}^{\nu(t)} J^A_{\lambda_\mathrm{i}}, \tag{2.13}$$

where $\nu(t) = \max\{n \in \mathbb{N} : \sigma_n \leq t\}$ and the products denotes composition of functions (in reverse order) and the empty composition is the identity. Then $\big(U(t,s)\big)_{0\le s\le t}$ is a nonexpansive evolution system on X.

Asymptotic equivalence. The notion of asymptotic equivalence between continuous trajectories associated to (2.12) and its discrete trajectories given by (2.4) was introduced by Passty in [63] for m-accretive operators in Banach spaces setting. It means that the continuous trajectories (2.12) and the discrete ones (2.4) describe the same asymptotic behavior. In order to prove it, Passty first introduced the notion of *asymptotically equality* between a nonexpansive evolution system $(U(t,s))_{0\le s\le t}$ and the semigroup $\{S(t)x : t \geq 0\}$ governed by $-A$, which means:

$$\lim_{t\to\infty} \sup_{h\geq 0} \|U(t+h,s)x - S(h)U(t,s)x\| = 0 \quad \text{for all} \quad s \geq 0, \tag{2.14}$$

and

$$\lim_{t\to\infty} \sup_{h\geq 0} \|U(t+h,t)\mathcal{S}(t)x - \mathcal{S}(t+h)x\| = 0. \tag{2.15}$$

for all $x \in \overline{D(A)}$. The following result from [63, Lemma 1] reveals the importance of the asymptotic equality:

Lemma 2.25 *Let A be an m-accretive operator and let $\{S(t)x : t \geq 0\}$ be the semigroup generated by $-A$ on $\overline{D(A)}$. Let $\big(U(t,s)\big)_{0\le s\le t}$ be a nonexpansive evolution system which is asymptotically equal to $S(t)$ on $D(A)$. Then $S(t)x$ converges strongly (respectively diverges) as $t \to +\infty$ if and only if $U(t,s)x$ converges strongly (respectively weakly) as $t \to +\infty$ for all $x \in \overline{D(A)}$.*

Note that the proximal sequence $\{x_n\}_{n\in\mathbb{N}}$ can be described by mean the nonexpansive evolution system (2.13) as $x_n = U(0, \sigma_n)$. Passty proved that $\{x_n\}_{n\in\mathbb{N}}$ is asymptotically equals to $\{S(t)x : t \geq 0\}$. Thus, using Lemma 2.25 for A m-accretive and Lipschitz and for each $x \in D(A)$, the continuous trajectory $\{S(t)x : t \geq 0\}$ associated to the differential inclusion (2.2) converges strongly (respectively weakly) when $t \to \infty$ if and only if every sequence $\{x_n\}$ given by (2.4) with stepsizes $\{\lambda_n\}_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1$ converges strongly (respectively weakly) when $n \to \infty$ (see [63, Theorem 2]). In other words, $\{S(t)x : t \geq 0\}$ and $\{x_n\}_{n\in\mathbb{N}}$ are asymptotically. Later, Sugimoto-Koizumi in [47, Theorem] were able to get rid of Passty's condition on the operator A and obtained an analogue result on asymptotic equivalence between continuous trajectory associated to (2.12) and discrete trajectories generated by (2.4). The key that allows to get is is the Kobayashi's inequality 2.11 (see also[44, Lemma 2.1] for the original reference).

Definition 2.26 *Let U be a nonexpansive evolution system. A function $\phi : [0, +\infty) \to X$ is an almost-orbit of U if*

$$\lim_{t\to+\infty} \sup_{h\geq 0} \|\phi(t+h) - U(t+h,t)\phi(t)\| = 0. \tag{2.16}$$

The notion of almost orbit was introduced by Miyadera-Kobayashi in [46]. This function is a kind of approximate solution to the differential inclusion $-\dot{x} \in Ax$. The authors used the Kobayashi's inequality in order to prove that the continuous path constructed by linear interpolations of some proximal iterations is an almost orbit of the semigroup generated by A. A converse result can be found in [47].

The following result from [2, Theorem 3.3] reveals the usefulness of the concept of almost-orbit. It keeps some essential ideas of [63, Theorem 2], but in a more general context.

Proposition 2.27 *Let U be a nonexpansive evolution system and let ϕ be an almost-orbit of U. If, for each $x \in H$ and $s \geq 0$, $U(t,s)x$ converges weakly (resp. strongly) as $t \to +\infty$, then so does $\phi(t)$.*

Thus, from the notion of almost-orbit and Proposition 2.27 is easier to obtain the asymptotic equality described in (2.14) and (2.15). Therefore, the notion of asymptotic equivalence can be reformulated in terms of almost-orbits.

The notion asymptotic equivalence is a useful tool that allows to exploit the information of the continuous system represented for the continuous trajectory $\{S(t)x : t \geq 0\}$, in order to get information about the asymptotic behavior of the discrete trajectory given by (2.4), such as is showed in the following examples:

Example 2.28 Suppose that X is a Hilbert space, $f \in \Gamma_0(X)$ and $A = \partial f$. If $\{x_n\}$ is given by (2.4) with stepsizes $\{\lambda_n\} \notin \ell^1$ and bounded, then for each $x \in D(A)$, $\{S(t)x : t \geq 0\}$ converges weakly (resp. strongly) if and only if $\{x_n\}$ converges weakly (resp. strongly). This result was proved in [39, Theorem 5.1]. As an important consequence of this result, the author proved that $\{x_n\}$ not always converges strongly (see [39, Corollary 5.1]).

Example 2.29 Let us consider X a Banach, $T : X \to X$ a nonexpansive operator, $A = I - T$ and given $z \in X$, let $\{z_n\}$ be the sequence generated by (2.11). Then, the continuous trajectories $\{S(t)z : t \geq 0\}$ associated to (2.2) converges weakly (resp. strongly) when $t \to \infty$ if and only if $\{z_n\}$ converges weakly (resp. strongly) when $n \to \infty$. It was proved in [47, Theorem 1.1].

So far, only asymptotic equivalence between continuous trajectories associated to a differential inclusion gorverned by maximally monotone operator and the sequences generated by PROX is known. Our main goal in Chapters 3 and 4 is to extend this results to the case when the monotone operator has an additive structure.

Chapter 3

Asymptotic Equivalence of Evolution Equations Governed by Cocoercive Operators and their Forward Discretizations

A joint work with Peypouquet, J. [30]

3.1 Euler sequences governed by cocoercive operators

Throughout this paper, H will denote a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$, and B will denote a *cocoercive operator*, which is a function $B : H \to H$ such that

$$\langle Bx - By, x - y \rangle \geq \theta \, \|Bx - By\|^2$$

for all $x, y \in H$ and some $\theta > 0$, known as the *cocoercivity constant* of B. It follows that B is maximally monotone (see [65, 8]). Also, by the Cauchy-Schwarz inequality, B is Lipschitz continuous with constant $\frac{1}{\theta}$.

Given an initial point $x_0 \in H$, along with a sequence $\{\lambda_n\}_{n \in \mathbb{N}}$ of positive numbers, called *step sizes*, we define the *Euler sequence* $\{x_n\}$ recursively by

$$x_n = x_{n-1} - \lambda_n B x_{n-1}, \quad n \geq 1. \tag{3.1}$$

To simplify the notation, we denote $E_\lambda = I - \lambda B$ for $\lambda > 0$, so that the equality in (3.1) becomes

$$x_n = E_{\lambda_n} x_{n-1}.$$

Lemma 3.1 *If $\lambda \leq 2\theta$, then E_λ is nonexpansive.*

PROOF. For all $x, y \in H$ and $\lambda > 0$, we have

$$\begin{array}{rcl} \|E_\lambda x - E_\lambda y\|^2 & = & \|x - y - \lambda(Bx - By)\|^2 \\ & \leq & \|x - y\|^2 + \lambda^2\|Bx - By\|^2 - 2\lambda\langle Bx - By, x - y\rangle \\ & = & \|x - y\|^2 + (\lambda^2 - 2\lambda\theta)\|Bx - By\|^2 \end{array}$$

by cocoercivity. If $\lambda \leq 2\theta$, then $\|E_\lambda x - E_\lambda y\| \leq \|x - y\|$, as claimed. □

The rule given by (3.1) defines a discrete-time dynamical system, which is *dissipative* in the following sense:

Lemma 3.2 *Let $\{x_n\}_{n\in\mathbb{N}}$ satisfy (3.1) with $\lambda_n \leq 2\theta$ for all $n \in \mathbb{N}$. The sequence $\{\|Bx_n\|\}$ is nonincreasing.*

PROOF. Consider the velocity at stage n, namely

$$y_n = -Bx_n = \frac{x_{n+1} - x_n}{\lambda_{n+1}}.$$

Since B is cocoercive, using (3.1), we obtain

$$\langle y_n - y_{n+1}, \lambda_{n+1}y_n\rangle = \langle Bx_{n+1} - Bx_n, x_{n+1} - x_n\rangle \geq \theta\|y_{n+1} - y_n\|^2.$$

Thus

$$\theta\|y_{n+1}\|^2 + \theta\|y_n\|^2 - \lambda_{n+1}\|y_n\|^2 \leq (2\theta - \lambda_{n+1})\langle y_{n+1}, y_n\rangle \leq \left(\frac{2\theta - \lambda_{n+1}}{2}\right)\left(\|y_{n+1}\|^2 + \|y_n\|^2\right).$$

Rearranging the terms, we obtain $\|y_{n+1}\| \leq \|y_n\|$. □

The following result establishes the weak convergence of Euler sequences. Although it is a well-known fact (see, for instance, [8]), we include it for the sake of completeness:

Proposition 3.3 *Let*

$$\{x_n\}_{n\in\mathbb{N}}$$

satisfy (3.1) with $\{\lambda_n\}_{n\in\mathbb{N}} \notin \ell^1$ and $\lambda := \sup_{n\in\mathbb{N}} \lambda_n < 2\theta$. If $B^{-1}(0) \neq \emptyset$, then x_n converges weakly, as $n \to +\infty$, to a point in $B^{-1}(0)$.

PROOF. Let $p \in B^{-1}(0)$. Using the cocoercivity and rearranging the terms, we obtain

$$\left(\frac{2\theta}{\lambda_{n+1}} - 1\right)\|x_{n+1} - x_n\|^2 + \|x_{n+1} - p\|^2 \leq \|x_n - p\|^2. \tag{3.2}$$

Since $\lambda_{n+1} \leq \lambda < 2\theta$, this implies $\|x_{n+1} - p\| \leq \|x_n - p\|$ and $\sum_{n=0}^{\infty}\|x_{n+1} - x_n\|^2 < +\infty$. Using (3.2) again, we deduce that

$$\sum_{n=0}^{\infty}\lambda_{n+1}\|Bx_n\|^2 = \sum_{n=0}^{\infty}\frac{\|x_{n+1} - x_n\|^2}{\lambda_{n+1}} < +\infty.$$

Since $(\lambda_n) \notin \ell^1$, Lemma 3.2 gives $\lim_{n\to+\infty} \|Bx_n\| = 0$. The maximal monotonicity implies that every weak limit point of the sequence

$$\{x_n\}_{n\in\mathbb{N}}$$

must belong to $B^{-1}(0)$, and the result follows from Opial's Lemma (see, for instance, [65, Lemma 4.1]). □

3.1.1 Discrete approximation of the evolution equation

The following Kobayashi-type inequality provides an estimation of the distance between the terms of two Euler sequences.

Lemma 3.4 *Let $\{x_n\}_{k\in\mathbb{N}}$, $\{\hat{x}_l\}_{l\in\mathbb{N}}$ be two sequences generated by* (3.1)*, with step sizes $\{\lambda_k\}_{k\in\mathbb{N}}$ and $\{\hat{\lambda}_l\}_{l\in\mathbb{N}}$, respectively. Assume $\lambda_k, \hat{\lambda}_l \leq 2\theta$ for all $k, l \in \mathbb{N}$, and fix $z \in H$. Then, for each $k, l \in \mathbb{N}$, we have*

$$\|x_k - \hat{x}_l\| \leq \|x_0 - z\| + \|\hat{x}_0 - z\| + \|Bz\|\sqrt{(\sigma_k - \hat{\sigma}_l)^2 + \tau_k + \hat{\tau}_l}, \tag{3.3}$$

where $\sigma_k = \sum_{\mathrm{i}=1}^{k} \lambda_{\mathrm{i}}$ and $\tau_k = \sum_{\mathrm{i}=1}^{k} \lambda_{\mathrm{i}}^2$ (similarly for $\hat{\sigma}_l$ and $\hat{\tau}_l$).

Proof. First, observe that

$$E_\lambda = I - \lambda B = I - \left[\frac{\lambda}{2\theta}\right](2\theta B) = I - \left[\frac{\lambda}{2\theta}\right](I - E_{2\theta}),$$

and recall that $E_{2\theta}$ is nonexpansive. Using [70, Proposition 3.11], we deduce that

$$\|x_k - \hat{x}_l\| \leq \|x_0 - z\| + \|\hat{x}_0 - z\| + \|2\theta Bz\|\sqrt{\left(\frac{\sigma_k}{2\theta} - \frac{\hat{\sigma}_l}{2\theta}\right)^2 + \frac{\tau_k}{4\theta^2} + \frac{\hat{\tau}_l}{4\theta^2}},$$

which is precisely (3.3). □

3.2 Approximation in finite horizon

A curious fact about Lemma 3.4 is that it provides existence and regularity results for the evolution equation

$$\begin{cases} -\dot{u}(t) &= B(u(t)), \quad t > 0, \\ u(0) &= x_0, \end{cases} \tag{3.4}$$

by means of an approximation scheme, which is different from that of Picard. For each $t \geq 0$ and $m \geq 1$, set

$$u_m(t) = \left[E_{\frac{t}{m}}\right]^m x_0. \tag{3.5}$$

In other words, $u_m(t)$ is the m-th term of the Euler sequence generated from x_0 using the constant step size $\lambda_k \equiv t/m$. We have the following:

Proposition 3.5 *The sequence $\{u_m\}_{m\in\mathbb{N}}$ converges pointwise on $[0,+\infty)$ and uniformly on $[0,T]$ for each $T>0$, to a function $u:[0,+\infty)\to H$ satisfying (3.4). Moreover, $\|u(t)-u(s)\|\le\|Bx_0\|\,|t-s|$ for all $t,s>0$.*

Proof. Given $t,s>0$ and $n,m\in\mathbb{N}$, define $u_m(t)$ and $u_n(s)$ as above. By Lemma 3.4, we have

$$\|u_m(t)-u_n(s)\|\le\|Bx_0\|\sqrt{(t-s)^2+\frac{t^2}{m}+\frac{s^2}{n}}. \tag{3.6}$$

For $s=t$, this gives

$$\|u_m(t)-u_n(t)\|\le t\,\|Bx_0\|\sqrt{\frac{1}{m}+\frac{1}{n}}.$$

It follows that (u_m) converges pointwise on $[0,+\infty)$, and uniformly on $[0,T]$ for each $T>0$, to a function $u:[0,+\infty)\to H$. Passing to the limit in (3.6), we obtain $\|u(t)-u(s)\|\le\|Bx_0\|\,|t-s|$ for all $t,s>0$. It remains to prove that u satisfies (3.4). First, take $s,t>0$, along with a sequence $\{n_m\}_{m\in\mathbb{N}}$ of positive integers such that $\lim_{m\to+\infty}\frac{n_m}{m}=\frac{s}{t}$, and use Lemma 3.4 to obtain

$$\left\|\left[E_{\frac{t}{m}}\right]^{n_m}x_0-\left[E_{\frac{s}{m}}\right]^{m}x_0\right\|\le\|Bx_0\|\sqrt{\left(\frac{n_m t}{m}-s\right)^2+\frac{n_m t^2}{m^2}+\frac{s^2}{m^2}}.$$

It follows that

$$\lim_{m\to+\infty}\left[E_{\frac{t}{m}}\right]^{n_m}x_0=u(s).$$

On the other hand, take $m\in\mathbb{N}$, set $\lambda_n\equiv\frac{t}{m}$, iterate (3.1) and sum for $n=0,\dots,m-1$, to obtain

$$\left[E_{\frac{t}{m}}\right]^{m}x_0-x_0=-\sum_{n=0}^{m-1}\frac{t}{m}B\left(\left[E_{\frac{t}{m}}\right]^{n}x_0\right).$$

The right-hand side is an approximate Riemann sum for $Bu(\cdot)$ on $[0,t]$. Passing to the limit, we deduce that

$$u(t)-x_0=-\int_0^t Bu(s)\,\mathrm{d}s,$$

and conclude that u satisfies (3.4). □

The Lipschitz constant provided in Proposition 3.5 also implies

Corollary 3.6 *The function $t\mapsto\|Bu(t)\|$ is nonincreasing.*

Remark 3.7 From (3.6), we also obtain

$$\|u_m(t)-u(t)\|\le\frac{t}{\sqrt{m}}\|Bx_0\|.$$

This estimation is linear on the length of the interval, which is consistent with the results obtained for backward discretizations (see, for instance, [32, p. 272]). In turn, the bound for forward discretizations, obtained following the standard argument (see, for instance, [14, Chapter II.1]), is exponential on the length of the interval.

From Lemma 3.4, we also obtain the following:

Proposition 3.8 *Let $(\hat{\lambda}_l)$ be bounded by 2θ. Then, for each $l \in \mathbb{N}$, we have*

$$\|\hat{x}_l - u(t)\| \leq \|x_0 - \hat{x}_0\| + \min\{\|Bx_0\|, \|B\hat{x}_0\|\}\sqrt{(\hat{\sigma}_l - t)^2 + \hat{\tau}_l}.$$

3.3 Approximation in infinite horizon I: general case

In this section, we show that Euler sequences have the same asymptotic behavior as the solutions of the evolution equation (3.4). The key argument is the idea of asymptotic equality introduced by Passty [63], closely related to the notion of almost-orbit, introduced by Miyadera and Kobayasi [46]. Further commentaries on this topic can be found in [1, 2, 3].

To simplify the notation, given $x \in H$ and $t \geq 0$, we write

$$\mathcal{S}_t x = u(t), \tag{3.7}$$

where u satisfies (3.4) with $x_0 = x$. Also, for $0 \leq s \leq t$, we write

$$U_{\mathcal{S}}(t,s) = \mathcal{S}(t-s). \tag{3.8}$$

In a similar fashion, if $n \in \mathbb{N}$ and $x \in H$, we denote

$$\mathcal{E}_n x = E_{\lambda_n} \circ \cdots \circ E_{\lambda_1} x. \tag{3.9}$$

In other words, $\mathcal{E}_n x$ is the n-th term of the Euler sequence starting from x. Assume $(\lambda_n) \notin \ell^1$, and write $\nu(t) = \max\{n \in \mathbb{N} : \sigma_n \leq t\}$. For $0 \leq s \leq t$, we set

$$U_{\mathcal{E}}(t,s) = \prod_{\mathrm{i}=\nu(s)+1}^{\nu(t)} E_{\lambda_{\mathrm{i}}}, \tag{3.10}$$

where the product denotes composition of functions (in reverse order) and the empty composition is the identity.

Example 3.9 The families $\big(U_{\mathcal{S}}\big)$ and $\big(U_{\mathcal{E}}\big)$, defined in (3.8) and (3.10), respectively, are nonexpansive evolution systems. Actually, the same is true if $\mathcal{S}$ is replaced by any other semigroup of nonexpansive functions on H, and if each $E_{\lambda_{\mathrm{i}}}$ is replaced by any other nonexpansive function on H.

The following result is part of [2, Theorem 3.3]:

Proposition 3.10 *Let U be a nonexpansive evolution system and let ϕ be an almost-orbit of U. If, for each $x \in H$ and $s \geq 0$, $U(t,s)x$ converges weakly (resp. strongly) as $t \to +\infty$, then so does $\phi(t)$.*

The following result establishes a relationship between the trajectories generated by $U_{\mathcal{S}}$ and $U_{\mathcal{E}}$:

Theorem 3.11 *Let $(\lambda_n) \in \ell^2 \setminus \ell^1$, and fix $x \in H$. For each $t > 0$, define $\phi_{\mathcal{S}}(t) = \mathcal{S}_t x$ and $\phi_{\mathcal{E}}(t) = \mathcal{E}_{\nu(t)} x$[1]. Then, $\phi_{\mathcal{S}}$ is an almost-orbit of $U_{\mathcal{E}}$ and $\phi_{\mathcal{E}}$ is an almost-orbit of $U_{\mathcal{S}}$.*

Proof. We first prove that $\phi_{\mathcal{S}}$ is an almost-orbit of $U_{\mathcal{E}}$. By Lemma 3.4, we have

$$\begin{aligned}
\left\| \left[\prod_{k=1}^{m} E_{\frac{h}{m}} \right] \mathcal{S}_t x - \left[\prod_{\mathrm{i}=\nu(t)+1}^{\nu(t+h)} E_{\lambda_{\mathrm{i}}} \right] \mathcal{S}_t x \right\| &\leq \|B\mathcal{S}_t x\| \sqrt{\left(\sigma_{\nu(t)+1}^{\nu(t+h)} - h\right)^2 + \tau_{\nu(t)+1}^{\nu(t+h)} + \frac{h^2}{m}} \\
&\leq \|Bx\| \sqrt{4\rho^2(t) + \tau_{\nu(t)+1}^{\infty} + \frac{h^2}{m}},
\end{aligned}$$

where $\rho(t) := \sup\{\lambda_n : n \geq \nu(t) - 1\}$ vanishes as $t \to +\infty$. Passing to the limit as $m \to +\infty$, we obtain

$$\|\mathcal{S}_h \mathcal{S}_t x - U_{\mathcal{E}}(t+h,t)\mathcal{S}_t x\| \leq \|Bx\| \sqrt{4\rho^2(t) + \tau_{\nu(t)+1}^{\infty}},$$

which is uniform in $h \geq 0$ (we can also arrive here directly using Proposition 3.8). It follows that

$$\lim_{t \to +\infty} \sup_{h \geq 0} \|\phi_{\mathcal{S}}(t+h) - U_{\mathcal{E}}(t+h,t)\phi_{\mathcal{S}}(t)\| = 0.$$

To prove that $\phi_{\mathcal{E}}$ is an almost-orbit of $U_{\mathcal{S}}$, we proceed in a similar fashion, using Lemma 3.4 to obtain

$$\left\| \prod_{\mathrm{i}=\nu(t)+1}^{\nu(t+h)} E_{\lambda_{\mathrm{i}}} \mathcal{E}_{\nu(t)} x - \prod_{k=1}^{m} E_{\frac{h}{m}} \mathcal{E}_{\nu(t)} x \right\| \leq \|Bx\| \sqrt{4\rho^2(t) + \tau_{\nu(t)+1}^{\infty} + \frac{h^2}{m}},$$

then passing to the limit as $m \to +\infty$ to deduce that

$$\|\phi_{\mathcal{E}}(t+h) - \mathcal{S}_h \phi_{\mathcal{E}}(t)\| \leq \|Bx\| \sqrt{4\rho^2(t) + \tau_{\nu(t)+1}^{\infty}},$$

and conclude. $\square$

If $n \in \mathbb{N}$ and $x \in H$, we denote

$$\mathcal{P}_n x = (I + \lambda_n B)^{-1} \circ \cdots \circ (I + \lambda_1 B)^{-1} x, \tag{3.11}$$

so that $\mathcal{P}_n x$ is the n-th term of the *proximal sequence* starting from x (see [67, 11, 65]).

As a consequence of Theorem 3.11, we obtain the following result, which extends [63, Theorem 2] and [47, Theorem], to include Euler sequences:

Theorem 3.12 *The following are equivalent:*

i) For every $x \in H$, $\mathcal{S}_t x$ converges weakly (resp. strongly), as $t \to +\infty$.

ii) For every $\{\lambda_n\}_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1$ and every $x \in H$, $\mathcal{P}_n x$ converges weakly (resp. strongly), as $n \to +\infty$.

[1] This is a piecewise constant interpolation of the sequence $\mathcal{E}_n x$.

iii) *There is* $\{\lambda_n\}_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1$ *such that, for every* $x \in H$, $\mathcal{P}_n x$ *converges weakly (resp. strongly), as* $n \to +\infty$.

iv) *For every* $\{\lambda_n\}_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1$ *and every* $x \in H$, $\mathcal{E}_n x$ *converges weakly (resp. strongly), as* $n \to +\infty$.

v) *There is* $\{\lambda_n\}_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1$ *such that, for every* $x \in H$, $\mathcal{E}_n x$ *converges weakly (resp. strongly), as* $n \to +\infty$.

In the next section, we shall prove that the ℓ^2 condition on the step sizes can be dropped if B is the gradient of a convex function. This fact will be crucial in finding a smooth version of Baillon's and Güler's counterexamples for strong convergence (see [39, 6]).

3.4 Approximation in infinite horizon II: the potential setting

Let $f : H \to \mathbb{R}$ be convex and differentiable. According to the Baillon-Haddad Theorem (see, for instance, [64, Theorem 3.13]), ∇f is Lipschitz continuous with constant L if, and only if, it is cocoercive with constant $\theta = L^{-1}$. We shall see that, in this case, the asymptotic equality between the semigroup and the evolution system can be obtained without the condition $(\lambda_n) \in \ell^2$. This will allow us to find a smooth convex function for which the gradient method, the proximal point algorithm and the steepest descent dynamic fail to produce strongly convergent trajectories.

3.4.1 Preliminary estimations

Given $x_0 \in H$ and a sequence $\{\lambda_n\}_{n\in\mathbb{N}}$ of step sizes, consider the (Euler) sequence $\{x_n\}_{n\in\mathbb{N}}$ generated by the gradient method:

$$x_{n+1} = x_n - \lambda_{n+1}\nabla f(x_n), \quad n \geq 1. \tag{3.12}$$

Proposition 3.13 *Let* $\lambda = \sup_{n\in\mathbb{N}} \lambda_n < 2/L$*, and set* $\alpha = 1 - \frac{L\lambda}{2}$*. For all* $n \in \mathbb{N}$ *and* $y \in H$*, we have*

$$\begin{aligned}
f(x_{n+1}) - f(y) &\leq \frac{\|y-x_0\|^2}{2\sigma_{n+1}} - \frac{\|y-x_{n+1}\|^2}{2\sigma_{n+1}} - \frac{\alpha\sigma_{n+1}}{2}\|\nabla f(x_n)\|^2 + \frac{\lambda L V_n}{4\sigma_{n+1}} && (3.13)\\
\|\nabla f(x_n)\|^2 &\leq \frac{\|x_{n+1}-x_0\|^2}{\alpha\sigma_{n+1}^2} + \frac{\lambda L V_n}{2\alpha\sigma_{n+1}^2} && (3.14)\\
\|\nabla f(x_n)\|^2 &\leq \frac{\|y-x_0\|^2}{\alpha\sigma_{n+1}^2} + \frac{\|\nabla f(y)\|^2}{\alpha} + \frac{\lambda L V_n}{2\alpha\sigma_{n+1}^2}, && (3.15)
\end{aligned}$$

where $V_n = \sum_{k=0}^{n} \|x_{k+1} - x_k\|^2$*. Moreover, if* f *is bounded from below, then* $\sup_{n\in\mathbb{N}} V_n \leq \frac{2}{\alpha L}\big(f(x_0) - \inf(f)\big)$.

Proof. Using the Descent Lemma (see, for instance [64, Lemma 1.30]), for each $k \in \mathbb{N}$, we have

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leq \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2 \\ &= \left[\frac{\lambda_{k+1}L - 2}{2}\right] \lambda_{k+1}\|\nabla f(x_k)\|^2 \\ &\leq -\alpha\lambda_{k+1}\|\nabla f(x_k)\|^2. \end{aligned} \tag{3.16}$$

Multiplying by $2\sigma_{k+1}$, we obtain

$$2\sigma_{k+1}f(x_{k+1}) - 2\sigma_k f(x_k) - 2\lambda_{k+1}f(x_k) \leq -2\alpha\sigma_{k+1}\lambda_{k+1}\|\nabla f(x_k)\|^2.$$

Summing for $k = 0, \ldots, n$, we deduce that

$$2\sigma_{n+1}f(x_{n+1}) - 2\sum_{k=0}^{n}\lambda_{k+1}f(x_k) \leq -2\alpha\sum_{k=0}^{n}\sigma_{k+1}\lambda_{k+1}\|\nabla f(x_k)\|^2. \tag{3.17}$$

On the other hand, for each $y \in H$, we have

$$f(y) \geq f(x_k) + \langle \nabla f(x_k), y - x_k \rangle = f(x_k) + \frac{1}{\lambda_{k+1}}\langle x_k - x_{k+1}, y - x_k \rangle.$$

In other words,

$$2\lambda_{k+1}(f(y) - f(x_k)) \geq \|y - x_{k+1}\|^2 - \|x_k - x_{k+1}\|^2 - \|y - x_k\|^2.$$

Summing for $k = 0, \ldots, n$, and reversing the inequality, we obtain

$$-2\sigma_{n+1}f(y) + 2\sum_{k=0}^{n}\lambda_{k+1}f(x_k) \leq \|y - x_0\|^2 - \|y - x_{n+1}\|^2 + V_n. \tag{3.18}$$

Adding (3.17) and (3.18), we have

$$\begin{aligned} 2\sigma_{n+1}(f(x_{n+1}) - f(y)) &\leq \|y - x_0\|^2 - \|y - x_{n+1}\|^2 + V_n - 2\alpha\sum_{k=0}^{n}\sigma_{k+1}\lambda_{k+1}\|\nabla f(x_k)\|^2 \\ &= \|y - x_0\|^2 - \|y - x_{n+1}\|^2 + (1-\alpha)V_n \\ &- \alpha\sum_{k=0}^{n}(2\sigma_{k+1}\lambda_{k+1} - \lambda_{k+1}^2)\|\nabla f(x_k)\|^2 \\ &= \|y - x_0\|^2 - \|y - x_{n+1}\|^2 + (1-\alpha)V_n \\ &- \alpha\sum_{k=0}^{n}(\sigma_{k+1}\lambda_{k+1} + \sigma_k\lambda_{k+1})\|\nabla f(x_k)\|^2 \\ &\leq \|y - x_0\|^2 - \|y - x_{n+1}\|^2 + (1-\alpha)V_n - \alpha\sigma_{n+1}^2\|\nabla f(x_n)\|^2, \end{aligned} \tag{3.19}$$

since $\{\|\nabla f(x_k)\|\}$ is nonincreasing, by Lemma 3.2, and

$$\sum_{k=0}^{n}(\sigma_{k+1}\lambda_{k+1} + \sigma_k\lambda_{k+1}) = \sigma_{n+1}^2.$$

Inequality (3.19) is exactly (3.13). Writing $y = x_{n+1}$ in (3.13) gives (3.14). Finally, to obtain (3.15), use the gradient inequality

$$f(x_{n+1}) - f(y) \geq \langle \nabla f(y), x_{n+1} - y \rangle$$

and the Cauchy-Schwarz inequality to deduce that

$$2\sigma_{n+1}(f(y) - f(x_{n+1})) \leq \sigma_{n+1}^2 \|\nabla f(y)\|^2 + \|y - x_{n+1}\|^2.$$

Adding this inequality to (3.19), and rearranging the terms, we obtain (3.15). For the last statement, notice that

$$V_n = \sum_{k=0}^{n} \|x_{k+1} - x_k\|^2 = \sum_{k=0}^{n} \lambda_{k+1}^2 \|\nabla f(x_k)\|^2 \leq \frac{2}{\alpha L} \sum_{k=0}^{n} \big(f(x_k) - f(x_{k+1}) \big) \leq \frac{2}{\alpha L} \big(f(x_0) - \inf(f) \big),$$

for all $n \in \mathbb{N}$, in view of (3.16). □

Now, consider the *steepest descent* evolution equation:

$$\begin{cases} -\dot{u}(t) &= \nabla f(u(t)), \quad t > 0 \\ u(0) &= x_0, \end{cases} \tag{3.20}$$

which is a particular case of (3.4). Taking $\lambda_k \equiv \frac{t}{n}$ and passing to the limit in (3.14) and (3.15), we obtain (see [39, Corollary 2.2] and [10, Théorème 2.3.2]) the following:

Proposition 3.14 *Let $u : [0, +\infty) \to H$ satisfy* (3.20). *For each $t > 0$ and each $y \in H$, we have*

$$\|\nabla f(u(t))\| \leq \frac{\|u(t) - x_0\|}{t} \tag{3.21}$$

$$\|\nabla f(u(t))\| \leq \|\nabla f(y)\| + \frac{\|y - x_0\|}{t}. \tag{3.22}$$

3.4.2 Preservation of the asymptotic behavior

Consider a strictly increasing function $\kappa : \mathbb{N} \to \mathbb{N}$ such that $\kappa(0) = 0$ (its relevance will become apparent later on). Given a sequence $\{\lambda_j\}_{j \in \mathbb{N}}$ of positive real numbers, set

$$\sigma_m^n = \sigma_{\kappa(n)} - \sigma_{\kappa(m)} = \sum_{j=\kappa(m)+1}^{\kappa(n)} \lambda_j, \quad \text{whenever } m < n. \tag{3.23}$$

Analogously, given a sequence (J_j) of functions from H to itself, and a point $x \in H$, write

$$\prod_m^n x = \prod_{j=\kappa(m)+1}^{\kappa(n)} J_j x, \quad \text{if } m < n, \tag{3.24}$$

where the last product denotes composition of functions in the appropriate order. If $n \leq m$, we define $\sigma_m^n = 0$ and $\prod_m^n x = x$. Although the following result is stated in a more general context, the proof is exactly the same as that of [39, Lemma 5.2].

Lemma 3.15 *Let $(R(t))$ be a nonexpansive semigroup and let $\{J_j\}_{j\in\mathbb{N}}$ be a sequence of nonexpansive functions. With the notation introduced above, for every $n, p \in \mathbb{N}$ and every $x \in H$, we have*

$$\left\| R(\sigma_n^{n+p})x - \prod_n^{n+p} x \right\| \leq \sum_{m=n+1}^{n+p} \left\| R(\sigma_{m-1}^m) \prod_n^{m-1} x - \prod_{m-1}^{m} \prod_n^{m-1} x \right\| \tag{3.25}$$

$$\left\| R(\sigma_n^{n+p})x - \prod_n^{n+p} x \right\| \leq \sum_{m=n+1}^{n+p} \left\| R(\sigma_{m-1}^m) R(\sigma_n^{m-1})x - \prod_{m-1}^{m} R(\sigma_n^{m-1})x \right\| \tag{3.26}$$

We are now in a position to prove the main result of this section. In what follows, $f : H \to R$ is convex and differentiable. Its gradient ∇f is Lipschitz continuous with constant L, and assume that the set $S = argmin(f)$ is not empty. Finally, we write $B = \nabla f$, and use the notation introduced at the beginning of Section 3.3 concerning the semigroup $\mathcal{S}$, the sequence of mappings $\{\mathcal{E}_n\}_{n\in\mathbb{N}}$, the function ν, and the evolution systems $U_{\mathcal{S}}$ and $U_{\mathcal{E}}$.

Theorem 3.16 *Fix $x \in H$, and let $\lambda = \sup_{n\in\mathbb{N}} \lambda_n < 2/L$, and $(\lambda_n) \notin \ell^1$. For each $t > 0$, define $\phi_{\mathcal{S}}(t) = \mathcal{S}_t x$ and $\phi_{\mathcal{E}}(t) = \mathcal{E}_{\nu(t)} x$. Then, $\phi_{\mathcal{E}}$ is an almost-orbit of $U_{\mathcal{S}}$, and $\phi_{\mathcal{S}}$ is an almost-orbit of $U_{\mathcal{E}}$.*

Proof. Let us prove that $\phi_{\mathcal{E}}$ is an almost-orbit of $U_{\mathcal{S}}$. Notice that

$$\begin{aligned} \|\phi_{\mathcal{E}}(t+h) - U_{\mathcal{S}}(t+h,t)\phi_{\mathcal{E}}(t)\| &= \left\| \left[\prod_{\mathrm{i}=\nu(t)}^{\nu(t+h)} E_{\lambda_{\mathrm{i}}} \right] \mathcal{E}_{\nu(t)} x - \mathcal{S}_h\, \mathcal{E}_{\nu(t)} x \right\| \\ &\leq \left\| \left[\prod_{\mathrm{i}=\nu(t)}^{\nu(t+h)} E_{\lambda_{\mathrm{i}}} \right] \mathcal{E}_{\nu(t)} x - \mathcal{S}_\eta\, \mathcal{E}_{\nu(t)} x \right\| \\ &+ \left\| \mathcal{S}_\eta\, \mathcal{E}_{\nu(t)} x - \mathcal{S}_h\, \mathcal{E}_{\nu(t)} x \right\|, \end{aligned} \tag{3.27}$$

where $\eta = \sigma_{\nu(t+h)} - \sigma_{\nu(t)}$. For the first term on the right-hand side, we partition the interval $[0, \sigma_{\nu(t+h)}]$ into subintervals of the form $[\sigma_{\kappa(\mathrm{i})}, \sigma_{\kappa(\mathrm{i}+1)}]$, $\mathrm{i} = 0, \ldots, n+p$, where $\{\kappa(\mathrm{i})\}$ is any strictly increasing sequence satisfying $k(0) = 0$ and $\sigma_{\kappa(\mathrm{i})-1} < \mathrm{i}^2 \leq \sigma_{\kappa(\mathrm{i})}$, for each i. In other words, each subinterval contains exactly one perfect square. The numbers n and p are chosen so that $\kappa(n) = \nu(t)$ and $\kappa(n+p) = \nu(t+h)$.

Using the notation in (3.23) and (3.24) with $R(t) = \mathcal{S}_t$ and $J_j = E_{\lambda_j}$, along with inequality (3.25), we obtain

$$\begin{aligned} \left\| \left[\prod_{\mathrm{i}=\nu(t)}^{\nu(t+h)} E_{\lambda_{\mathrm{i}}} \right] \mathcal{E}_{\nu(t)} x - \mathcal{S}_\eta\, \mathcal{E}_{\nu(t)} x \right\| &= \left\| \left(\prod_n^{n+p} \right) \left(\prod_0^n x \right) - \mathcal{S}_{\sigma_n^{n+p}} \left(\prod_0^n x \right) \right\| \\ &\leq \sum_{m=n+1}^{n+p} \left\| \mathcal{S}_{\sigma_{m-1}^m} \prod_n^{m-1} \left(\prod_0^n x \right) - \prod_{m-1}^{m} \prod_n^{m-1} \left(\prod_0^n x \right) \right\| \\ &= \sum_{m=n+1}^{n+p} \left\| \mathcal{S}_{\sigma_{m-1}^m} \left(\prod_0^{m-1} x \right) - \prod_{m-1}^{m} \left(\prod_0^{m-1} x \right) \right\| \\ &\leq \sum_{m=n+1}^{n+p} \sqrt{\frac{2}{L} \sigma_{m-1}^m} \left\| \nabla f \left(\prod_0^{m-1} x \right) \right\|, \end{aligned}$$

where the last inequality follows from Lemma 3.4 and the fact that $\lambda_j \leq \frac{2}{L}$ for all j.
Next, we use (3.15) with $y = \mathrm{Proj}_S x$, to obtain

$$\left\| \left[\prod_{\mathrm{i}=\nu(t)}^{\nu(t+h)} E_{\lambda_{\mathrm{i}}} \right] \mathcal{E}_{\nu(t)} x - \mathcal{S}_\eta \, \mathcal{E}_{\nu(t)} x \right\| \leq \sqrt{\frac{2\alpha \mathrm{dist}(x,S)^2 + 2\lambda(f(x) - \min(f))}{\alpha^2 L}} \sum_{m=n+1}^{n+p} \frac{\sqrt{\sigma_{m-1}^m}}{\sigma_{\kappa(m-1)}}$$

But $(m-1)^2 \leq \sigma_{\kappa(m-1)}$, while $\sigma_{m-1}^m = \sigma_{\kappa(m)} - \sigma_{\kappa(m-1)} \leq \sigma_{\kappa(m+1)-1} - \sigma_{\kappa(m-1)} \leq (m+1)^2 - (m-1)^2 = 4m$, by construction. Therefore,

$$\sum_{m=n+1}^{n+p} \frac{\sqrt{\sigma_{m-1}^m}}{\sigma_{\kappa(m-1)}} \leq \sum_{m=n+1}^{n+p} \frac{2\sqrt{m}}{(m-1)^2} \leq 2\sqrt{2} \sum_{m=n}^{\infty} m^{-3/2} \leq 2\sqrt{2} \int_{n-1}^{\infty} x^{-3/2}\,\mathrm{d}x = \frac{4\sqrt{2}}{\sqrt{n-1}}.$$

We conclude that

$$\left\| \left[\prod_{\mathrm{i}=\nu(t)}^{\nu(t+h)} E_{\lambda_{\mathrm{i}}} \right] \mathcal{E}_{\nu(t)} x - \mathcal{S}_\eta \, \mathcal{E}_{\nu(t)} x \right\| \leq \sqrt{\frac{64\alpha \mathrm{dist}(x,S)^2 + 64\lambda(f(x) - \min(f))}{\alpha^2 L \kappa^{-1}(\nu(t) - 1)}} \tag{3.28}$$

To estimate the second term on the right-hand side of (3.27), we first use the Lipschitz continuity of the semigroup (see Proposition 3.5), and then (3.15) with $y = \mathrm{Proj}_S x$, to obtain

$$\begin{aligned} \left\| \mathcal{S}_\eta \, \mathcal{E}_{\nu(t)} x - \mathcal{S}_h \, \mathcal{E}_{\nu(t)} x \right\| &\leq |\eta - h| \, \|\nabla f(\mathcal{E}_{\nu(t)} x)\| \\ &\leq \left| \sigma_{\nu(t)}^{\nu(t+h)} - h \right| \frac{\sqrt{\alpha \mathrm{dist}(x,S)^2 + \lambda(f(x) - \min(f))}}{\alpha \sigma_{\nu(t)+1}} \\ &\leq \frac{4\sqrt{\alpha \mathrm{dist}(x,S)^2 + \lambda(f(x) - \min(f))}}{\alpha L \sigma_{\nu(t)+1}}. \end{aligned} \tag{3.29}$$

Finally, (3.27), (3.28) and (3.29) together imply

$$\lim_{t \to +\infty} \sup_{h \geq 0} \|\phi_{\mathcal{E}}(t+h) - U_{\mathcal{S}}(t+h,t)\phi_{\mathcal{E}}(t)\| = 0.$$

Using essentially the same arguments, we show that $\phi_{\mathcal{S}}$ is an almost-orbit of $U_{\mathcal{E}}$. We use (3.26) instead of (3.25), alng with the bounds given in Proposition 3.14, in place of those from Proposition 3.13. The details are left to the reader. □

The following result complements Theorem 3.12, establishing the equivalence between the convergence of trajectories/sequences generated by the steepest descent dynamics, the proximal point algorithm and the gradient method:

Theorem 3.17 *The following are equivalent:*

i) For every $x \in H$, $\mathcal{S}_t x$ converges weakly (resp. strongly), as $t \to +\infty$.

ii) For every $(\lambda_n) \notin \ell^1$ and every $x \in H$, $\mathcal{P}_n x$ converges weakly (resp. strongly), as $n \to +\infty$.

iii) There is $(\lambda_n) \notin \ell^1$ such that, for every $x \in H$, $\mathcal{P}_n x$ converges weakly (resp. strongly), as $n \to +\infty$.

iv) For every $(\lambda_n) \notin \ell^1$ *with* $\sup_{n\in\mathbb{N}} \lambda_n < 2/L$, *and every* $x \in H$, $\mathcal{E}_n x$ *converges weakly (resp. strongly), as* $n \to +\infty$.

v) There is $(\lambda_n) \notin \ell^1$ *with* $\sup_{n\in\mathbb{N}} \lambda_n < 2/L$ *such that, for every* $x \in H$, $\mathcal{E}_n x$ *converges weakly (resp. strongly), as* $n \to +\infty$.

3.4.3 Baillon's counterexample revisited

Let $f : H \to \mathbb{R} \cup \{+\infty\}$ be proper, lower-semicontinuous and convex. Bruck [12] proved that if f has minimizers, and $u : [0, +\infty) \to H$ satisfies

$$-\dot{u}(t) \in \partial f(u(t)) \tag{3.30}$$

for almost every $t > 0$, then $u(t)$ converges weakly, as $t \to +\infty$, to a minimizer of f. A few years later, Baillon [6] constructed a proper lower-semicontinuous convex function $f : \ell^2 \to \mathbb{R} \cup \{+\infty\}$ for which (3.30) has solutions that do not converge strongly. Baillon's function is not continuous, and its domain is not all of ℓ^2.

In [54], Martinet introduced the proximal point algorithm, and showed, for a constant sequence of step sizes $\lambda_n \equiv \lambda$, that z_n converges weakly, as $n \to +\infty$, to a minimizer of f. This result was extended to the case where $\{\lambda_n\}_{n\in\mathbb{N}}$ is bounded from below by a positive number [67], and to the case where $\{\lambda_n\}_{n\in\mathbb{N}} \notin \ell^1$ [11]. Using Baillon's counterexample and the concept of asymptotic semigroup from Passty [63] (see Theorem 3.17 above), Güler [39] showed that sequences generated by the proximal point algorithm do not always converge strongly.

We shall present a family of *smooth* convex functions for which the steepest descent dynamics, the proximal point algorithm and the gradient method all produce trajectories/sequences that do not converge strongly. Related results have been found in [4] by a different (constructive) argument.

Theorem 3.18 *Let* $\{\Lambda_n\}_{n\in\mathbb{N}} \notin \ell^1$ *be a bounded sequence. There is a convex function* $f : H \to \mathbb{R}$, *with Lipschitz continuous gradient, such that*

i) There is $u : [0, +\infty) \to H$ *satisfying* (3.20), *that converges weakly but not strongly as* $t \to +\infty$ *to a minimizer of* f.

ii) There is a proximal sequence $\{z_n\}_{n\in\mathbb{N}}$, *generated using step sizes* $\{\Lambda_n\}_{n\in\mathbb{N}}$, *that converges weakly but not strongly as* $n \to +\infty$ *to a minimizer of* f.

iii) There is an Euler sequence $\{x_n\}_{n\in\mathbb{N}}$, *generated using step sizes* $\{\Lambda_n\}_{n\in\mathbb{N}}$, *that converges weakly but not strongly as* $n \to +\infty$ *to a minimizer of* f.

Proof. Let $\mathcal{H} = \ell^2$, and let $\varphi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be the function in Baillon's example. Fix $\omega > 0$, and let $\varphi_\omega : \mathcal{H} \to \mathbb{R}$ be the Moreau envelope of φ with index ω, defined by

$$\varphi_\omega(x) = \inf_{y\in\mathcal{H}} \left\{ \varphi(y) + \frac{1}{2\omega}\|x - y\|^2 \right\}$$

for $x \in \mathcal{H}$ (see, for instance, [64, Section 3.5.4]). The function φ_ω is differentiable and $\nabla\varphi_\omega$ is cocoercive with constant ω. According to [39, Corollary 5.1], for every bounded sequence $\{\Lambda_n\}_{n\in\mathbb{N}} \notin \ell^1$, there exists $\tilde{x} \in \mathcal{H}$ such that the proximal sequence defined by

$$\begin{cases} x_{n+1} &= (I + \lambda_{n+1}\partial\varphi)^{-1}x_n, \quad n \geq 0 \\ x_0 &= \tilde{x}, \end{cases}$$

converges weakly, but not strongly, to 0, the only minimizer of φ. In particular, for the constant sequence $\tilde{\lambda}_n \equiv \omega$, there is at least one such $\tilde{x}$. But

$$(I + \omega\partial\varphi)^{-1}y = y - \omega\nabla\varphi_\omega(y)$$

for all $y \in \mathcal{H}$. It follows that (x_n) is also an Euler (gradient) sequence, generated from $\tilde{x}$ with step sizes identically ω. We have found a sequence $\{\tilde{\lambda}_n\}_{n\in\mathbb{N}} \notin \ell^1$ and an initial point $\tilde{x}$ such that the corresponding Euler sequence does not converge strongly. If $\{\Lambda_n\}_{n\in\mathbb{N}}$ is bounded by Λ, it suffices to pick $f = \varphi_\omega$ with $\omega > \Lambda/2$ to conclude, by virtue of Theorem 3.17. $\square$

3.5 The Banach space setting

Some of the results presented here can be extended to a class of Banach spaces in a rather straightforward manner. Let X be a real Banach space with topological dual X^*. The duality product is denoted $\langle\cdot,\cdot\rangle_{X^*,X}$.

From the proof of Lemma 3.4, which uses [70, Proposition 3.11], we observe that Kobayashi's inequality holds in arbitrary Banach spaces provided the sequences of step sizes are bounded by some $\Lambda > 0$, and E_Λ is nonexpansive. A simple sufficient condition for this can be given in 2-*uniformly smooth* spaces[2], reflexive spaces where $\mathcal{J}$ is single valued (see [50]), and where there is a constant $\kappa > 0$ such that

$$\|u + v\|^2 \leq \|u\|^2 + 2\langle\mathcal{J}(u), v\rangle_{X^*,X} + \kappa\|v\|^2$$

for all $u, v \in X$ (see [71, Corollary 1']). Hilbert spaces are 2-uniformly smooth with $\kappa = 1$. If $p \geq 2$, the L^p spaces are 2-uniformly smooth with constant $\kappa = p - 1$ (see [71, Corollary 2]). For $\lambda > 0$, we have

$$\begin{aligned} \|E_\lambda x - E_\lambda y\|^2 &= \|(x-y) - \lambda(Bx - By)\|^2 \\ &\leq \|x-y\|^2 - 2\lambda\langle\mathcal{J}(x-y), Bx - By\rangle_{X^*,X} + \kappa\lambda^2\|Bx - By\|^2 \\ &\leq \|x-y\|^2 + \lambda(\kappa\lambda - 2\theta)\|Bx - By\|^2. \end{aligned}$$

Therefore, it suffices to take $\Lambda \leq \frac{2\theta}{\kappa}$. As a consequence, we have the following:

Remark 3.19 Let X be 2–uniformly smooth with constant κ, and let $B : X \to X$ be cocoercive with constant θ. From Lemma 3.4 until Theorem 3.12, all results remain valid in X, replacing 2θ by $\frac{2\theta}{\kappa}$, when relevant.

[2]The arguments also hold for *q-cocoercive* operators in *q-uniformly smooth* spaces (see, for instance, [52]).

Chapter 4

Forward-backward approximation of evolution equations in finite and infinite horizon

A joint work with J. Peypouquet [29]

4.1 Sequences governed by a sum of accretive operators

Throughout this chapter, X will be a Banach space with topological dual X^*. Their norms and the duality product are denoted by $\|\cdot\|$, $\|\cdot\|_*$ and $\langle\cdot,\cdot\rangle$, respectively. In what follows, we assume that X^* is 2-uniformly convex, which implies that X is reflexive, the duality mapping is single valued, and there is a constant $\kappa > 0$ such that

$$\|u+v\|^2 \leq \|u\|^2 + 2\langle \mathcal{J}(u), v\rangle + \kappa\|v\|^2, \tag{4.1}$$

for all $u, v \in X$ (see [50, 71]). For instance, L^p spaces have this property for $p \geq 2$. Let $A : X \to 2^X$ be a *m-accretive* operator and $B : X \to X$ be a *cocoercive* with parameter $\theta > 0$. Then $A + B$ is m-accretive, the operator

$$E_\lambda = I - \lambda B \tag{4.2}$$

is nonexpansive for all $\lambda \in [0, \frac{2\theta}{\kappa}]$ and the *forward backward splitting operator* $T_\lambda : X \to X$ defined by

$$T_\lambda = J_\lambda^A \circ E_\lambda,$$

is single-valued, everywhere defined and nonexpansive.

Let us consider the following *forward-backward* iterations defined by

$$x_k = T_{\lambda_k}(x_{k-1}) = J^A_{\lambda_k}(E_{\lambda_k}(x_{k-1})), \qquad k \in \mathbb{N}, \tag{4.3}$$

where $\{\lambda_k\}_{k\in\mathbb{N}}$ is a sequence of positive number, called step sizes, $x_0 \in X$.

These *forward-backward iterations* are fundamental in the numerical analysis of optimization problems, since they serve as building blocks for first order methods. The gradient method, originally introduced by Cauchy in [15], and its variant, the projected gradient method [38, 49], the proximal point algorithm introduced by Martinet [54] and further extended by Rockafellar [67] and Brézis-Lions [11], and the proximal-gradient algorithm [63, 51], with applications in image and signal processing, such as the *iterative shrinkage thresholding algorithm* [33, 24], are keynote particular cases. Moreover, some primal dual methods [16, 27, 69] can be reduced to these types of iterations.

Although we will not study the convergence of the iterations (4.3) in this section (which will be studied in detail in Chapter 5), our main purpose here is to analyze them as discrete approximations of an evolution equation governed by the sum $A+B$. To this end, it is useful to rewrite (4.3) in a more general way

$$-\frac{x_n - x_{n-1}}{\lambda_n} + \varepsilon_n \in Ax_n + Bx_{n-1}, \quad n \in \mathbb{N}, \tag{4.4}$$

where ε_n accounts for possible perturbations or computational errors. In the notation of (4.3), this is

$$x_k = J^A_{\lambda_k}(E_{\lambda_k}(x_{k-1}) + \lambda_k \varepsilon_k). \qquad k \in \mathbb{N} \tag{4.5}$$

When $\varepsilon_k = 0$ for all $k \in \mathbb{N}$, the left-hand in (4.4) side can be interpreted as a discretization of the velocity for a trajectory $t \mapsto u(t)$, so (4.4) can be related to the differential inclusion

$$-\dot{u}(t) \in Au(t) + Bu(t), \tag{4.6}$$

for $t > 0$. In the following sections, we shall establish the nature of this relationship. On the one hand, we shall prove that the iterations described in (4.4) can be used, in at least two different ways, to construct a sequence of curves that approximate the solutions of (4.6) uniformly on each compact time interval. The existence of such solutions is obtained as a byproduct. On the other hand, we shall show that, given A and B, the trajectories satisfying (4.6) will have the same convergence properties, when $t \to +\infty$, as the sequences satisfying (4.4), when $k \to +\infty$, provided the step sizes are sufficiently small.

For $\lambda, \mu \in \left(0, \frac{\theta}{\kappa}\right]$, let $E^a_\lambda = E_\lambda(x) + \lambda a$, $E^b_\mu = E_\mu(x) + \mu b$ and $T^a_\lambda(x) = J^A_\lambda(E^a_\lambda(x))$, $T^b_\mu(y) = J^A_\mu(E^b_\mu(y))$, with $a, b, x, y \in X$. We have the following auxiliary result:

Lemma 4.1 *Write* $\Theta = \frac{\theta}{\kappa}$. *For* $\lambda, \mu \in (0, \Theta]$ *and* $a, b, x, y \in X$, *we have:*

$$\|T^a_\lambda(x) - T^b_\mu(y)\| \le \alpha\|y - T^a_\lambda(x)\| + \beta\|x - T^b_\mu(y)\| + \gamma\|x - y\| + \gamma\Theta\|a - b\|, \tag{4.7}$$

where

$$\alpha = \frac{\mu(\Theta - \lambda)}{\Theta(\lambda + \mu) - \lambda\mu}, \ \beta = \frac{\lambda(\Theta - \mu)}{\Theta(\lambda + \mu) - \lambda\mu}, \ \gamma = \frac{\lambda\mu}{\Theta(\lambda + \mu) - \lambda\mu}. \tag{4.8}$$

Proof. Setting

$$\begin{aligned}
(\lambda+\mu)\|T^a_\lambda(x) - T^b_\mu(x)\|^2 &= (\lambda+\mu)\langle T^a_\lambda(x) - T^b_\mu(y), \Delta\rangle_{X,X^*} \\
&= \mu\langle E^a_\lambda(x) - T^b_\mu(y), \Delta\rangle - \lambda\langle E^b_\mu(y) - T^a_\lambda(x), \Delta\rangle \\
&+ \lambda\mu\left\langle \frac{1}{\mu}(E^b_\mu(y) - T^b_\mu(y)) - \frac{1}{\lambda}(E^a_\lambda(x) - T^a_\lambda(x)), \Delta\right\rangle \\
&\le \mu\langle E^a_\lambda(x) - T^b_\mu(y), \Delta\rangle - \lambda\langle E^b_\mu(y) - T^a_\lambda(x), \Delta\rangle,
\end{aligned} \tag{4.9}$$

since A is accretive and

$$\frac{E^b_\mu(y) - T^b_\mu(y)}{\mu} \in A(T^b_\mu(y)) \quad ; \frac{E^a_\lambda(x) - T^a_\lambda(x)}{\lambda} \in A(T^a_\lambda(x))$$

for all $\lambda, \mu \in (0, \Theta]$ and $a, b, x, y \in X$. On the other hand

$$-\lambda\mu\|T^a_\lambda(x) - T^b_\mu(y)\|^2 \;\; = -\lambda\mu\langle x - T^b_\mu(y), \Delta\rangle + \lambda\mu\langle y - T^a_\lambda(x), \Delta\rangle + \lambda\mu\langle x - y, \Delta\rangle. \quad (4.10)$$

Combining (4.9) and (4.10), we obtain:

$$\begin{aligned}[\Theta(\lambda+\mu) - \lambda\mu]\|T^a_\lambda(x) - T^b_\mu(y)\|^2 \;\; \leq \;\; & \mu(\Theta - \lambda)\langle x - T^b_\mu(y), \Delta\rangle - \lambda(\Theta - \mu)\langle y - T^a_\lambda(x), \Delta\rangle \\ + \;\; & \lambda\mu\langle E_\Theta(x) - E_\Theta(y), \Delta\rangle + \Theta\lambda\mu\langle a - b, \Delta\rangle.\end{aligned}$$

Since $\Theta(\lambda+\mu) - \lambda\mu > 0$ and E_Θ is nonexpansive, we obtain

$$\|T^a_\lambda(x) - T^b_\mu(y)\| \;\; \leq \;\; \alpha\|y - T^a_\lambda(x)\| + \beta\|x - T^b_\mu(y)\| + \gamma\|x - y\| + \gamma\Theta\|a - b\|.$$

□

There exist previous versions of Kobayashi's inequiality–types related to m-accretive operators (maximally monotone operators in Hilbert space setting). The first version was proposed in [44, Lemma 2.1] (see also [65] for the version in Hilbert space setting) and allows to estimate the distance between arbitrary iterates of two independent sequences generated by (4.3) with $B = 0$. A new version of Kobayashi's inequality was presented in Lemma 3.4 when the sequence are generated by (4.3) when $A = 0$ and $\varepsilon_n = 0$ for all $n \in \mathbb{N}$. In the following result, we provide a new version of Kobayashi's inequality to forward–backward iterations (4.5).

Theorem 4.2 *Let $\{x_k\}_{k\in\mathbb{N}}$, $\{\widehat{x}_l\}_{l\in\mathbb{N}}$ be two sequences generated by (4.5), with stepsizes $\{\lambda_k\}_{k\in\mathbb{N}}$ and $\{\widehat{\lambda}_l\}_{l\in\mathbb{N}}$, respectively and $0 < \lambda_k, \widehat{\lambda}_l \leq \Theta$ for all $k, l \in \mathbb{N}$. Then, for $u \in D(A)$*

$$\|x_k - \widehat{x}_l\| \leq \|x_0 - u\| + \|\widehat{x}_0 - u\| + \|\|(A+B)u\|\|\sqrt{(\sigma_k - \widehat{\sigma}_l)^2 + \tau_k + \widehat{\tau}_l} + S_k + \widehat{S}_l, \quad (4.11)$$

where $S_k = \sum_{\mathrm{i}=1}^{k} \lambda_\mathrm{i}\|\varepsilon_\mathrm{i}\|$, $\|\|Au\|\| = \inf_{v\in Au}\|v\|$, $\sigma_k = \sum_{\mathrm{i}=1}^{k}\lambda_\mathrm{i}$ and $\tau_k = \sum_{\mathrm{i}=1}^{k}\lambda_\mathrm{i}^2$ (similarly for $\widehat{S}_l$, $\widehat{\sigma}_l$ and $\widehat{\tau}_l$).

Proof. To simplify notation, set

$$C_{k,l} = \sqrt{(\sigma_k - \widehat{\sigma}_l)^2 + \tau_k + \widehat{\tau}_l}.$$

In view the characterization (4.4) of the sequence $\{x_k\}_{k\in\mathbb{N}}$, we have

$$y_k = \frac{E_{\lambda_k}(x_{k-1}) + \lambda_k\varepsilon_k - x_k}{\lambda_k} \in Ax_k \quad \text{for each} \quad k \in \mathbb{N}.$$

Thus, for $v \in Au$ and the accretivity of A we obtain:

$$\begin{aligned}\|x_k - u\| \;\; & \leq \;\; \|x_k - u + \lambda_k(y_k - v)\| = \|E_{\lambda_k}(x_{k-1}) - E_{\lambda_k}(u) + \lambda_k\varepsilon_k - \lambda_k(Bu + v)\| \\ & \leq \;\; \|E_{\lambda_k}(x_{k-1}) - E_{\lambda_k}(u)\| + \lambda_k\|v + Bu\| + \lambda\|\varepsilon_k\|.\end{aligned}$$

Since E_{λ_k} is nonexpansive and $v \in Au$ is arbitrary, we obtain

$$\|x_k - u\| \leq \|x_{k-1} - u\| + \lambda_k \||(A+B)u\|| + \lambda_k \|\varepsilon_k\|.$$

Iterating this inequality we have

$$\|x_k - u\| \leq \|x_0 - u\| + \sigma_k \||(A+B)u\|| + S_k,$$

where we get

$$\begin{aligned} \|x_k - \widehat{x}_0\| &= \|x_k - u\| + \|\widehat{x}_0 - u\| \\ &\leq \|x_0 - u\| + \|\widehat{x}_0 - u\| + \sigma_k \||(A+B)u\|| + S_k \\ &\leq \|x_0 - u\| + \|\widehat{x}_0 - u\| + C_{k,0} \||(A+B)u\|| + S_k \end{aligned}$$

Thus, the inequality holds for $(k, 0)$. In a similar fashion we prove the inequality for $(0, l)$, for $l \geq 0$. The proof will continue using induction on the pair (k, l). For the inductive step, we assume the inequality holds for the pairs $(k-1, l-1)$, $(k, l-1)$ and $(k-1, l)$, and prove that it also holds for the pair (k, l). To this end, we use the inequality (4.7) with $x = x_{k-1}$, $y = \widehat{x}_{l-1}$, $\lambda = \lambda_k$, $\mu = \widehat{\lambda}_l$, $a = \lambda_k \varepsilon_k$ and $b = \widehat{\lambda}_l \widehat{\varepsilon}_l$:

$$\|x_k - \widehat{x}_l\| \leq \alpha_{k,l} \|x_k - \widehat{x}_{l-1}\| + \beta_{k,l} \|x_{k-1} - \widehat{x}_l\| + \gamma_{k,l} \|x_{k-1} - \widehat{x}_{l-1}\| + \gamma_{k,l} \Theta \|\varepsilon_k - \widehat{\varepsilon}_l\|. \quad (4.12)$$

Now, using the induction hypothesis in (4.12) and $\alpha_{k,l} + \beta_{k,l} + \gamma_{k,l} = 1$, we deduce that

$$\begin{aligned} \|x_k - \widehat{x}_l\| &\leq \|x_0 - u\| + \|\widehat{x}_0 - u\| + \||(A+B)u\|| \left(\alpha_{k,l} c_{k,l-1} + \beta_{k,l} c_{k-1,l} + \gamma_{k,l} c_{k-1,l-1}\right) \\ &+ \gamma_{k,l} \Theta(\|\varepsilon_k\| + \|\widehat{\varepsilon}_l\|) + \alpha_{k,l}(S_k + \widehat{S}_{l-1}) + \beta_{k,l}(S_{k-1} + \widehat{S}_l) + \gamma_{k,l}(S_{k-1} + \widehat{S}_{l-1}). \\ &= \|x_0 - u\| + \|\widehat{x}_0 - u\| + \||(A+B)u\|| \left(\alpha_{k,l} c_{k,l-1} + \beta_{k,l} c_{k-1,l} + \gamma_{k,l} c_{k-1,l-1}\right) \\ &+ S_{k-1} + \widehat{S}_{l-1} + (\alpha_{k,l} \lambda_k + \gamma_{k,l} \Theta) \|\varepsilon_k\| + (\beta_{k,l} \widehat{\lambda}_a + \gamma_{k,l} \Theta) \|\widehat{\varepsilon}_l\| \\ &= \|x_0 - u\| + \|\widehat{x}_0 - u\| + \||(A+B)u\|| \left(\alpha_{k,l} c_{k,l-1} + \beta_{k,l} c_{k-1,l} + \gamma_{k,l} c_{k-1,l-1}\right) + S_k + \widehat{S}_l \end{aligned} \quad (4.13)$$

since $\alpha_{k,l} \lambda_k + \gamma_{k,l} \Theta = \lambda_k$ and $\beta_{k,l} \widehat{\lambda}_l + \gamma_{k,l} \Theta = \widehat{\lambda}_l$. On the other hand, notice that

$$\begin{aligned} \alpha_{k,l} c_{k,l-1} + \beta_{k,l} c_{k-1,l} + \gamma_{k,l} c_{k-1,l-1} &\leq \sqrt{\alpha_{k,l} + \beta_{k,l} + \gamma_{k,l}} \sqrt{\alpha_{k,l} c_{k,l-1}^2 + \beta_{k,l} c_{k-1,l}^2 + \gamma_{k-1,l-1} c_{k,l}^2} \\ &= \sqrt{\alpha_{k,l} c_{k,l-1}^2 + \beta_{k,l} c_{k-1,l}^2 + \gamma_{k-1,l-1} c_{k,l}^2}, \end{aligned} \quad (4.14)$$

On the other hand

$$\begin{aligned} c_{k,l-1}^2 &= c_{k,l}^2 + 2\widehat{\lambda}_l(\sigma_k - \widehat{\sigma}_l) \\ c_{k-1,l}^2 &= c_{k,l}^2 + 2\lambda_k(\sigma_k - \widehat{\sigma}_l) \\ c_{k-1,l-1}^2 &= c_{k,l}^2 + 2(\widehat{\lambda}_l - \lambda_k)(\sigma_k - \widehat{\sigma}_l) - 2\lambda_k \widehat{\lambda}_l. \end{aligned}$$

Therefore,

$$\alpha_{k,l} c_{k,l-1}^2 + \beta_{k,l} c_{k-1,l}^2 + \gamma_{k,l} c_{k-1,l-1}^2 = c_{k,l}^2 - 2\gamma_{k,l} \lambda_k \widehat{\lambda}_l \leq c_{k,l}^2. \quad (4.15)$$

Combining (4.13), (4.14) and (4.15), we obtain (4.11). □

4.2 Approximation in finite horizon

Theorem 4.2 provides existence and regularity results for the evolution equation

$$\begin{cases} -\dot{u}(t) & \in & (A+B)u(t), \quad \text{for almost every } \; t>0, \\ u(0) & = & u_0 \in \overline{D(A)}, \end{cases} \tag{4.16}$$

by means of an approximation scheme. For each $t \geq 0$ and $m \geq 1$, set

$$u_m(t) = \left[T_{\frac{t}{m}}\right]^m u_0. \tag{4.17}$$

In other words, $u_m(t)$ is the m-th term of the forward-backward sequence generated by (4.3) from u_0 using the constant step size $\lambda_k \equiv t/m$ for all $k \in \mathbb{N}$. We shall prove that $\{u_m\}$ converges uniformly on compact intervals to a Lipschitz-continuous function satisfying (4.16).

Proposition 4.3 *The sequence $\{u_m\}_{m\in\mathbb{N}}$ converges pointwise on $[0,+\infty)$ and uniformly on $[0,S]$ for each $S>0$, to a function $u:[0,+\infty)\to X$ which is globally Lipschitz-continuous with constant $|||(A+B)u_0|||$.*

Proof. We may assume that $u_0 \in D(A)$. Extension to $\overline{D(A)}$ will then be possible in view of the Lipschitz (thus uniform) continuity. Given $t,s>0$ and $n,m \in \mathbb{N}$, define $u_m(t)$ and $u_n(s)$ as above. By Theorem 4.2, we have:

$$\|u_m(t)-u_n(s)\| \leq |||(A+B)u_0|||\sqrt{(t-s)^2+\frac{t^2}{m}+\frac{s^2}{n}}. \tag{4.18}$$

For $s=t$, this gives

$$\|u_m(t)-u_n(t)\| \leq t\,|||(A+B)u_0|||\sqrt{\frac{1}{m}+\frac{1}{n}}.$$

It follows that $\{u_m\}_{m\in\mathbb{N}}$ converges pointwise on $[0,+\infty)$, and uniformly on $[0,S]$ for each $S>0$, to a function $u:[0,+\infty)\to X$. Passing to the limit in (4.18), we obtain

$$\|u(t)-u(s)\| \leq |||(A+B)u_0|||\cdot|t-s|$$

for all $t,s>0$. □

Remark 4.4 Given $S>0$ and $m\geq 1$, define $v_m:[0,S]\to X$ by

$$v_m(t) = \left[T_{\frac{S}{m}}\right]^{\mu(t)} u_0, \quad \text{where} \quad \mu(t) = \left\lfloor m\frac{t}{S}\right\rfloor \quad \text{and} \quad t\in[0,S]. \tag{4.19}$$

This is a piecewise constant interpolation of the forward-backward sequence generated with $\frac{S}{m}$ as step sizes, and initial point u_0 for $k=1,\dots m$. In order to estimate the distance between v_m and u_m defined in (4.17) and (4.19) respectively, we use (4.11) to obtain

$$\|u_m(t)-v_m(t)\| \leq |||(A+B)u_0|||\sqrt{\frac{S^2}{m^2}+\frac{t^2}{m}+\frac{tS}{m}} \leq \frac{3S}{\sqrt{m}}|||(A+B)u_0|||.$$

Whence, as $m\to\infty$, v_m also converges uniformly on $[0,S]$, for each $S>0$, to the same function u.

Theorem 4.5 *The function u, given by Proposition 4.3, satisfies* (4.16).

PROOF. In this case, it suffices to verify that it is an integral solution in the Benilan's sense [9, Proposition 2.5], which means that, whenever $y \in (A+B)x$ and $S \geq t > s \geq 0$, we have

$$\|u(t)-x\|^2 - \|u(s)-x\|^2 \leq 2\int_s^t \langle \mathcal{J}(x-u(\tau)), y\rangle \mathrm{d}\tau. \tag{4.20}$$

If $\{x_n\}_{n\in\mathbb{N}}$ is any sequence generated by (4.3) with steps sizes $\{\lambda_n\}_{n\in\mathbb{N}}$, then

$$-(x_n - x_{n-1}) - \lambda_n Bx_{n-1} + \lambda_n Bx_n \in \lambda_n Ax_n + \lambda_n Bx_n$$

for each $n \in \mathbb{N}$. In view of the monotonicity of $A+B$, we have

$$\langle \mathcal{J}(x-x_n), \lambda_n y + x_n - x_{n-1} + \lambda_n Bx_{n-1} - \lambda_n Bx_n\rangle \geq 0,$$

whenever $y \in Ax + Bx$. Whence,

$$\begin{aligned}
2\lambda_n\langle \mathcal{J}(x-x_n), y\rangle &\geq 2\langle \mathcal{J}(x-x_n), x_{n-1}-x_n\rangle + 2\lambda_n\langle \mathcal{J}(x-x_n), Bx_n - Bx_{n-1}\rangle \\
&= 2\|x_n - x\|^2 + 2\langle \mathcal{J}(x-x_n), x_{n-1}-x\rangle + 2\lambda_n\langle \mathcal{J}(x-x_n), Bx_n - Bx_{n-1}\rangle \\
&\geq \|x_n - x\|^2 - \|x_{n-1}-x\|^2 + 2\lambda_n\langle \mathcal{J}(x-x_n), Bx_n - Bx_{n-1}\rangle \\
&\geq \|x_n - x\|^2 - \|x_{n-1}-x\|^2 - 2\theta^{-1}\lambda_n\|x-x_n\|\|x_n - x_{n-1}\|.
\end{aligned}$$

Thus

$$\|x_n - x\|^2 - \|x_{n-1}-x\|^2 \leq 2\lambda_n\langle \mathcal{J}(x-x_n), y\rangle + 2\theta^{-1}\lambda_n\|x-x_n\|\|x_n-x_{n-1}\|. \tag{4.21}$$

Now, let us choose $x_0 = u_0$, $\lambda_n \equiv \frac{S}{m}$, where m is fixed but arbitrary. In view of Remark 4.4, there is a constant $K > 0$ such that $2\theta^{-1}\|x - x_n\| \leq K$ for $n = 1, \ldots, m$. Summing up for $n = \mu(s), \cdots, \mu(t)$ in (4.21), we obtain

$$\begin{aligned}
\|v_m(t)-x\|^2 - \|u(s)-x\|^2 &\leq 2\sum_{n=\mu(s)}^{\mu(t)} \frac{S}{m}\left[\langle \mathcal{J}(x-x_n), y\rangle + K\|x_n - x_{n-1}\|\right] \\
&\leq 2\sum_{n=\mu(s)}^{\mu(t)} \frac{S}{m}\langle \mathcal{J}(x-x_n), y\rangle + \sum_{n=\mu(s)}^{\mu(t)} \frac{6S^2K|\!|\!|(A+B)u_0|\!|\!|}{m\sqrt{m}} \\
&= 2\sum_{n=\mu(s)}^{\mu(t)} \frac{S}{m}\langle \mathcal{J}(x-x_n), y\rangle + (\mu(t)-\mu(s))\frac{6S^2K|\!|\!|(A+B)u_0|\!|\!|}{m\sqrt{m}} \\
&\leq 2\sum_{n=\mu(s)}^{\mu(t)} \frac{S}{m}\langle \mathcal{J}(x-x_n), y\rangle + \frac{6S^2K|\!|\!|(A+B)u_0|\!|\!|}{\sqrt{m}}.
\end{aligned}$$

We obtain (4.20) by letting $m \to \infty$. □

As a consequence of Theorem (4.2) and Theorem 4.5, we have

Corollary 4.6 . *Let $\{x_k\}_{k\in\mathbb{N}}$ be a sequence generated by* (4.3) *with $\varepsilon_k = 0$ for all $k \in \mathbb{N}$ and $u : [0,T] \to X$ a solution of* (4.17). *Then*

(i) The function $t \mapsto |\!|\!|(A+B)u(t)|\!|\!|$ is nonincreasing.

(ii) $\|x_k - u(t)\| \leq \|x_0 - u_0\| + \min\{|\!|\!|(A+B)x_0|\!|\!|, |\!|\!|(A+B)u_0|\!|\!|\}\sqrt{(\sigma_k - t)^2 + \tau_k}$.

4.3 Approximation in infinite horizon

In this section, we show that forward–backward sequence presented in (4.3) have the same asymptotic behavior as the number of the iterations goes to infinity, as the solutions of the evolution equation (4.16). The key argument is the idea of asymptotic equality introduced by Passty [63], closely related to the notion of almost-orbit, introduced by Miyadera and Kobayasi [46]. Further commentaries on this topic can be found in [1, 2, 3]. To simplify the notation, given $x \in \overline{D(A)}$ and $t \geq 0$, we write

$$\mathcal{S}_t x = u(t), \tag{4.22}$$

where u satisfies (4.16) with $x_0 = x$. Also, for $0 \leq s \leq t$, we write

$$U_{\mathcal{S}}(t,s) = \mathcal{S}(t-s). \tag{4.23}$$

In a similar fashion, if $n \in \mathbb{N}$ and $x \in X$, we denote

$$\mathcal{T}_n x = T_{\lambda_n} \circ \cdots \circ T_{\lambda_1} x. \tag{4.24}$$

In other words, $\mathcal{T}_n x$ is the n-th term of the forward-backward sequence (4.3) starting from $x \in \overline{D(A)}$. Assume $\{\lambda_n\}_{n\in\mathbb{N}} \notin \ell^1$, and write $\nu(t) = \max\{n \in \mathbb{N} : \sigma_n \leq t\}$. For $0 \leq s \leq t$, we set

$$U_{\mathcal{T}}(t,s) = \prod_{\mathrm{i}=\nu(s)+1}^{\nu(t)} T_{\lambda_{\mathrm{i}}}, \tag{4.25}$$

where the product denotes composition of functions and the empty composition is the identity.

Example 4.7 The families $\big(U_{\mathcal{S}}\big)$ and $\big(U_{\mathcal{T}}\big)$, defined in (4.23) and (4.25), respectively, are nonexpansive evolution systems. Actually, the same is true if $\mathcal{S}$ is replaced by any other semigroup of nonexpansive functions on X, and if each $T_{\lambda_{\mathrm{i}}}$ is replaced by any other nonexpansive function on X.

The following result from [2, Theorem 3.3] reveals the usefulness of the concept of almost-orbit.

Proposition 4.8 *Let U be a nonexpansive evolution system and let ϕ be an almost-orbit of U. If, for each $x \in X$ and $s \geq 0$, $U(t,s)x$ converges weakly (resp. strongly) as $t \to \infty$, then so does $\phi(t)$. The same holds if the word "converges" is replaced by "almost-converges" or "converges in average".*

The following result establishes a relationship between the trajectories generated by $U_{\mathcal{S}}$ and $U_{\mathcal{T}}$:

Theorem 4.9 *Let $\{\lambda_n\}_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1$, and fix $x \in X$. For each $t > 0$, define $\phi_{\mathcal{S}}(t) = \mathcal{S}_t x$ and $\phi_{\mathcal{T}}(t) = \mathcal{T}_{\nu(t)} x$[1]. Then, $\phi_{\mathcal{S}}$ is an almost-orbit of $U_{\mathcal{T}}$, and $\phi_{\mathcal{T}}$ is an almost-orbit of $U_{\mathcal{S}}$.*

[1] This is a piecewise constant interpolation of the sequence $\mathcal{T}_n x$.

Proof. We first prove that $\phi_{\mathcal{S}}$ is an almost-orbit of $U_{\mathcal{T}}$. By Theorem 4.2, we have

$$\begin{aligned}\left\|\left[\prod_{k=1}^{m} T_{\frac{h}{m}}\right]\mathcal{S}_t x-\left[\prod_{\mathrm{i}=\nu(t)+1}^{\nu(t+h)} T_{\lambda_{\mathrm{i}}}\right]\mathcal{S}_t x\right\| &\leq \|\!|(A+B)\mathcal{S}_t x|\!\|\sqrt{\left(\sigma_{\nu(t)+1}^{\nu(t+h)}-h\right)^2+\tau_{\nu(t)+1}^{\nu(t+h)}+\frac{h^2}{m}}\\ &\leq \|\!|(A+B)x|\!\|\sqrt{4\rho^2(t)+\tau_{\nu(t)+1}^{\infty}+\frac{h^2}{m}},\end{aligned}$$

where $\sigma_k^n=\sigma_n-\sigma_k$, $\tau_k^n=\tau_n-\tau_k$ and $\rho(t):=\sup\{\lambda_n : n\geq \nu(t)-1\}$, which vanishes as $t\to\infty$. Passing to the limit as $m\to+\infty$, we obtain

$$\|\mathcal{S}_h\mathcal{S}_t x-U_{\mathcal{T}}(t+h,t)\mathcal{S}_t x\|\leq \|\!|(A+B)x|\!\|\sqrt{4\rho^2(t)+\tau_{\nu(t)+1}^{\infty}},$$

which is uniform in $h\geq 0$. It follows that

$$\lim_{t\to+\infty}\sup_{h\geq 0}\|\phi_{\mathcal{S}}(t+h)-U_{\mathcal{T}}(t+h,t)\phi_{\mathcal{S}}(t)\|=0.$$

To prove that $\phi_{\mathcal{T}}$ is an almost-orbit of $U_{\mathcal{S}}$, we proceed in a similar fashion, using Theorem 4.2 to obtain

$$\left\|\prod_{\mathrm{i}=\nu(t)+1}^{\nu(t+h)} T_{\lambda_{\mathrm{i}}}\mathcal{T}_{\nu(t)}x-\prod_{k=1}^{m} T_{\frac{h}{m}}\mathcal{T}_{\nu(t)}x\right\|\leq \|\!|(A+B)x|\!\|\sqrt{4\rho^2(t)+\tau_{\nu(t)+1}^{\infty}+\frac{h^2}{m}},$$

then passing to the limit as $m\to+\infty$ to deduce that

$$\|\phi_{\mathcal{T}}(t+h)-\mathcal{S}_h\phi_{\mathcal{T}}(t)\|\leq \|\!|(A+B)x|\!\|\sqrt{4\rho^2(t)+\tau_{\nu(t)+1}^{\infty}},$$

and conclude. □

Theorem 4.9 implies [63, Lemmas 4 & 6], [47, Proposition 2.3], [46, Proposition 7.4],[65, Proposition 8.6 i) & 8.7], [30, Theorem 3.1]. Combining Theorem 4.9 with Proposition 4.8, and using [74, Lemma 5.3], we obtain:

Theorem 4.10 *The following statements are equivalent:*

i) For every $z\in\overline{D(A)}$, $\mathcal{S}_t z$ converges strongly (weakly) as $t\to\infty$.

ii) For every initial point $x_0\in X$, every sequence of step sizes $\{\lambda_n\}_{n\in\mathbb{N}}\in\ell^2\setminus\ell^1$ and every sequence of errors $\{\varepsilon_n\}_{n\in\mathbb{N}}$ such that $\sum_{n=1}^{\infty}\|\varepsilon_n\|<+\infty$, the sequence $\{x_n\}_{n\in\mathbb{N}}$ generated by (4.5), converges strongly (weakly), as $n\to\infty$.

iii) There exist a sequence of step sizes $\{\lambda_n\}_{n\in\mathbb{N}}\in\ell^2\setminus\ell^1$ such that, for every initial point $x_0\in X$, the sequence $\{x_n\}_{n\in\mathbb{N}}$ generated by (4.4), converges strongly (weakly), as $n\to\infty$.

Theorem 4.10 implies [63, Theorem 1 & Theorem 2], [47, Theorem], [46, Theorem 7.5], as well as [30, 3.2].

New convergence results for forward–backward sequences on Banach spaces.

Recall from Section 4.1 that X is a real Banach sace with 2-uniformly convex dual, A is m-accretive, B is cocoercive. Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence representing computational errors and let $\{x_n\}_{n\in\mathbb{N}}$ satisfy (4.5). We assume that $\sum_{n=1}^{\infty}\|\varepsilon_n\| < \infty$. Finally, set $\mathcal{A} = A + B$ and $\Sigma = \mathcal{A}^{-1}(0)$, and assume that $\mathcal{A} \neq \emptyset$.. To simplify the statements and arguments, suppose X is uniformly convex. We know that Σ is closed and convex, and the projection P_Σ is well defined, single-valued and continuous.

Theorem 4.11 *Let $\{\lambda_n\}_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1$. Assume one of the following conditions holds:*

i) There is $\alpha > 0$ such that for every $x \notin \Sigma$ and every $y \in \mathcal{A}(x)$,

$$\langle y, \mathcal{J}(x - P_\Sigma(x))\rangle \geq \alpha\|x - P_\Sigma(x)\|^2$$

.

ii) J_1^A is compact and, for every every $y \in \mathcal{A}(x)$, $\langle y, \mathcal{J}(x - P_\Sigma(x))\rangle \geq 0$; or

iii) The interior of Σ is not empty.

Then, $\{x_n\}_{n\in\mathbb{N}}$ converges strongly, as $n \to \infty$, to a point in Σ.

Proof. In all three cases, we first prove that for each $z \in \overline{D(\mathcal{A})}$, $\mathcal{S}_t z$ converges strongly, as $t \to +\infty$, to a point in Σ.

i) The hypothesis of [59, Theorem 1] are easily verified.

ii) It suffices to combine [59, Proposition 1] and [59, Theorem 1].

iii) We use [59, Theorem 4].

We conclude by applying Theorem (4.10). □

Chapter 5

Work in progress 1: Analysis of the nonautonomous case

Álvarez and Peypouquet in [1] provide a generalization to the nonautonomous case of the Kobayashi's inequality for the proximal iterates governed by nonautonomous maximally monotone operator and the asymptotic equivalence between continuous trajectories associated to the differential inclusion governed by a family of nonautonomous m-accretive operators and the corresponding proximal iterates. Our main goal is to extend these results to the case where the continuous and discretes trajectories are governed by the sum of a maximally monotone operator with a cocoercive operator in the Hilbert space setting.

Advanced work

The advanced work is Kobayashi-type inequality which enable us to get an estimation of the distance between two iterates of independent sequences governed by two families of nonautonomous maximal monotones with a certain structure in Hilbert space setting, as we explain below.

Throughout this chapter, X will be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. For $D \subset X$ a nonempty set, let $\{A(t)\}_{t \geq 0}$ and $\{\widehat{A}(t)\}_{t \geq 0}$ be two families of maximally montone operators with common domain D and let $B : X \to X$ be a cocoercive operator of parameter $\theta \geq 0$, defined on X.

Consider two forward–backward schemes governed by $A_n, \widehat{A}_m$ and B and respectively:

$$-\frac{x_{n+1} - x_n}{\lambda_{n+1}} \in A_{n+1} x_{n+1} + B x_n, \; -\frac{\widehat{x}_{m+1} - \widehat{x}_m}{\widehat{\lambda}_{m+1}} \in \widehat{A}_{m+1} \widehat{x}_{m+1} + B \widehat{x}_m, \quad n, m \geq 1, \tag{5.1}$$

with initial points $x_0, \widehat{x}_0 \in D$ respectively, $A_n = A(\sigma_n)$, $\widehat{A}_m = \widehat{A}(\widehat{\sigma}_m)$, $\sigma_n = \sum_{\mathrm{i}=1}^{n} \lambda_{\mathrm{i}}$ and $\tau_n = \sum_{\mathrm{i}=1}^{n} \lambda_{\mathrm{i}}^2$ (similarly for $\widehat{\sigma}_m$ and $\widehat{\tau}_m$ respectively), with $\lambda_n, \widehat{\lambda}_m \in (0, \theta]$, for all $n, m \in \mathbb{N}$. According with Lemma 3.1, the function $E_\lambda = I - \lambda B$ is a nonexpansive function for all

$\lambda \in (0, \theta]$, which allows us rewrite the iterations defined in (5.1) by

$$x_{n+1} = J_{\lambda_n}^{A_n}(E_{\lambda_n}(x_n)),\ \widehat{x}_{m+1} = J_{\widehat{\lambda}_m}^{A_m}(E_{\widehat{\lambda}_m}(\widehat{x}_m)). \tag{5.2}$$

Such as in [1], we shall assume that the following condition holds:

$$\forall m, n \in \mathbb{N}, \exists \Theta_{n,m} \geq 0, \forall [x, v] \in A_n, \forall [\widehat{x}, \widehat{v}] \in \widehat{A}_m, \quad \frac{\langle x - \widehat{x}, \widehat{v} - v \rangle}{\|x - \widehat{x}\|} \leq \Theta_{n,m}. \tag{5.3}$$

The condition (5.3) was introduced by [45] to determinate the existance of weak solutions for nonautonomous differential inclusions.

The main goal in this section is provide a global estimation of the distance between two iterates of the independent sequences generated by (5.2). First, we shall need the following technical result. For $\lambda > 0$, let $T_\lambda : X \to X$ be the mapping defined by $T_\lambda = J_\lambda^A \circ E_\lambda$.

Lemma 5.1 *For $\lambda, \mu \in (0, \theta]$, we have*

$$\|T_\lambda(u) - \widehat{T}_\mu(v)\| \leq \alpha \|v - T_\lambda(u)\| + \beta \|u - \widehat{T}_\mu(v)\| + \gamma \|u - v\| + \gamma\theta \left\langle w - \widehat{w}, \frac{\Delta}{\|\Delta\|} \right\rangle,$$

where

$$\alpha = \frac{\mu(\theta - \lambda)}{\theta(\lambda + \mu) - \lambda\mu},\ \beta = \frac{\lambda(\theta - \mu)}{\theta(\lambda + \mu) - \lambda\mu},\ \gamma = \frac{\lambda\mu}{\theta(\lambda + \mu) - \lambda\mu}. \tag{5.4}$$

Proof. For $\lambda, \mu \in (0, \theta]$ and setting $\Delta = T_\lambda(v) - \widehat{T}_\mu(u)$ we have

$$\begin{aligned}
\theta(\lambda + \mu)\|T_\lambda(u) - \widehat{T}_\mu(v)\|^2 &= \theta\mu\langle E_\lambda(u) - \widehat{T}_\mu(v), \Delta\rangle - \theta\lambda\langle E_\mu(v) - T_\lambda(u), \Delta\rangle \\
&+ \theta\lambda\mu \left\langle \frac{1}{\mu}(E_\mu(v) - \widehat{T}_\mu(v)) - \frac{1}{\lambda}(E_\lambda(u) - T_\lambda(u)), \Delta \right\rangle \\
&= \theta\mu \left\langle u - \widehat{T}_\mu(v), \Delta \right\rangle - \theta\lambda \langle v - T_\lambda(u), \Delta\rangle \\
&+ \theta\lambda\mu \langle Bv - Bu, \Delta\rangle + \theta\lambda\mu\langle \widehat{w} - w, \Delta\rangle,
\end{aligned} \tag{5.5}$$

where $w = \dfrac{1}{\lambda}(E_\lambda(u) - T_\lambda(u))$ and $\widehat{w} = \dfrac{1}{\mu}(E_\mu(v) - \widehat{T}_\mu(v))$. On the other hand,

$$-\lambda\mu\|T_\lambda(u) - \widehat{T}_\mu(v)\|^2 = -\lambda\mu\langle u - \widehat{T}_\mu(v), \Delta\rangle + \lambda\mu\langle v - T_\lambda(u), \Delta\rangle + \lambda\mu\langle u - v, \Delta\rangle \tag{5.6}$$

From (5.5) and (5.6), we have

$$\begin{aligned}
[\theta(\lambda + \mu) - \lambda\mu]\|T_\lambda(u) - T_\mu(v)\|^2 &\leq \mu(\theta - \lambda)\langle v - T_\lambda(u), \Delta\rangle - \lambda(\theta - \mu)\langle u - \widehat{T}_\mu(v), \Delta\rangle \\
&+ \lambda\mu\langle E_\theta(v) - E_\theta(u), \Delta\rangle + \theta\lambda\mu\langle \widehat{w} - w, \Delta\rangle.
\end{aligned}$$

From Cauchy-Shwartz inequality and using that E_θ is a nonexpansive operator, we have

$$\|T_\lambda(u) - \widehat{T}_\mu(v)\| \leq \alpha\|v - T_\lambda(u)\| + \beta\|u - \widehat{T}_\mu(v)\| + \gamma\|u - v\| + \gamma\theta \left\langle \widehat{w} - w, \frac{\Delta}{\|\Delta\|} \right\rangle.$$

□

Theorem 5.2 *Let $\{x_n\}_{n\in\mathbb{N}}$ and $\{\widehat{x}_m\}_{m\in\mathbb{N}}$ be two sequences generated* (5.1) *with stepsizes $\{\lambda_n\}_{n\in\mathbb{N}}$ and $\{\widehat{\lambda}_m\}_{m\in\mathbb{N}}$ respectively and $0<\lambda_n,\widehat{\lambda}_m\leq\theta$. Then, for $z\in D$ we have:*

$$\|x_n-\widehat{x}_m\|\leq\|x_0-z\|+\|\widehat{x}_0-z\|+\xi_{n,m}+\sqrt{(\sigma_n(z)-\widehat{\sigma}_m(z))^2+\tau_n(z)+\widehat{\tau}_m+\eta_{n,m}(z)},\quad(5.7)$$

where $\sigma_n(z)=\sum_{\mathrm{i}=1}^{n}\lambda_{\mathrm{i}}\|(A_{\mathrm{i}}+B)^0z\|$, $\tau_n(z)=\sum_{\mathrm{i}=1}^{n}\lambda_{\mathrm{i}}^2\|(A_{\mathrm{i}}+B)^0z\|^2$ (similarly $\widehat{\sigma}_m(z)$ and $\widehat{\tau}_m(z)$),

$$\begin{cases}\xi_{n,0} &= \xi_{0,m}=0,\quad \textit{and}\\ \xi_{n,m} &= \alpha_{n,m}\xi_{n,m-1}+\beta_{n,m}\xi_{n-1,m}+\gamma_{n,m}\xi_{n-1,m-1}+\theta\gamma_{n,m}\Theta_{n,m}.\end{cases}\quad(5.8)$$

$$\begin{cases}\eta_{n,0}(z) &= \eta_{0,m}(z)=0,\quad \textit{and}\\ \eta_{n,m}(z) &= \alpha_{n,m}\eta_{n,m-1}(z)+\beta_{n,m}\eta_{n-1,m}(z)+\gamma_{n,m}\eta_{n-1,m-1}(z)\\ &+\ 2\theta\gamma_{n,m}\left(\sigma_n(z)-\widehat{\sigma}_m(z)\right)\left[\|(\widehat{A}_m+B)^0(z)\|-\|(A_n+B)^0(z)\|\right].\end{cases}\quad(5.9)$$

Here, $\Theta_{n,m}\geq 0$ is given by (5.3)

Proof. For $z\in D$, let

$$C_{n,m}(z)=\sqrt{(\sigma_n(z)-\widehat{\sigma}_m(z))^2+\tau_n(z)+\widehat{\tau}_m(z)+\eta_{n,m}(z)}.\quad(5.10)$$

For x_1 the first iteration of (5.1), $v_1=-\dfrac{x_1-E_{\lambda_1}(x_0)}{\lambda_1}$, $z\in D$ and using (2.1), we have:

$$\begin{aligned}\|x_1-z\| &\leq \|x_1-z+\lambda_1(v_1-A_1^0z)\|\\ &\leq \|x_0-z\|+\lambda_1\|(A_1+B)^0z\|\\ &= \|x_0-z\|+\sigma_1(z),\end{aligned}$$

where the second inequality holds because E_{λ_1} is nonexpansive. Inductively we have

$$\|x_n-z\|\leq\|x_0-z\|+\sigma_n(z)\leq\|x_0-z\|+C_{n,0}(z).\quad(5.11)$$

for all $z\in D$ and $n\in\mathbb{N}$. Now, from (5.11), we have

$$\|x_n-\widehat{x}_0\|\leq\|x_0-z\|+\|\widehat{x}_0-z\|+C_{n,0}(z),\quad(5.12)$$

for each $z\in D$ and the inequality (5.7) holds for $(n,0)$. In a similar fashion we can prove the inequality (5.7) holds for $(0,m)$, for $m\geq 0$.

For the inductive step, we assume the inequality (5.7) holds for the pairs $(n-1,m-1)$, $(n,m-1)$ and $(n-1,m-1)$, and prove that it also holds for the pair (k,l). To this end, we use the inequality (5.4) with $u=x_{n-1}$, $v=\widehat{x}_{m-1}$, $\lambda=\lambda_n$ and $\mu=\widehat{\lambda}_m$:

$$\begin{aligned}\|x_n-\widehat{x}_m\| &\leq \alpha_{n,m}\|x_n-\widehat{x}_{m-1}\|+\beta_{n,m}\|x_{n-1}-\widehat{x}_m\|+\gamma_{n,m}\|x_{n-1}-\widehat{x}_{m-1}\|+\gamma_{n,m}\\ &+\ \gamma_{n,m}\theta\left\langle w_n-\widehat{w}_m,\frac{\Delta_{n,m}}{\|\Delta_{n,m}\|}\right\rangle,\end{aligned}$$

where $\alpha_{n,m}$, $\beta_{n,m}$, $\gamma_{n,m}$ are given by (5.4). Since $\left[T_{\lambda_n}(x_{n-1}), \dfrac{E_{\lambda_n}(x_{n-1}) - T_{\lambda_n}(x_{n-1})}{\lambda_n}\right] \in A_n$ and $\left[\widehat{T}_{\widehat{\lambda}_m}(\widehat{x}_{m-1}), \dfrac{\widehat{E}_{\lambda_m}(\widehat{x}_{m-1}) - \widehat{T}_{\widehat{\lambda}_m}(\widehat{x}_{m-1})}{\widehat{\lambda}_m}\right] \in \widehat{A}_m$, then there exist $\Theta_{n,m} > 0$ such that, $\left\langle w_n - \widehat{w}_m, \frac{\Delta_{n,m}}{\|\Delta_{n,m}\|}\right\rangle \leq \Theta_{n,m}$ for all $m, n \geq 0$ (see [1]). Using induction hypothesis, we have

$$\begin{aligned}
\|x_n - \widehat{x}_m\| &\leq \alpha_{n,m}\|x_n - \widehat{x}_{m-1}\| + \beta_{n,m}\|x_{n-1} - \widehat{x}_m\| + \gamma_{n,m}\|x_{n-1} - \widehat{x}_{m-1}\| + \gamma_{n,m}\theta\Theta_{n,m} \\
&\leq (\alpha_{n,m} + \beta_{n,m} + \gamma_{n,m})(\|x_0 - z\| + \|\widehat{x}_0 - z\|) \\
&+ (\alpha_{n,m}\xi_{n,m-1} + \beta_{n,m}\xi_{n-1,m} + \gamma_{n,m}\xi_{n-1,m-1} + \theta\gamma_{n,m}\Theta_{n,m}) \\
&+ (\alpha_{n,m}C_{n,m-1}(z) + \beta_{n,m}C_{n-1,m}(z) + \gamma_{n,m}C_{n-1,m-1}(z)) \\
&= \|x_0 - z\| + \|\widehat{x}_0 - z\| + (\alpha_{n,m}\xi_{n,m-1} + \beta_{n,m}\xi_{n-1,m} + \gamma_{n,m}\xi_{n-1,m-1} + \theta\gamma_{n,m}\Theta_{n,m}) \\
&+ (\alpha_{n,m}C_{n,m-1}(z) + \beta_{n,m}C_{n-1,m}(z) + \gamma_{n,m}C_{n-1,m-1}(z)) \qquad (5.13)
\end{aligned}$$

On the other hand

$$\begin{aligned}
\alpha_{n,m}C_{n,m-1}(z) + \beta_{n,m}C_{n-1,m}(z) + \gamma_{n,m}C_{n-1,m-1}(z) &= \alpha_{n,m}^{\frac{1}{2}}(\alpha_{n,m}^{\frac{1}{2}}C_{n,m-1}(z)) + \beta_{n,m}^{\frac{1}{2}}(\beta_{n,m}^{\frac{1}{2}}C_{n-1,m}(z)) \\
&+ \gamma_{n,m}^{\frac{1}{2}}(\gamma_{n,m}^{\frac{1}{2}}C_{n-1,m-1}(z)) \\
&\leq (\alpha_{n,m}C^2_{n,m-1}(z) + \beta_{n,m}C^2_{n-1,m}(z) \\
&+ \gamma_{n,m}C^2_{n-1,m-1}(z))^{\frac{1}{2}}.
\end{aligned}$$

Let $\lambda_n(z) = \lambda_n\|(A_n + B)^0 z\|$ and $\widehat{\lambda}_m(z) = \widehat{\lambda}_m\|(\widehat{A}_m + B)^0 z\|$. Notice that

$$C^2_{n,m-1}(z) = (\sigma_n(z) - \widehat{\sigma}_m(z))^2 + \tau_n(z) + \widehat{\tau}_m(z) + \eta_{n,m-1}(z) + \widehat{\lambda}_m(z)(\sigma_n(z) - \widehat{\sigma}_m(z)), \quad (5.14)$$
$$C^2_{n-1,m}(z) = (\sigma_n(z) - \widehat{\sigma}_m(z))^2 + \tau_n(z) + \widehat{\tau}_m(z) + \eta_{n-1,m}(z) - \lambda_n(z)\left(\sigma_n(z) - \widehat{\sigma}_m(z)\right), \quad (5.15)$$

and

$$\begin{aligned}
C^2_{n-1,m-1}(z) &= (\sigma_n(z) - \widehat{\sigma}_m(z))^2 + \tau_n(z) + \widehat{\tau}_m(z) + \eta_{n-1,m-1}(z) \\
&- 2\left(\lambda_n(z) - \widehat{\lambda}_m(z)\right)(\sigma_n(z) - \widehat{\sigma}_m(z)) - 2\widehat{\lambda}_m(z)\lambda_n(z) \\
&\leq (\sigma_n(z) - \widehat{\sigma}_m(z))^2 + \tau_n(z) + \widehat{\tau}_m(z) + \eta_{n-1,m-1}(z) \\
&- 2\left(\lambda_n(z) - \widehat{\lambda}_m(z)\right)(\sigma_n(z) - \widehat{\sigma}_m(z)). \qquad (5.16)
\end{aligned}$$

Replacing (5.14), (5.15), (5.16) in (5.13) and setting

$$r_{n,m} = (\alpha_{n,m}C_{n,m-1}(z) + \beta_{n,m}C_{n-1,m}(z) + \gamma_{n,m}C_{n-1,m-1}(z))^2,$$

we have:

$$
\begin{aligned}
r_{n,m} &\leq (\sigma_n(z) - \widehat{\sigma}_m(z))^2 + \tau_n(z) + \widehat{\tau}_m + \alpha_{n,m}\eta_{n,m-1}(z) + \beta_{n,m}\eta_{n,m-1}(z) + \gamma_{n,m}\eta_{n-1,m-1}(z) \\
&+ 2(\sigma_n(z) - \widehat{\sigma}_m(z))\left[\widehat{\lambda}_m(z)(\alpha_{n,m} + \gamma_{n,m}) - \lambda_n(z)(\gamma_{n,m} + \beta_{n,m})\right] \\
&= (\sigma_n(z) - \widehat{\sigma}_m(z))^2 + \tau_n(z) + \widehat{\tau}_m + \alpha_{n,m}\eta_{n,m-1}(z) + \beta_{n,m}\eta_{n,m-1}(z) \\
&+ \gamma_{n,m}\eta_{n-1,m-1}(z) + \frac{2\theta}{\theta(\lambda_n + \widehat{\lambda}_m) - \lambda_n\widehat{\lambda}_m}\left[\lambda_n\widehat{\lambda}(z) - \widehat{\lambda}_m\lambda_n(z)\right] \\
&\leq (\sigma_n(z) - \widehat{\sigma}_m(z))^2 + \tau_n(z) + \widehat{\tau}_m + \alpha_{n,m}\eta_{n,m-1}(z) + \beta_{n,m}\eta_{n,m-1}(z) \\
&+ \gamma_{n,m}\eta_{n-1,m-1}(z) + \frac{2\theta\lambda_n\widehat{\lambda}_m}{\theta(\lambda_n + \widehat{\lambda}_m) - \lambda_n\widehat{\lambda}_m}\left[\|(\widehat{A}_m + B)^0(z)\| - \|(A_n + B)^0(z)\|\right] \\
&= (\sigma_n(z) - \widehat{\sigma}_m(z))^2 + \tau_n(z) + \widehat{\tau}_m + \alpha_{n,m}\eta_{n,m-1}(z) + \beta_{n,m}\eta_{n,m-1}(z) \\
&+ \gamma_{n,m}\eta_{n-1,m-1}(z) + 2\theta\gamma_{n,m}\left[\|(\widehat{A}_m + B)^0(z)\| - \|(A_n + B)^0(z)\|\right] \\
&= C^2_{n,m}(z).
\end{aligned}
$$

Therefore

$$\|x_n - \widehat{x}_m\| \leq \|x_0 - z\| + \|\widehat{x}_0 - z\| + \xi_{n,m} + C_{n,m}(z), \tag{5.17}$$

for all $z \in D$ and $n, m \in \mathbb{N}$. □

Remark 5.3 When $A = \widehat{A}$ and no dependent of a temporal parameter $t \geq 0$, (5.1) becomes to forward–backward scheme (4.3). Then we recover the Kobayashi's inequality (4.11) with $\varepsilon_n = \widehat{\varepsilon}_m = 0$ for all $n, m \in \mathbb{N}$ in Hilbert spaces setting: for $z \in D(A)$ and $n, m \in \mathbb{N}$

$$\|x_n - \widehat{x}_m\| \leq \|x_0 - z\| + \|\widehat{x}_0 - z\| + (\|(A + B)^0 z\|)\sqrt{(\sigma_n - \widehat{\sigma}_m)^2 + \tau_n + \widehat{\tau}_m}. \tag{5.18}$$

The inequality could reveal us the relationship between the corresponding continuous and discrete systems in a finite and infinite horizon. However, the inequality (5.7) has some terms expressed implicitly, which makes its implementation difficult. We plan to use combinatorial techniques to obtain estimates of these implicit terms in (5.7). This work represents a first interesting challenge, which could be developed in a post-doctoral project.

Part II

Forward Backward and Primal Dual algorithms

Foreword

In this second part of the manuscript, we show some results obtained on the acceleration of certain primal–dual algorithms for image processing by using the asymptotic properties of the preconditioned forward–backward algorithm.

Using the information about the asymptotic behavior of overrelaxed forward–backward algorithm, in Chapter 6 we deal specifically with the algorithms presented by Loris–Verhoeven [53] and Condat–Vũ [27, 69], when the cocoercive operator in both algorithms is affine. Finally, some numerical experiments related to the implementation of these algorithms in image recovery will be presented. This work was developed during a visit research at GIPSA-LAB of the Grenoble–Alpes University, under supervision of Professor Laurent Condat.

The Chapter 7 is dedicated to show a second current work on the convergence of the gradient and proximal–gradient algorithms under assumptions in [72], instead of the usual Lipschitz condition. This problem was proposed by proffesor Hong-Kun Xu during my visit research to the Hangzhou-Dianzi University in Hangzhou, China.

Chapter 6

Relaxed Forward–Backward Splitting and some Primal–Dual Algorithms

A joint work with L. Condat, D. Kitahara and A. Hirabayashi [26]

6.1 The forward–backward iteration

Throughout this chapter, H will be a real Hilbert space endowed with the inner product $\langle \cdot, \cdot \rangle$, $A : H \to 2^H$ a maximally monotone operator and $B : H \to H$ a θ-cocoercive operator, for some real $\theta > 0$. The forward–backward algorithm, proposed by Mercier [55] and further developed by many authors [51, 37, 68, 18, 23, 25], allows to approach the solutions of the monotone inclusion

$$0 \in Az + Bz. \tag{6.1}$$

For $z_0 \in H$ an initial estimate of a solution and $\gamma > 0$ a real parameter, the classical forward–backward iteration [51, 62, 8] to find a solution to (6.1), is given by:

$$\begin{array}{l} \text{For } n = 0, 1, \ldots \\ \lfloor\; z_{n+1} = J_\gamma^A (z_n - \gamma B z_n)\ , \end{array} \tag{6.2}$$

A not so well known iterative method to approximate solution of (6.1) is the *overrelaxed forward–backward* algorithm. The geometric intuition is quite simple: since z_{n+1} is closer to a solution than z_n, it is natural, starting at z_n, to move further in the direction $z_{n+1} - z_n$, which improves the estimate. This yields the relaxed forward–backward iteration. From now on, we will omit the word "relaxed" for simplicity.

For $\{\rho_n\}_{n\in\mathbb{N}}$ a sequence of relaxation parameters, the iterations are described as follow:

Forward–Backward iteration for (6.1)**:** for n=0,1,...

$$\left\lfloor \begin{array}{l} z_{n+\frac{1}{2}} = J_\gamma^A\Big(z_n - \gamma B z_n\Big) \\ z_{n+1} = z_n + \rho_n\Big(z_{n+\frac{1}{2}} - z_n\Big) \end{array} \right. . \tag{6.3}$$

While there is no interest in doing underrelaxation with ρ_n less than 1, it is expected that convergence of (6.3) is faster if doing overrelaxation with ρ_n larger than 1; this is what is most often observed in practice. On the other hand, the explicit mapping from z_n to $z_{n+\frac{1}{2}}$ in (6.3) can be equivalently written under the implicit form:

$$0 \in A z_{n+\frac{1}{2}} + B z_n + \frac{1}{\gamma}\Big(z_{n+\frac{1}{2}} - z_n\Big). \tag{6.4}$$

The standard convergence for the forward–backward iteration is given by the following result from [27, Lemma 4.4], see also [8, Theorem 26.14].

Theorem 6.1 ***(Forward–Backward algorithm*** (6.3)***)*** *Suppose* $0 < \gamma < 2\theta$ *and* $z_0 \in H$. *Set* $\delta = 2 - \gamma/(2\theta)$. *Suppose* $\{\rho_n\}_{n\in\mathbb{N}}$ *is a sequence in* $[0, \delta]$ *such that* $\sum_{n\in\mathbb{N}} \rho_n(\delta - \rho_n) = +\infty$. *Then the sequence* $\{z_n\}_{n\in\mathbb{N}}$ *defined by the iteration* (6.3) *converges weakly to a solution of* (6.1).

The proof relies on the Krasnosel'skiĭ–Mann theorem [8, Proposition 5.16 and Proposition 5.34], and the fact that the operator $T = J_\gamma^A \circ (\mathrm{I} - \gamma B)$, which maps z_n to $z_{n+\frac{1}{2}}$, is $(1/\delta)$-averaged. Moreover, the fixed points of T are the solutions to (6.1), as is clearly visible in (6.4).

Remark 6.2 If γ is close to 2θ, δ is close to 1, so that overrelaxation cannot be used. This explains why the relaxed forward–backward iteration is not so well known.

Now, let P be a bounded, self-adjoint, strongly positive, linear operator on H. Clearly, solving (6.1) is equivalent to solving

$$0 \in P^{-1} A z + P^{-1} B z. \tag{6.5}$$

Let H_P be the Hilbert space obtained by endowing H with the inner product $\langle x, x' \rangle_P = \langle x, P x' \rangle$, $(x, x' \in H)$. According with [8, Proposition 20.24], $P^{-1}A$ is maximally monotone in H_P. However, the cocoercivity of $P^{-1}B$ in H_P has to be checked on a case-by-case basis.

The *preconditioned* forward–backward iteration to solve (6.5) is

Preconditioned Forward–Backward iteration for (6.5) for $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} z_{n+\frac{1}{2}} = J_1^{P^{-1}A}\Big(z_n - P^{-1} B z_n\Big) \\ z_{n+1} = z_n + \rho_n\Big(z_{n+\frac{1}{2}} - z_n\Big) \end{array} \right. . \tag{6.6}$$

In the case when $P^{-1}B$ is a cocoercive operator, the corresponding convergence result follows:

Theorem 6.3 ***(Preconditioned Forward–Backward algorithm*** (6.6)***)*** *Suppose that* $P^{-1}B$ *is* χ*-cocoercive in* H_P, *with* $\chi > \frac{1}{2}$. *Set* $\delta = 2 - 1/(2\chi)$. *Suppose that* $\{\rho_n\}_{n\in\mathbb{N}}$ *is a sequence in* $[0, \delta]$ *such that* $\sum_{n\in\mathbb{N}} \rho_n(\delta - \rho_n) = +\infty$. *Then the sequence* $\{z_n\}_{n\in\mathbb{N}}$ *defined by the iteration* (6.6) *converges weakly to a solution of* (6.1).

Proof. This is a straightforward application of Theorem 6.1 in H_P instead of H, with $\gamma = 1$. Weak convergence in H_P is equivalent to weak convergence in H, and the solution sets of (6.1) and (6.5) are the same. □

Remark 6.4 The explicit mapping from z_n to $z_{n+\frac{1}{2}}$ in (6.6) can be equivalently written under the implicit form:

$$0 \in Az_{n+\frac{1}{2}} + Bz_n + P\Big(z_{n+\frac{1}{2}} - z_n\Big). \tag{6.7}$$

If $B = 0$, the forward–backward iteration reduces to the proximal point algorithm [67, 8]. In that case, weak convergence to a zero of A is obtained with any $\gamma > 0$ and $\delta = 2$, in the notations of Theorem 6.3. Let us give a formal convergence statement for the more general preconditioned proximal point algorithm. The problem is to solve

$$0 \in Az, \tag{6.8}$$

when the solution set is supposed nonempty. Now, let P be a bounded, self-adjoint, strongly positive, linear operator on H. Let $z_0 \in H$ and let $\{\rho_n\}_{n\in\mathbb{N}}$ be a sequence of relaxation parameters. The relaxed and preconditioned proximal point algorithm is the iteration:

Proximal point iteration for (6.8): for $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} z_{n+\frac{1}{2}} = J_1^{P^{-1}A} z_n \\ z_{n+1} = z_n + \rho_n\Big(z_{n+\frac{1}{2}} - z_n\Big) \end{array} \right. . \tag{6.9}$$

The convergence of the preconditioned proximal point algorithm can be stated from Theorem 6.3 as follow:

Theorem 6.5 ***(Proximal point iteration** (6.9)**)*** *Let $z_0 \in H$ and suppose that $\{\rho_n\}_{n\in\mathbb{N}}$ is a sequence in $[0, 2]$ such that $\sum_{n\in\mathbb{N}} \rho_n(2 - \rho_n) = +\infty$. Then the sequence $\{z_n\}_{n\in\mathbb{N}}$ defined by the iteration* (6.9)*, converges weakly to a solution of* (6.8)*.*

Proof. $P^{-1}A$ is maximally monotone operator in H_P [8, Proposition 20.24]. Thus, its resolvent $J_1^{P^{-1}A}$ is firmly nonexpansive in H_P and by virtue of the Krasnosel'skiĭ–Mann theorem [8, Proposition 5.16 and Proposition 5.34], the sequence $\{z_n\}_{n\in\mathbb{N}}$ converges weakly in H_P to some $z^* \in H$ with $0 \in P^{-1}Az^*$, so that $0 \in Az^*$. □

The case where B is affine

Let us suppose that, in addition to being θ–cocoercive, B is an affine operator. That means, $B : z \in H \mapsto Qz + c$, for some bounded, self-adjoint, positive, nonzero, linear operator Q on H and some element $c \in H$. Then, we can write (6.4) as

$$0 \in (A + B)z_{n+\frac{1}{2}} + P(z_{n+\frac{1}{2}} - z_n), \tag{6.10}$$

with

$$P = \frac{1}{\gamma}\mathrm{I} - Q.$$

Thus, the forward–backward iteration (6.6) can be interpreted as a preconditioned proximal point iteration (6.9) when $P = \frac{1}{\gamma}\mathrm{I} - Q$, applied to find a zero of $A + B$. Since P must be strongly positive, we must have $0 < \gamma < \theta$, so that the admissible range for γ is halved. But in return, we get the larger range $[0, 2]$ for relaxation. Hence, we have the following new convergence result:

Theorem 6.6 ***(Forward–Backward algorithm*** (6.3)***, affine case)*** *Suppose* $0 < \gamma < \theta$, $z_0 \in H$ *and* $\{\rho_n\}_{n\in\mathbb{N}}$ *is a sequence in* $[0, 2]$ *such that* $\sum_{n\in\mathbb{N}} \rho_n(2 - \rho_n) = +\infty$. *Then the sequence* $\{z_n\}_{n\in\mathbb{N}}$ *defined by the iteration* (6.9) *converges weakly to a solution of* (6.1).

Proof. In view of (6.10) and (6.9), this is Theorem 6.5 applied to the problem (6.8) with $A + B$, which is maximally monotone. □

Remark 6.7 If $\gamma = 0.99\theta$, we can set $\rho_n = 1.5$ according with Theorem 6.1. Now, From Theorem 5.6, we can do better and set $\rho_n = 1.99$. So, which value of γ should be used in practice? If Q is badly conditioned and we do not use overrelaxation ($\rho_n = 1$), γ close to 2θ is probably the best choice. Now, if Q is well conditioned and we still do not use overrelaxation, $\gamma = \theta$ may be a better choice. Indeed, if $Q = (1/\theta)\mathrm{I}$ and $\gamma = \theta$, then $z_1 = J_\gamma^A(z_0 - \gamma B z_0) = J_\gamma^A(0)$ and z_1 is a solution to (6.1), so that the algorithm converges in one iteration. In any case, a value of γ less than θ is not interesting. The recommendation for a given practical problem is to try the two settings: $\gamma = (2 - \varepsilon)\theta$, for a small $\varepsilon > 0$, and $\rho_n = 1$ on one hand, $\gamma = (1 - \varepsilon)\theta$ and $\rho_n = 2 - \varepsilon$ on the other hand.

Applications to convex optimization

In the following, we denote by $\Gamma_0(H)$ the set of convex, proper, lower semicontinuous functions from H to $\mathbb{R} \cup \{+\infty\}$ [8]. Let $f, h \in \Gamma_0(H)$ and suppose that h is a differentiable function with β–Lipschitz continuous gradient ∇h, for some real $\beta > 0$. We define the *proximity operator* of f [58] as

$$\mathrm{prox}_f(x) = \underset{x'\in H}{\arg\min} \left(f(x') + \frac{1}{2}\|x - x'\|^2 \right).$$

It is well known that for every $x \in H$, $\mathrm{prox}_f(x) = (I + \partial f)^{-1}(x)$, where ∂f is the subdifferential of f [8]. There are fast and exact methods to compute the proximity operator of a large class of functions [17, 20, 35]. Let us consider the convex optimization problem

$$\underset{x\in H}{\mathrm{minimize}}\ f(x) + h(x), \tag{6.11}$$

whose solution set is supposed nonempty. The well known Fermat's rule [8, Theorem 27.2] states that the problem (6.11) is equivalent to (6.1) with $A = \partial f$, which is maximally monotone, and $B = \nabla h$, which is θ-cocoercive, with $\theta = 1/\beta$ [8, Corollary 18.17]. Hence, it is natural to use the forward–backward iteration to approximate the solutions of (6.11): for $\gamma > 0$, $x_0 \in H$ and $\{\rho_n\}_{n\in\mathbb{N}}$ a sequence of relaxation parameters, the forward–backward iteration in this case is given by:

Proximal–gradient iteration for (6.11): for $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} x_{n+\frac{1}{2}} = \operatorname{prox}_{\gamma f}\Big(x_n - \gamma\nabla h(x_n)\Big) \\ x_{n+1} = x_n + \rho_n\Big(x_{n+\frac{1}{2}} - x_n\Big) \end{array} \right. . \tag{6.12}$$

As a direct consequence of Theorem 6.1, we have:

Theorem 6.8 ***(Proximal–gradient algorithm*** (6.12)***)*** *Let $0 < \gamma < 2$ and $x_0 \in H$. Set $\delta = 2-\gamma\beta/2$ and suppose that $\{\rho_n\}_{n\in\mathbb{N}}$ is a sequence in $[0, \delta]$ such that $\sum_{n\in\mathbb{N}} \rho_n(\delta-\rho_n) = +\infty$. Then the sequence $\{x_n\}_{n\in\mathbb{N}}$ defined by the iteration* (6.12) *converges weakly to a solution of* (6.11).

Now, we shall focus on the case where h is quadratic:

$$h : x \mapsto \tfrac{1}{2}\langle x, Qx\rangle + \langle x, c\rangle, \tag{6.13}$$

for some bounded, self-adjoint, positive, nonzero, linear operator Q on H and some element $c \in H$. A very common example is a least-squares penalty, in particular to solve inverse problems, that is,

$$h : x \mapsto \tfrac{1}{2}\|Kx - y\|^2, \tag{6.14}$$

for some bounded linear operator K from H to a real Hilbert space Y and some element $y \in Y$. Clearly, (6.14) is an instance of (6.13) with $Q = K^*K$, where K^* is the adjoint of K, and $c = K^*y$. In this case, for every $x \in H$, we have

$$\nabla h(x) = Qx + c, \tag{6.15}$$

with $\beta = \|Q\|$. Setting $P = \frac{1}{\gamma}\mathrm{I} - Q$, we can remark that the update in (6.12) can be written as

$$x_{n+\frac{1}{2}} = \operatorname*{arg\,min}_{x\in H} f(x) + h(x) + \tfrac{1}{2}\|x - x_n\|_P^2,$$

where we introduce the norm $\|\cdot\|_P : x \mapsto \sqrt{\langle x, Px\rangle}$. So, $x_{n+\frac{1}{2}}$ can be viewed as being obtained by applying the proximity operator of $f + h$ with the preconditioned norm $\|\cdot\|_P$. Hence, as a direct consequence of Theorem 6.6, we have:

Theorem 6.9 ***(Proximal–gradient algorithm, quadratic case)*** *Let $0 < \gamma < 1/\beta$ and $x_0 \in H$. Suppose that $\{\rho_n\}_{n\in\mathbb{N}}$ is a sequence in $[0, 2]$ such that $\sum_{n\in\mathbb{N}} \rho_n(2 - \rho_n) = +\infty$. Then, the sequence $\{x_n\}_{n\in\mathbb{N}}$ defined by the iteration* (6.12) *converges weakly to a solution of* (6.11).

6.2 The Loris–Verhoeven iteration

For U a real Hilbert space, let $g \in \Gamma_0(U)$, $h : H \to \mathbb{R}$ be a convex and differentiable function with β–Lipschitz continuous gradient ∇h, for some real $\beta > 0$ and let $L : H \to U$ be a bounded and linear operator. Often, the template problem (6.11) of minimizing the sum of two functions is too simple and we would like, instead, to sove the following optimization problem

$$\underset{x\in H}{\text{minimize}}\; g(Lx)+h(x), \tag{6.16}$$

where the solution set is supposed nonempty. We assume that there is no simple way to compute the proximity operator of $g\circ L$. The problem (6.16) is equivalent to the monotone inclusion

$$0\in L^*\partial g(Lx)+\nabla h(x).$$

To get rid of the annoying operator L, we can introduce an auxiliary variable $u\in\partial g(Lx)$, which shall be called the *dual* variable, so that the problem now consists in finding $x\in H$ and $u\in U$ such that

$$\begin{cases} u\in\partial g(Lx) \\ 0\in L^*u+\nabla h(x) \end{cases}.$$

The interest in increasing the dimension of the problem is that we obtain a system of two monotone inclusions, which are decoupled: ∂g and ∇h appear separately in the two inclusions. Then, equivalently, the problem is to find a pair of objects $z=(x,u)$ in $Z=H\times U$ such that

$$\begin{pmatrix}0\\0\end{pmatrix}\in\underbrace{\begin{pmatrix}L^*u\\ -Lx+(\partial g)^{-1}u\end{pmatrix}}_{Az}+\underbrace{\begin{pmatrix}\nabla h(x)\\0\end{pmatrix}}_{Bz}. \tag{6.17}$$

The operator $A:Z\to 2^Z,(x,u)\mapsto(L^*u,-Lx+(\partial g)^{-1}u)$ is maximally monotone [8, Proposition 26.32 (iii)] and $B:Z\to Z,(x,u)\mapsto(\nabla h(x),0)$ is θ-cocoercive, with $\theta=1/\beta$. Thus, it is natural to think of the forward–backward iteration to solve the problem (6.17). However, to make the resolvent of A computable with the proximity operator of g, we need preconditioning. The solution consists in the iteration, that we first write in implicit form:

$$\begin{pmatrix}0\\0\end{pmatrix}\in\underbrace{\begin{pmatrix}L^*u_{n+\frac12}\\ -Lx_{n+\frac12}+(\partial g)^{-1}u_{n+\frac12}\end{pmatrix}}_{A_{n+\frac12}}+\underbrace{\begin{pmatrix}\nabla h(x_n)\\0\end{pmatrix}}_{Bz_n}+\underbrace{\begin{pmatrix}\frac{1}{\tau}\mathrm{I} & 0\\ 0 & \frac{1}{\sigma}\mathrm{I}-\tau LL^*\end{pmatrix}}_{P}\underbrace{\begin{pmatrix}x_{n+\frac12}-x_n\\ u_{n+\frac12}-u_n\end{pmatrix}}_{z_{n+\frac12}-z_n}, \tag{6.18}$$

where $\tau>0$ and $\sigma>0$ are two real parameters, $z_n=(x_n,u_n)$ and $z_{n+\frac12}=(x_{n+\frac12},u_{n+\frac12})$. It is not straightforward to see that this yields an explicit iteration. The key is to remark that we have

$$x_{n+\frac12}=x_n-\tau\nabla h(x_n)-\tau L^*u_{n+\frac12} \tag{6.19}$$

so that we can update the primal variable $x_{n+\frac12}$, once the dual variable $u_{n+\frac12}$ is available. Thus, the first step of the algorithm is to construct $u_{n+\frac12}$. It depends on $Lx_{n+\frac12}$, which is not yet available, but using (6.19), we can express it using x_n and $LL^*u_{n+\frac12}$. This last term is canceled in the preconditioner P to make the update explicit. That is,

$$\begin{aligned}
&0 \in -Lx_{n+\frac{1}{2}} + (\partial g)^{-1}u_{n+\frac{1}{2}} + (\tfrac{1}{\sigma}\mathrm{I} - \tau LL^*)(u_{n+\frac{1}{2}} - u_n)\\
\Leftrightarrow\ &0 \in -Lx_n + \tau L\nabla h(x_n) + \tau LL^* u_{n+\frac{1}{2}} + (\partial g)^{-1}u_{n+\frac{1}{2}} + (\tfrac{1}{\sigma}\mathrm{I} - \tau LL^*)(u_{n+\frac{1}{2}} - u_n)\\
\Leftrightarrow\ &0 \in -Lx_n + \tau L\nabla h(x_n) + \tau LL^* u_n + (\partial g)^{-1}u_{n+\frac{1}{2}} + \tfrac{1}{\sigma}(u_{n+\frac{1}{2}} - u_n)\\
\Leftrightarrow\ &\Big(\sigma(\partial g)^{-1} + \mathrm{I}\Big)u_{n+\frac{1}{2}} \ni \sigma Lx_n - \tau\sigma L\nabla h(x_n) - \tau\sigma LL^* u_n + u_n\\
\Leftrightarrow\ &u_{n+\frac{1}{2}} = \operatorname{prox}_{\sigma g^*}\Big(\sigma L\Big(x_n - \tau\nabla h(x_n)\Big) + u_n - \tau\sigma LL^* u_n\Big),
\end{aligned}$$

where we recall that the conjugate function $g^* \in \Gamma_0(U) : u \mapsto \sup_{u' \in U} \langle u, u'\rangle - g(u')$ is such that $\partial g^* = (\partial g)^{-1}$ [8, Corollary 16.30]. We also recall the Moreau identity, which allows computing the proximity operator of g^* from the one of g, and conversely [20]:

$$\operatorname{prox}_{\sigma g^*}(u) = u - \sigma \operatorname{prox}_{g/\sigma}(u/\sigma). \tag{6.20}$$

Let us define the dual convex optimization problem associated to the primal problem (6.16):

$$\underset{u \in U}{\text{minimize}}\ g^*(u) + h^*(-L^*u). \tag{6.21}$$

If a pair $(x, u) \in H \times U$ is a solution to (6.17), then x is a solution to (6.16) and u is a solution to (6.21).

Let $\tau > 0$ and $\sigma > 0$, let $x_0 \in H$ and $u_0 \in U$, and let $\{\rho_n\}_{n\in\mathbb{N}}$ be a positive sequence of relaxation parameters. The primal–dual forward–backward iteration, which we call the Loris–Verhoeven iteration is:

Loris–Verhoeven iteration for (6.16) and (6.21): for $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l}
u_{n+\frac{1}{2}} = \operatorname{prox}_{\sigma g^*}\Big(u_n + \sigma L\Big(x_n - \tau\nabla h(x_n) - \tau L^* u_n\Big)\Big)\\
x_{n+1} = x_n - \rho_n\tau\Big(\nabla h(x_n) + L^* u_{n+\frac{1}{2}}\Big)\\
u_{n+1} = u_n + \rho_n\Big(u_{n+\frac{1}{2}} - u_n\Big)
\end{array}\right. . \tag{6.22}$$

This algorithm was first proposed by Loris and Verhoeven, in the case where h is a least-squares term [53]. It was then rediscovered several times and named *Primal–Dual Fixed-Point algorithm based on the Proximity Operator* (PDFP2O) [19] or *Proximal Alternating Predictor–Corrector* (PAPC) algorithm [34]. The above interpretation of the algorithm as a primal–dual forward–backward iteration has been presented in [22]. As an application of Theorem 6.3, we obtain the following convergence result for (6.22):

Theorem 6.10 ***(Loris–Verhoeven algorithm*** (6.22)***)*** *Let $x_0 \in H$ and $u_0 \in U$. Suppose that $0 < \tau < 2/\beta$, $\sigma\tau\|L\|^2 < 1$ and set $\delta = 2 - \tau\beta/2$. Suppose that $\{\rho_n\}_{n\in\mathbb{N}}$ is a sequence in $[0, \delta]$ such that $\sum_{n\in\mathbb{N}} \rho_n(\delta - \rho_n) = +\infty$. Then the sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{u_n\}_{n\in\mathbb{N}}$ defined by the iteration* (6.22) *converge weakly to a solution of* (6.16) *and to a solution of* (6.21)*, respectively.*

Proof. In view of (6.18) and (6.7), this is the Theorem 6.3 applied to (6.17). For this, P must be strongly positive, which is the case if and only if $\sigma\tau\|L\|^2 < 1$. Moreover, $P^{-1}B$ is $1/(\tau\beta)$–cocoercive in H_P. □

The following result makes it possible to have $\sigma\tau\|L\|^2 = 1$, which is a consequence of the analysis in [60] of the PD3O algorithm [73]. See also [19, Theorem 3.4 and Theorem 3.5] for the same result, but without relaxation.

Theorem 6.11 ***(Loris–Verhoeven algorithm*** (6.22)***)*** *Suppose that H and U are of finite dimension and let $x_0 \in H$ and $u_0 \in U$. Suppose that $0 < \tau < 2/\beta$, $\sigma\tau\|L\|^2 \leq 1$, and $\rho_n = 1$, for all $n \in \mathbb{N}$. Then the sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{u_n\}_{n\in\mathbb{N}}$ defined by the iteration* (6.22) *converge to a solution of* (6.16) *and to a solution of* (6.21)*, respectively.*

Remark 6.12 According to Theorem 6.11, if $H = U$ has finite dimension, $L = \mathrm{I}$, we can set $\sigma = 1/\tau$ and the Loris–Verhoeven iteration becomes

$$\left\lfloor \begin{array}{l} u_{n+\frac{1}{2}} = \mathrm{prox}_{g^*/\tau}\Big(x_n/\tau - \nabla h(x_n)\Big) \\ x_{n+\frac{1}{2}} = x_n - \tau\nabla h(x_n) - \tau u_{n+1} \\ \qquad\;\; = \mathrm{prox}_{\tau g}\Big(x_n - \tau\nabla h(x_n)\Big) \\ x_{n+1} = x_n + \rho_n(x_{n+\frac{1}{2}} - x_n) \\ u_{n+1} = u_n + \rho_n(u_{n+\frac{1}{2}} - u_n) \end{array}\right. . \tag{6.23}$$

Then, we can discard the dual variable and we recover the forward–backward iteration (6.12). It is interesting that in this primal algorithm, there is an implicit dual variable $u_{n+1} = -\nabla h(x_n) + (x_n - x_{n+1})/\tau$, which converges to a solution of the dual problem; that is, to a minimizer of $h^*(-u) + g^*(u)$.

Again, let us focus on the case where h is quadratic; that is, $h : x \mapsto \frac{1}{2}\langle x, Qx\rangle + \langle x, c\rangle$, for some bounded, self-adjoint, positive, nonzero, linear operator Q on H and $c \in H$. We have $\beta = \|Q\|$ and we can rewrite the primal–dual inclusion (6.18), which characterizes the Loris–Verhoeven iteration 6.22, as

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \underbrace{\begin{pmatrix} \nabla h(x_{n+\frac{1}{2}}) + L^* u_{n+\frac{1}{2}} \\ -Lx_{n+\frac{1}{2}} + (\partial g)^{-1} u_{n+\frac{1}{2}} \end{pmatrix}}_{Az_{n+\frac{1}{2}}} + \underbrace{\begin{pmatrix} \frac{1}{\tau}\mathrm{I} - Q & 0 \\ 0 & \frac{1}{\sigma}\mathrm{I} - \tau LL^* \end{pmatrix}}_{P} \underbrace{\begin{pmatrix} x_{n+\frac{1}{2}} - x_n \\ u_{n+\frac{1}{2}} - u_n \end{pmatrix}}_{z_{n+\frac{1}{2}} - z_n}. \tag{6.24}$$

As an application of Theorem 6.6, we have:

Theorem 6.13 ***(Loris–Verhoeven algorithm*** (6.22)***, quadratic case)*** *Let $x_0 \in H$ and $u_0 \in U$. Suppose that $0 < \tau < 1/\beta$ and that $\sigma\tau\|L\|^2 < 1$. Suppose that $\{\rho_n\}_{n\in\mathbb{N}}$ is a sequence in $[0, 2]$ such that $\sum_{n\in\mathbb{N}} \rho_n(2 - \rho_n) = +\infty$. Then the sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{u_n\}_{n\in\mathbb{N}}$ defined by the iteration* (6.22) *converge weakly to a solution of* (6.16) *and to a solution of* (6.21)*, respectively.*

6.3 The Condat–Vũ iteration

Let us consider the primal optimization problem:

$$\underset{x\in H}{\text{minimize}}\ f(x) + g(Lx) + h(x), \tag{6.25}$$

where $f \in \Gamma_0(H)$, $g \in \Gamma_0(U)$, $h : H \to \mathbb{R}$ is convex and differentiable function with β-Lipschitz continuous gradient ∇h, for some real $\beta > 0$ and $L : H \to U$ is a bounded linear operator. Roughly speaking, the problem (6.25) is equivalent to solving

$$0 \in \partial f(x) + L^*\partial g(Lx) + \nabla h(x), \tag{6.26}$$

where the solution set is supposed nonempty. More precisely, a solution to (6.26) is a solution to (6.25), but the converse may not be true. Under mild qualification constraints on f, g, h, however, the solution sets of (6.25) and (6.26) are the same. For instance, this is the case if

$$0 \in \operatorname{sri}\left(L(\operatorname{dom}(f)) - \operatorname{dom}(g)\right), \tag{6.27}$$

where sri denotes the strong relative interior. See [8, Proposition 27.5, Corollary 27.6] for other examples of qualification constraints. Like in the previous sections, let us introduce a dual variable u, so that we can rewrite the problem (6.26) as the search of a pair of objects $z = (x, u)$ in $Z = H \times U$ such that such that

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \underbrace{\begin{pmatrix} \partial f(x) + L^*u \\ -Lx + (\partial g)^{-1}u \end{pmatrix}}_{Az} + \underbrace{\begin{pmatrix} \nabla h(x) \\ 0 \end{pmatrix}}_{Bz}. \tag{6.28}$$

A pair $(x, u) \in Z$ is a solution to (6.28) if and only if x is a solution to (6.26) and $u \in \partial g(Lx)$ is a solution to the dual problem associated to (6.25):

$$\underset{u\in U}{\text{minimize}}\,(f + h)^*(-L^*u) + g^*(u). \tag{6.29}$$

The operator $A : Z \to 2^Z, (x, u) \mapsto (\partial f(x) + L^*u, -Lx + (\partial g)^{-1}u)$ is maximally monotone [8, Proposition 26.32 (iii)] and the operator $B : Z \to Z, (x, u) \mapsto (\nabla h(x), 0)$ is θ-cocoercive, with $\theta = 1/\beta$. Thus, it is again natural to think of the forward–backward iteration, with preconditioning. The difference with the construction in Section 6.2 is the presence of the nonlinear operator ∂f, which prevents us to express $x_{n+\frac{1}{2}}$ in terms of x_n and $u_{n+\frac{1}{2}}$. Instead, the iteration is made explicit by canceling the dependence of $x_{n+\frac{1}{2}}$ from $u_{n+\frac{1}{2}}$ in P. That is, the iteration, written in implicit form, is:

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \underbrace{\begin{pmatrix} \partial f(x_{n+\frac{1}{2}}) + L^*u_{n+\frac{1}{2}} \\ -Lx_{n+\frac{1}{2}} + (\partial g)^{-1}u_{n+\frac{1}{2}} \end{pmatrix}}_{Az_{n+\frac{1}{2}}} + \underbrace{\begin{pmatrix} \nabla h(x_n) \\ 0 \end{pmatrix}}_{Bz_n} + \underbrace{\begin{pmatrix} \frac{1}{\tau}\mathrm{Id} & -L^* \\ -L & \frac{1}{\sigma}\mathrm{Id} \end{pmatrix}}_{P} \underbrace{\begin{pmatrix} x_{n+\frac{1}{2}} - x_n \\ u_{n+\frac{1}{2}} - u_n \end{pmatrix}}_{z_{n+\frac{1}{2}} - z_n}, \tag{6.30}$$

where $\tau > 0$ and $\sigma > 0$ are two real parameters, $z_n = (x_n, u_n)$ and $z_{n+\frac{1}{2}} = (x_{n+\frac{1}{2}}, u_{n+\frac{1}{2}})$. Thus, let $\tau > 0$ and $\sigma > 0$, let $x_0 \in H$ and $u_0 \in U$, and let $\{\rho_n\}_{n\in\mathbb{N}}$ be a sequence of relaxation parameters. The primal–dual forward–backward iteration, which we call the Condat–Vũ iteration is:

Condat–Vũ iteration form I for 6.25 and 6.29: for $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} x_{n+\frac{1}{2}} = \operatorname{prox}_{\tau f}\big(x_n - \tau\nabla h(x_n) - \tau L^* u_n\big) \\ u_{n+\frac{1}{2}} = \operatorname{prox}_{\sigma g^*}\big(u_n + \sigma L(2x_{n+\frac{1}{2}} - x_n)\big) \\ x_{n+1} = x_n + \rho_n(x_{n+\frac{1}{2}} - x_n) \\ u_{n+1} = u_n + \rho_n(u_{n+\frac{1}{2}} - u_n) \end{array} \right. . \tag{6.31}$$

This algorithm was proposed independently by the first author [27] and by B. C. Vũ [69]. An alternative is to update u before x, instead of the converse. This yields a different algorithm, characterized by the primal–dual inclusion

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \underbrace{\begin{pmatrix} \partial f(x_{n+\frac{1}{2}}) + L^* u_{n+\frac{1}{2}} \\ -Lx_{n+\frac{1}{2}} + (\partial g)^{-1} u_{n+\frac{1}{2}} \end{pmatrix}}_{Az_{n+\frac{1}{2}}} + \underbrace{\begin{pmatrix} \nabla h(x_n) \\ 0 \end{pmatrix}}_{Bz_n} + \underbrace{\begin{pmatrix} \frac{1}{\tau}\mathrm{Id} & L^* \\ L & \frac{1}{\sigma}\mathrm{Id} \end{pmatrix}}_{P} \underbrace{\begin{pmatrix} x_{n+\frac{1}{2}} - x_n \\ u_{n+\frac{1}{2}} - u_n \end{pmatrix}}_{z_{n+\frac{1}{2}} - z_n}. \tag{6.32}$$

The corresponding primal–dual forward–backward iteration in this case is given by:

Condat–Vũ iteration form II for (6.25) **and** (6.29): for $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} u_{n+\frac{1}{2}} = \operatorname{prox}_{\sigma g^*}\big(u_n + \sigma L x_n\big) \\ x_{n+\frac{1}{2}} = \operatorname{prox}_{\tau f}\big(x_n - \tau\nabla h(x_n) - \tau L^*(2u_{n+\frac{1}{2}} - u_n)\big) \\ u_{n+1} = u_n + \rho_n(u_{n+\frac{1}{2}} - u_n) \\ x_{n+1} = x_n + \rho_n(x_{n+\frac{1}{2}} - x_n) \end{array} \right. . \tag{6.33}$$

As an application of Theorem 6.3, we obtain the following convergence result [27, Theorem 3.1]:

Theorem 6.14 ***(Condat–Vũ algorithm*** (6.31) ***and*** (6.33)***)*** *Let $x_0 \in H$ and $u_0 \in U$. Suppose that $\tau > 0$ and $\sigma > 0$ satisfy $\tau\big(\sigma\|L\|^2 + \beta/2\big) < 1$ and consider $\delta = 2 - (\beta/2)\left(\frac{1}{\tau} - \sigma\|L\|^2\right)^{-1} > 1$. Suppose that $\{\rho_n\}_{n\in\mathbb{N}}$ is a sequence in $[0, \delta]$ such that $\sum_{n\in\mathbb{N}} \rho_n(\delta - \rho_n) = +\infty$. Then the sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{u_n\}_{n\in\mathbb{N}}$ defined by either the iteration* (6.31) *or the iteration* (6.33) *converge weakly to a solution of* (6.25) *and to a solution of* (6.29)*, respectively.*

Proof. In view of (6.30) and (6.32), this is Theorem 6.3 applied to the problem (6.28). The condition on τ and σ implies that $\sigma\tau\|L\|^2 < 1$, so that P is strongly positive, For this, P must be strongly positive, which is the case if and only if $\sigma\tau\|L\|^2 < 1$, by virtue of the properties of the Schur complement. Let us establish the cocoercivity of $P^{-1}B$ in Z_P. Set $\chi = \dfrac{1}{\beta}\left(\dfrac{1}{\tau} - \sigma\|L\|^2\right)$. In both cases (6.30) and (6.32), we have, for every $z = (x, u)$ and

$z' = (x', u')$ in Z,

$$
\begin{aligned}
\|P^{-1}Bz - P^{-1}Bz'\|_P^2 &= \langle P^{-1}Bz - P^{-1}Bz', Bz - Bz' \rangle \\
&= \left\langle \tfrac{1}{\sigma} \left(\tfrac{1}{\sigma\tau}\mathrm{Id} - L^*L \right)^{-1} \big(\nabla h(x) - \nabla h(x')\big), \nabla h(x) - \nabla h(x') \right\rangle \\
&\leq \left(\tfrac{1}{\tau} - \sigma\|L\|^2 \right)^{-1} \big\langle \nabla h(x) - \nabla h(x'), \nabla h(x) - \nabla h(x') \big\rangle \\
&\leq \beta \left(\tfrac{1}{\tau} - \sigma\|L\|^2 \right)^{-1} \big\langle x - x', \nabla h(x) - \nabla h(x') \big\rangle \\
&= \beta \left(\tfrac{1}{\tau} - \sigma\|L\|^2 \right)^{-1} \big\langle z - z', Bz - Bz' \big\rangle \\
&= \beta \left(\tfrac{1}{\tau} - \sigma\|L\|^2 \right)^{-1} \big\langle z - z', P^{-1}Bz - P^{-1}Bz' \big\rangle_P.
\end{aligned}
$$

So, $P^{-1}B$ is χ-cocoercive in Z_P. Moreover, $\chi > 1/2$ if and only if $\tau(\sigma\|L\|^2 + \beta/2) < 1$. Finally, $\delta = 2 - 1/(2\chi)$. □ □

We can observe that if $h = 0$, the Condat–Vũ iteration reverts to the Chambolle–Pock iteration. So, the former can be viewed as a generalization of the latter. Accordingly, if we set $\beta = 0$ in Theorem 6.14, we recover Theorem 6.13. If $f = 0$, the Condat–Vũ iteration and the Loris–Verhoeven iteration are different. One can expect the larger range of parameters of the latter to be beneficial to the convergence speed in practice.

For the Condat–Vũ algorithm, let us focus on the case where h is a quadratic function; that is, $h : x \mapsto \frac{1}{2}\langle x, Qx \rangle + \langle x, c \rangle$, for some bounded, self-adjoint, positive, nonzero, linear operator Q on H and some element $c \in H$. We have $\beta = \|Q\|$. We can rewrite the primal–dual inclusion (6.30), which characterizes the Condat–Vũ iteration (6.31), as

$$
\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \underbrace{\begin{pmatrix} (\partial f + \nabla h)(x_{n+\frac{1}{2}}) + L^* u_{n+\frac{1}{2}} \\ -Lx_{n+\frac{1}{2}} + (\partial g)^{-1} u_{n+\frac{1}{2}} \end{pmatrix}}_{Az_{n+\frac{1}{2}}} + \underbrace{\begin{pmatrix} \frac{1}{\tau}\mathrm{I} - Q & -L^* \\ -L & \frac{1}{\sigma}\mathrm{I} \end{pmatrix}}_{P} \underbrace{\begin{pmatrix} x_{n+\frac{1}{2}} - x_n \\ u_{n+\frac{1}{2}} - u_n \end{pmatrix}}_{z_{n+\frac{1}{2}} - z_n}. \tag{6.34}
$$

Similarly, we can rewrite the primal–dual inclusion (6.32), which characterizes the second form of the Condat–Vũ iteration (6.33), as (6.34), with L replaced by $-L$. In both cases, using the properties of the Schur complement, P is strongly positive if and only if

$$
\tau\|Q + \sigma L^* L\| < 1 \tag{6.35}
$$

(which implies that $\tau < 1/\beta$). Thus, a sufficient condition for this inequality holds is $\tau(\sigma\|L\|^2 + \beta) < 1$. However, in some applications, $\|Q + \sigma L^* L\|$ may be smaller than $\sigma\|L\|^2 + \beta$, so that larger stepsizes τ and σ may be used when h is quadratic, for the benefit of convergence speed. Thus, when h is quadratic, the Condat–Vũ iteration can be viewed as a preconditioned Chambolle–Pock iteration. Accordingly, as an application of Theorem 6.6, we have:

Theorem 6.15 ***(Condat–Vũ algorithm*** (6.31) ***and*** (6.33)***, quadratic case)*** *Let $x_0 \in H$ and $u_0 \in U$. Suppose that $\tau\|Q + \sigma L^* L\| < 1$. Suppose that $\{\rho_n\}_{n\in\mathbb{N}}$ is a sequence in $[0, 2]$ such that $\sum_{n\in\mathbb{N}} \rho_n(2-\rho_n) = +\infty$. Then the sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{u_n\}_{n\in\mathbb{N}}$ defined by either the iteration* (6.31) *or the iteration* (6.33) *converge weakly to a solution of* (6.25) *and to a solution of* (6.29)*, respectively.*

Remark 6.16 The problem (6.25) was first considered in [21] using a forward–backward–forward scheme. This method is not as popular as the Condat–Vũ algorithm, since it involves more auxiliary variables and is generally slower.

6.4 Numerical experiment

In this section, we implement the instance of Loris–Verhoeven algorithm and Condat–Vũ algorithm, described in (6.22) and (6.31) respectively in image recovering setting and make a comparisson of the obtained results. For the numerical experiments and following a similar scheme as [28], we solve the following optimization problem:

$$\text{Find } \widehat{x} \in \arg\min_{x\in\Omega} \frac{1}{2}\|Ax-y\|^2 + \lambda \cdot \mathrm{TV}(x), \tag{6.36}$$

where:

- $H = \mathbb{R}^{N_h} \times \mathbb{R}^{N_v}$ of a gray scale images of size N_h columns times N_h rows, endowed with the usual Euclidean inner product.
- The data y, which lives in a real Hilbert space U, represents the available image.
- $A : X \to U$ is a linear and bounded operator modeling the acquisition process. As an image corruption A, we used a Gaussian blur of 25×25 and standard desviation 4 (applied by Mathlab function *fspecial*), followed by additive zero mean white Gaussian noise with standard desviation 5.
- Ω is a closed and convex subset of H.
- $\lambda > 0$ is a tradeoff parameter to tune, depending of the properties od A and the noisy level.

The discrete total variation, which is denoted by TV in (6.36), is defined as follow (see also [28] and the references therein): We define the discrete gradient operator $D : H \to H^2$, which maps an image x to a pair of images (u_h, u_v) with, for every $k_h = 1, \cdots, N_h$, $k_v = 1, \cdots, N_v$,

$$\begin{aligned} u_h[k_h,k_v] &= \{x[k_h,k_v] - x[k_h-1,k_v],\ \text{if } k_h \geq 2,\ \ 0\ \text{ else}\} \\ u_v[k_h,k_v] &= \{x[k_h,k_v] - x[k_h,k_v-1],\ \text{if } k_v \geq 2,\ \ 0\ \text{ else}\}\,. \end{aligned}$$

Thus, we have $\mathrm{TV}(x) = \|Dx\|_{1,2}$, with $\|(u_h,u_v)\|_{1,2} = \displaystyle\sum_{k_h=1}^{N_h}\sum_{k_v=1}^{N_v}\sqrt{u_h[k_h,k_v]^2 + u_v[k_h,k_v]^2}$.
According with [16], $\|D^*D\| \leq 8$. Hence, the problem (6.36) becomes to the problem (6.16) with:

- $h(x) = \frac{1}{2}\|Ax-y\|^2$.
- $g(x) = \iota_\Omega(x)$, where ι_Ω in the indicator function on Ω.
- $U = H \times H$, $\lambda \cdot \mathrm{TV} = g \circ L$, with $h = \lambda\|\cdot\|_{1,2}$ and $L = D$.

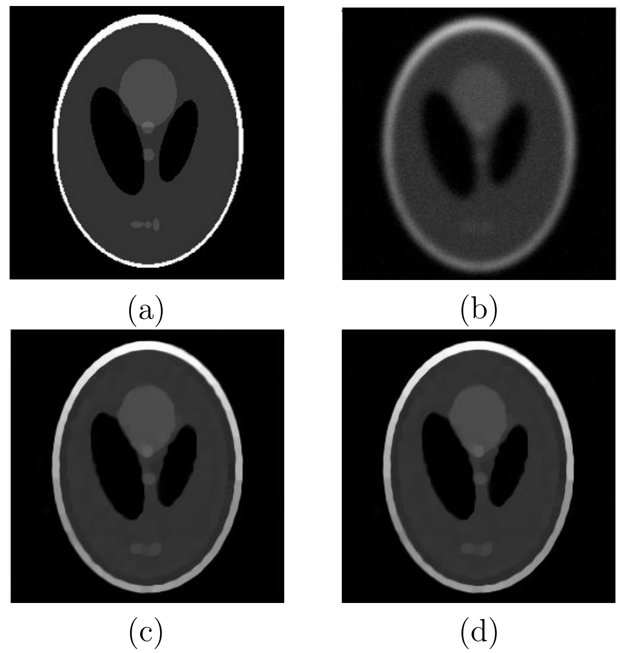

Figure 6.1: (a) correspond to the original image, (b) correspond to the corrupted image, (c) is the recovered image by Condat–Vũ (6.31) with $\sigma = 0.04$, $\tau = \frac{0.99}{0.5+8\sigma}$, $\rho = 1$, $\|L\|^2 = 8$, $\lambda = 0.5$ and (d) is the image recovered by Loris–Verhoeven (6.22) with $\sigma = 0.125$, $\tau = 0.99$, $\rho = 1.5$, $\|L\|^2 = 8$ and $\lambda = 0.5$, after 3000 iterations.

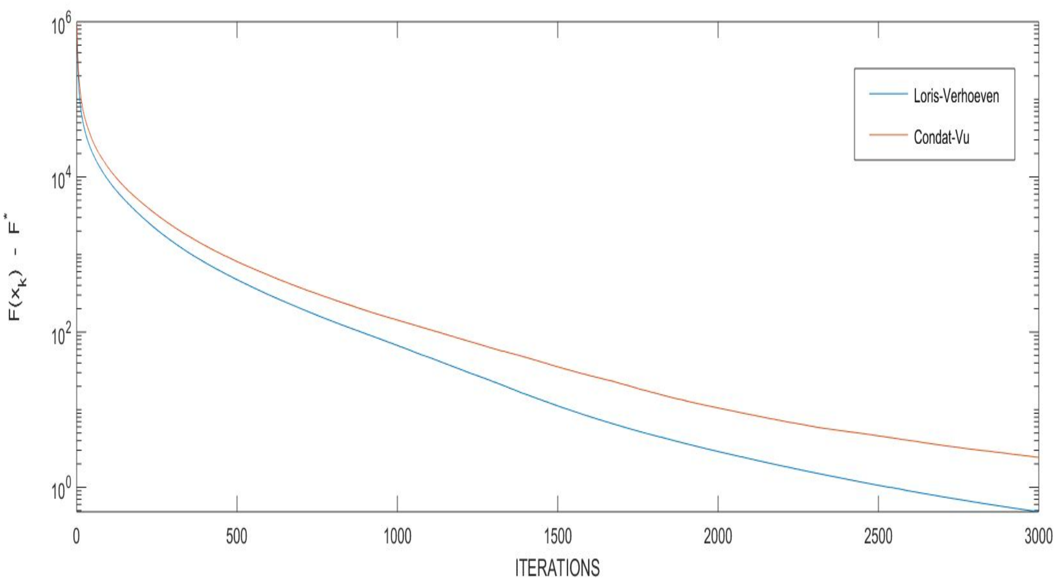

Figure 6.2: Comparisson of the convergence rate between Loris–Verhoeven algorithm (blue) and Condat–Vũ algorithm (orange), after 3000 iterations.

Chapter 7

Work in progress 2: Study of the convergence of the gradient algorithm in absence Lipschitz-continuity condition.

The main goal of this work is to investigate whether it is possible to transfer the convergence properties of the Frank–Wolf algorithm obtained in [72] to the gradient algorithm by cosidering the assumptions in [72] instead of the usual Lipschitz condition. This problem was proposed by professor Hong-Kun Xu during my visit research to the Hahgzhou-Dianzi University in Hangzhou, China.

Convergence of the Frank–Wolf algorithm

The Frank–Wolf Algorithm (FWA) [36] is a first order optimization method also known as conditional gradient method [66], which allows to aproximate the solutions of the constrained optimization problem:

$$\underset{x\in C}{\text{minimize}}\ f(x), \tag{7.1}$$

whose solution set is supposed nonempty, $C \subset \mathbb{R}^n$ nonempty, compact and convex set, $f \in \Gamma_0(\mathbb{R}^n)$ differentiable, with ∇f Lipschitz continuous on C with constant $L > 0$. Starting with an intial $x_0 \in C$, the (FWA) is described as follow:

$$\left\{\begin{array}{rcl} \overline{x}_n & = & \arg\min\limits_{x\in C}\langle \nabla f(x_n), x\rangle, , \\ x_{n+1} & = & x_n + \dfrac{2}{n+2}\left(\overline{x}_n - x_n\right). \end{array}\right. \tag{7.2}$$

Moreover, it is proved that $f(x_n) - f^* \leq O\left(\dfrac{1}{n}\right)$, where $f^* = \min\{f(x) : x \in C\}$ (see [41, 42, Theorem 1]). On the other hand, Xu [72] considere the problem (7.1) in general Banach space setting, with $C \subset X$ nonempty, convex and weakly compact in X. Thus, a more general version of Frank–Wolf algorithm (7.2) is considered in order to approach the solution of (7.1), namely:

$$\begin{cases} \overline{x}_n &= \arg\min_{x\in C}\langle f'(x_n), x\rangle, , \\ x_{n+1} &= x_n + \gamma_n\left(\overline{x}_n - x_n\right), \end{cases} \tag{7.3}$$

where f is a continuously Fréchet differentiable, where f' is supposed uniformly continuous on C and the step size γ_n is chosen by:

- *Linear Search:*
$$\gamma_n = \arg\min_{\gamma\in[0,1]} f\left(x_n - \gamma(\overline{x}_n - x_n)\right), \tag{7.4}$$
- *Open Loop Rule:* The sequence $\{\gamma_n\}_{n\in\mathbb{N}}$ satisfies one of the follwoing two conditions
  (a) $\lim\limits_{n\to\infty}\gamma_n = 0$.
  (b) $\sum\limits_{n=1}^{\infty}\gamma_n = \infty$.

In order to correct the abscense of Lipschitz condition on f', [72] considered the notion of *curvature constant* of f of order $\sigma\in(1,2]$ on C, which is given by:

$$C_f^{(\sigma)} = \sup\left\{\frac{\sigma}{\gamma^\sigma}\left(f(y) - f(x) - \langle y - x, \nabla f(x)\rangle\right) : x, z\in C, \gamma\in(0,1), y = x + \gamma(z - x)\right\}.$$

Thus, $C_f^{(\sigma)}\geq 0$ is the least nonnegtive number such that

$$f(y)\leq f(x) + \langle f'(x), y - x\rangle + \frac{\gamma^\sigma}{\sigma}C_f^{(\sigma)}, \tag{7.5}$$

for all $x, y\in C$ such that $y = x + \gamma(z - x)$ for all $\gamma\in[0,1]$ and $z\in C$. The following result comes from [72, Theorem 4.9 and Theorem 4.10]:

Theorem 7.1 *Let $\{x_n\}_{n\in\mathbb{N}}$ be generated by (7.3). Suppose that there exist $\sigma > 1$ such that $C_f^{(\sigma)}$ is finite. If the step size $\{\gamma_n\}_{n\in\mathbb{N}}$ is selected by linear search (7.4), then*

$$f(x_n) - f^*\leq \frac{\theta}{\left(1 + \frac{1}{\sigma}\theta^{\frac{1}{\sigma-1}}\left(C_f^{(\sigma)}\right)^{\frac{1}{\sigma-1}}\cdot n\right)^{\sigma-1}} = O\left(\frac{1}{n^{\sigma-1}}\right) \tag{7.6}$$

where $\theta = f(x_0) - f^*$. In particular:

- *If f' is ν-Hölder continuous with constant $L_\nu > 0$ and $\nu\in(0,1]$, then*
$$f(x_n) - f^*\leq \frac{\theta}{\left(1 + \frac{1}{\nu+1}\theta^{\frac{1}{\nu}}(L_\nu\delta^{\nu+1})^{-\frac{1}{\nu}}\cdot n\right)^{\nu}} = O\left(\frac{1}{n^\nu}\right),$$
- *If f' is Lipschitz continuous with constant $L > 0$, then*
$$f(x_n) - f^*\leq \frac{\theta}{1 + \frac{\theta}{2L\delta^2}\cdot n} = O\left(\frac{1}{n}\right), \tag{7.7}$$

with $\delta = diam(C)$.

If the step sizes $\{\gamma_n\}_{n\in\mathbb{N}} \subset (0,1]$ satisfies the open loop rule, then

$$f(x_n) - f^* \leq \frac{\sigma^\sigma \Delta}{n^{\sigma-1}} \quad \textit{for all} \quad n \geq 1, \tag{7.8}$$

where $\Delta = \max\left\{f(x_0) - f^*, \frac{1}{\sigma}C_f^{(\sigma)}\right\}$.

- *If f' is ν-Hölder continuous and $\nu \in (0,1]$, then*

$$f(x_n) - f^* \leq O\left(\frac{1}{n^\nu}\right).$$

- *If f' is Lipschitz continuous with constant C, then*

$$f(x_n) - f^* \leq O\left(\frac{1}{n}\right).$$

Advanced work on the gradient method

Consider X a real Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$, $f \in \Gamma_0(X)$ differentiable, such that $\inf(f) > -\infty$ and ∇f is uniformly continuous on X. Our main goal is to extend the result obtained in [72] for Frank–Wolf algorithm to gradient algorithm:

$$x_n = x_{n-1} - \gamma_n \nabla f(x_{n-1}), \;\; n \in \mathbb{N}, \tag{7.9}$$

with initial point $x_0 \in C$. From now on, let us suppose that $C_f^{(\sigma)}$ is finite for some $\sigma > 1$, and the step size $\gamma_n \in [0,1]$ is chosen by linear search (7.4). Then, for each $n \in \mathbb{N}$ we have

$$f(x_{n+1}) = f(x_n - \gamma_n \nabla f(x_n)) = \min_{\gamma\in[0,1]} f(x_n - \gamma\nabla f(x_n)) \leq f(x_n). \tag{7.10}$$

Therefore $\{f(x_n)\}_{n\in\mathbb{N}}$ is nonincreasing.

On the other hand

$$\begin{aligned} f(x_{n+1}) &= f(x_n - \gamma_n \nabla f(x_n)) \\ &= \min_{\gamma\in[0,1]} f(x_n - \gamma\nabla f(x_n)) \\ &\leq \min_{\gamma\in[0,1]} \left\{f(x_n) - \gamma\|\nabla f(x_n)\|^2 + \frac{\gamma^\sigma}{\sigma}C_f^{(\sigma)}\right\} \end{aligned} \tag{7.11}$$

where the inequality holds by (7.5). Let $\phi(\gamma) = f(x_n) - \gamma\|\nabla f(x_n)\|^2 + \frac{\gamma^\sigma}{\sigma}C_f^{(\sigma)}$ and suppose that $\widehat{\gamma} \in [0,1]$ is the minimizer of ϕ in $[0,1]$ and consider the following cases:

Case 1: If $\widehat{\gamma} = 1$, we obtain

$$C_f^{(\sigma)} \leq \frac{(1-\gamma)\sigma}{1-\gamma^\sigma}\|\nabla f(x_n)\|^2, \quad \text{for all} \quad \gamma \in [0,1). \tag{7.12}$$

Combining (7.11) and (7.12) we have:

$$f(x_{n+1}) \leq f(x_n) - \frac{C_f^{(\sigma)}}{\sigma}\left[\frac{1-2\gamma^\sigma+\gamma^{\sigma+1}}{(1-\gamma)\sigma}\right], \quad \text{for all} \quad \gamma \in [0,1). \tag{7.13}$$

Passing to the limit $\gamma \to 0$ in (7.13) we obtain

$$f(x_{n+1}) \leq f(x_n) - \frac{C_f^{(\sigma)}}{\sigma}. \tag{7.14}$$

If (7.14) holds for an infinity number of iterations and since $\{f(x_n)\}_{n\in\mathbb{N}}$ is decreasing, then $\inf(f) = -\infty$, which is a contradiction because f is bounded from below. Therefore $\widehat{\gamma} = 1$ in a finite number of iterations.

Case 2: Now, let us suppose that $0 < \widehat{\gamma} < 1$. By first order optimality condition, we have

$$0 = \phi'(\widehat{\gamma}) = -\|\nabla f(x_n)\|^2 + \widehat{\gamma}^{\sigma-1}C_f^{(\sigma)}.$$

Thus $\widehat{\gamma} = \dfrac{\|\nabla f(x_n)\|^{\frac{2}{\sigma-1}}}{C_f^{(\sigma)\frac{1}{\sigma-1}}}$. From (7.11) we obtain

$$f(x_{n+1}) \leq f(x_n) - \frac{\|\nabla f(x_n)\|^{\frac{2\sigma}{\sigma-1}}(\sigma-1)}{\sigma C_f^{(\sigma)\frac{1}{\sigma-1}}} \tag{7.15}$$

By subgradient inequality, we have

$$-\|\nabla f(x_n)\|^{\frac{2\sigma}{\sigma-1}} \leq -\frac{(f(x_n)-f(x_{n+1}))^{\frac{\sigma}{\sigma-1}}}{\gamma_n^{\frac{\sigma}{\sigma-1}}}. \tag{7.16}$$

Replacing (7.16) in (7.15), we get

$$f(x_{n+1}) \leq f(x_n) - \frac{(\sigma-1)\left(f(x_n)-f(x_{n+1})\right)^{\frac{\sigma}{\sigma-1}}}{\sigma C_f^{(\sigma)\frac{1}{\sigma-1}}\gamma_n^{\frac{\sigma}{\sigma-1}}} \leq f(x_n) - \frac{(\sigma-1)\left(f(x_n)-f(x_{n+1})\right)^{\frac{\sigma}{\sigma-1}}}{\sigma C_f^{(\sigma)\frac{1}{\sigma-1}}\Sigma_n^{\frac{\sigma}{\sigma-1}}}, \tag{7.17}$$

where $\Sigma_n = \sum_{k=1}^{n}\gamma_k$. Thus, for $\theta_n := f(x_n) - f^*$ we have

$$\theta_{n+1} \leq \theta_n - \frac{(\sigma-1)\left(\theta_n-\theta_{n+1}\right)^{\frac{\sigma}{\sigma-1}}}{\sigma C_f^{(\sigma)\frac{1}{\sigma-1}}\Sigma_n^{\frac{\sigma}{\sigma-1}}}. \tag{7.18}$$

It is necessary to obtain a recurrence formula for the sequence θ_n which enable us to obtain information on the convergence rate for the gradient algorithm without Lipschitz condition. On the other hand, there is not information about the convergence of the iterations $\{x_n\}_{n\in\mathbb{N}}$ described by (7.9). Another interesting problem is to study the convergence properties the proximal gradient algorithm under assumption described in [72] for the smooth part, instead of Lipschitz condition. This work represents an interesting challenge, which could be developed in a post-doctoral project.

Conclutions

The conclutions of this thesis are summarize as follow:

1. The continuous trajectories associated to a differential inclusion governed by the sum of a maximally monotone operator with a cocoercive operator and the discrete ones generated by means of a forward backward discretization of the continuous system, can be compared looking at their qualitative differences and similarities in terms of the convergence of these trajectories in finite and infinity horizon. Although the existing theory for resolvent iterations can be applied in this setting, their implementation becomes computationally intractable, since computing the resolvent of a sum of two operator is a difficult task. The new theory presented in this manuscript for forward backward iteration is more adequate because the two operator involved are treated independently.

2. The relationship between continuous and discrete systems described above allows to obtain new asymptotic properties of the forward backward algorithm, such as was showed in Theorem 4.11, where we provided new results on the strong convergence for forward backward algorithm. In the particular case of the explicit iterations governed by an operator deriving from a potential, we provided important results concerning to the strong convergence of gradient algorithm, such as was showed in Theorem 3.18.

3. Relaxed versions of the preconditioned forward backward algorithm can be obtained when the cocoercive operator is affine. These new versions allow to obtain relaxed versions of some outstanding primal dual algorithms, such as Loris-Verhoeven and Condat-Vũ algorithms. The relaxation of these algorithms can be interpreted as accelerated versions of the original ones, which can be appreciated by means of numerical experiments in image recovery.

Bibliography

[1] Álvarez, F.; Peypouquet, J. Asymptotic equivalence and Kobayashi-type estimates for nonautonomous monotone operators in Banach spaces. *Discrete and Continuous Dynamical Systems 25*, 4 (2009), 1109–1128.

[2] Álvarez, F.; Peypouquet, J. Asymptotic almost-equivalence of lipschitz evolution systems in Banach spaces. *Nonlinear Analysis: Theory, Methods & Applications 73*, 9 (2010), 3018–3033.

[3] Álvarez, F.; Peypouquet, J. A unified approach to the asymptotic almost-equivalence of evolution systems without Lipschitz conditions. *Nonlinear Analysis: Theory, Methods & Applications 74*, 11 (2011), 3440–3444.

[4] Attouch, H.; Baillon, J. Weak versus strong convergence of a regularized newton dynamic for maximal monotone operators. *Nonlinear Analisis 46*, 1 (2018), 177–195.

[5] Baillon, JB.; Haddad, G. Quelques propiétés des opéreteurs angles-bornés et n-cyclement monotones. *Isr. J. Math. 26*, 2 (1977), 137–150.

[6] Baillon, J. Un exemple concernant le comportement asymptotique de la solution du problème $\mathrm{d}u/\mathrm{d}t + \partial\varphi(u) \ni 0$. *J. Funct. Anal, 28* (1978), 369–376.

[7] Barbu, V. *Nonlinear semigroups and differential equations in Banach spaces*. Nordhoff, Leyden., 1976.

[8] Bauschke, H.; Combettes, P. *Convex analysis and monotone operator theory in Hilbert spaces, CMS Books in Mathematics*, 2nd ed. Springer, New York, 2017.

[9] Bénilan, P. Équations d'Évolution san un Espace de Banach quelconque et applications. *Thése, Orsay* (1972).

[10] Brézis, H. Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert.

[11] Brézis, H.; Lions, P. Produits infinis de résolvantes. *Israel J. Math. 29* (1978), 329–345.

[12] Bruck, R. Asymptotic convergence of nonlinear contraction semigroup in Hilbert spaces. *J. Funct. Anal. 8* (1975), 15–26.

[13] Bruck, RE.; Reich, S. A general convergence principle in nonlinear functional analysis. *Nonlinear Anal. 4* (1980), 939–950.

[14] Cartan, H. *Cours de calcul différentiel.* Paris, Hermann, 1967.

[15] Cauchy, A. Méthode générale pour la résolution des systemes d'équations simultanées.

[16] Chambolle, A.; Pock, T. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision 40*, 1 (may 2011), 120–145.

[17] Chaux, C.; Combettes, P. L., and Pesquet, J.-C.; Wajs, V. R. A variational formulation for frame-based inverse problems. *Inverse Problems 23*, 4 (2007), 1495–1518.

[18] Chen, H.-G. *Forward–Backward Splitting Techniques: Theory and Applications.* PhD thesis, University of Washington, Seattle, WA, 1994.

[19] Chen, P.; Huang, J., and Zhang, X. A primal–dual fixed point algorithm for convex separable minimization with applications to image restoration. *Inverse Problems 29*, 2 (2013).

[20] Combettes, P.L.; Pesquet, J.-C. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, R. Bauschke, H. H.; Burachik, V. Combettes, P. L.; Elser, and H. Luke, D. R.; Wolkowicz, Eds. Springer-Verlag, New York, 2010.

[21] Combettes, P.L.; Pesquet, J.-C. Primal–dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. *Set-Valued and Variational Analysis 20*, 2 (2012), 307–330.

[22] Combettes, P. L.; Condat, L., and Pesquet, J.-C.; Vũ, B. C. A forward–backward view of some primal–dual optimization methods in image recovery. In *Proc. of IEEE ICIP* (Paris, France, Oct. 2014).

[23] Combettes, P. Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization 53*, 5-6 (2004), 475–504.

[24] Combettes, P.L.; Wajs, V. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul. 4*, 4 (2005), 1168–1200.

[25] Combettes, P. L.; Wajs, V. R. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul. 4* (2005), 1168–1200.

[26] Condat, L.; Contreras, A., and Kitahara, D.; Hirabayashi, A. Proximal Splitting Algorithms: Relax them all! *Submitted.*

[27] Condat, L. A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl. 158*, 2 (2013), 460–479.

[28] Condat, L. A Generic Proximal Algorithm for Convex Optimization – Application to Total Variation Minimization. *IEEE Signal Proc. Letters 21*, 8 (2014), 1054–1057.

[29] Contreras, A.; Peypouquet, J. Forward–Backward Approximation of Evolution Equations in Finite and Infinite Horizon. *Submitted*.

[30] Contreras, A.; Peypouquet, J. Asymptotic equivalence of evolution equation governed by cocoercive operators and their forward discretizations. *Journal of Optimization, Theory and Applications. 182*, 1 (2019), 30–48.

[31] Crandall, MG.; Pazy, A. Semigroups of linear contractions and dissipative sets. *J. Funct. Anal. 3* (1969), 376–418.

[32] Crandall, M.G.; Liggett, T. Generation of semigroups of nonlinear transformations on general banach spaces. *Am. J. Math 93* (1971), 265–298.

[33] Daubechies, I.; Defrise, M., and De Mol, C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math. 57*, 11 (2004), 1413–1457.

[34] Drori, Y.; Sabach, S., and Teboulle, M. A simple algorithm for a class of nonsmooth convex concave saddle-point problems. *Oper. Res. Lett. 43*, 2 (2015), 209–214.

[35] El Gheche, M.; Chierchia, G., and Pesquet, J.-C. Proximity operators of discrete information divergences. *IEEE Trans. Inform. Theory 64*, 2 (2018), 1092–1104.

[36] Frank, M.; Wolf, P. An algorithm for quadratic programming. *Naval Research Logistic Quaterly 3* (1956), 95–110.

[37] Gabay, D. Applications of the method of multipliers to variational inequalities. In *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*, M. Fortin and R. Glowinski, Eds. Amsterdam, Netherlands, 1983.

[38] Goldstein, A. Convex programming in hilbert space. *Bull. Amer. Math. Soc. 70* (1964), 709–710.

[39] Güler, O. On the convergence of the proximal point algorithm for convex optimization. *SIAM J. Control Opt. 29* (1991), 403–419.

[40] Hirsch, F. *Familles resolvantes, generateurs, cogénérateurs, potentiels*. These de Doctorat d'État, Orsay, 1971.

[41] Jaggi, M., H. Z. The recent revival of The Frank-Wolfe algorithm. *Optima 95* (September 2014), 2–8.

[42] Jaggi, M. Revisiting Frank Wolf: Projections-free sparse convex optimization, Proceedings of the 30th International Conference on Machine Learning. Atlanta, Georgia. *JMLR:WCP 28* (2013), 427–435.

[43] Kato, T. Nonlinear semi-groups and evolution equations. *J. Math. Soc. Japan 19* (1967), 508–520.

[44] Kobayashi, Y. Difference aproximation of Cauchy problems for quasi-dissipative operators and generation of nonlinear semigroups. *J. Math Soc. Japan 27* (1975), 640–665.

[45] Kobayashi, K.; Kobayashi, Y., and Oharu, S. Nonlinear evolution operators in Banach spaces. *Osaka J. Math 21* (1984), 281–310.

[46] Kobayasi, K.; Miyadera, I. On the asymptotic behavior of almost-orbits of nonlinear contractions in Banach spaces. *Nonlinear Analisis 6* (1982), 349–365.

[47] Koizumi, M.; Sugimoto, T. On the asymptotic behavior of a nonlinear contraction semigroup and the resolvente iteration. *Proc. Japan Acad. Ser. A. Math. Sci. 59*, 6 (1983), 238–240.

[48] Komura, Y. Nonlinear semi-groups in Hilbert spaces. *Nonlinear Analysis: Theory, Methods & Applications 19* (1967), 493–507.

[49] Levitin, E.S.; Poljak, B. Minimization methods in the presence of constraints. *(Russian) Z. Vycisl. Mat i Mat. Fiz. 6* (1966), 787–823.

[50] Lindenstrauss, J.; Tzafriri, L. Classical banach spaces.ii. function spacesergebnisse der mathematik und ihrer grenzgebiete [results in mathematics and related areas]. *Berlin-New York: Springer-Verlag 97* (1979), pp. x+243.

[51] Lions, P.L.; Mercier, B. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal. 16*, 6 (1979), 964–979.

[52] López G.; Martín Márquez V.; Wang F.; Xu, H. Forward-backward splitting methods for accretive operators in Banach spaces. *Abstr. Appl. Anal. 16* (2012), 25 pp.

[53] Loris, I.; Verhoeven, C. On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty. *Inverse Problems 27*, 12 (2011).

[54] Martinet, B. Regularisation, d'inéquations variationelles par approximations succesives. *Rev.fr.autom.inform.rech.Opér 4* (1970), 154–158.

[55] Mercier, B. Topics in finite element solution of elliptic problems. No. 63 in lectures on mathematics.

[56] Minty, G. Monotone (nonlinear) operators in Hilbert space. *Duke Math. J. 29* (1962), 341–346.

[57] Moreau, J. Proximité et dualité dans un espace Hilbertien. *Bull. Soc. Math., France 93* (1965), 273–299.

[58] Moreau, J. J. Fonctions convexes duales et points proximaux dans un espace Hilbertien. *C. R. Acad. Sci. Paris Sér. A Math. 255* (1962), 2897–2899.

[59] NEVALINA, O.; REICH, S. Strong convergence of contraction semigroups and of iterative methods for accretive operators in Banach spaces. *Israel J. Math. 32* (1979), 44–58.

[60] O'CONNOR, D., AND VANDENBERGHE, L. On the equivalence of the primal-dual hybrid gradient method and Douglas–Rachford splitting. *Mathematical Programming* (2018). To be published.

[61] OPIAL, Z. Convergence of the sequence of succesive approximations for nonexpansive mappings. *Bull. Amer. Math. Soc. 73* (1967), 591–597.

[62] PASSTY, G. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *J. Math. Anal. Appl. 72* (1979), 383–390.

[63] PASSTY, G. Preservation of the asymptotic behavior of a nonlinear contraction semigroup by backward differencing. *Houston J. Math. 7* (1981), 103–110.

[64] PEYPOUQUET, J. *Convex optimization in normed spaces: theory, methods and examples.* Springer, 2015.

[65] PEYPOUQUET, J.; SORIN, S. Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time. *J. Convex Analysis 17* (2010), 1113–1163.

[66] POLYAK, B. *Introduction to Optimization.* New York: Optimization Software, Inc., 1987.

[67] ROCKAFELLAR, R. Monotone operators and the proximal point algorithm,. *SIAM J. Control Optim. 14* (1976), 877–898.

[68] TSENG, P. Applications of splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM J. Control Optim. 29* (1991), 119–138.

[69] VŨ, B. C. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math. 38*, 3 (2013), 667–681.

[70] VIGERAL, G. Evolution equations in discrete and continuous time for nonexpansive operators in Banach spaces. *ESAIM, Control Optim. Calc. Var. 16* (2010), 809–832.

[71] XU, H. Inequalities in Banach spaces with applications. *Nonlinear Analisis 16* (1991), 1127–1138.

[72] XU, H. Convergence Analysis of the Frank-Wolfe Algorithm and Its Generalization in Banach Spaces.

[73] YAN, M. A New Primal–Dual Algorithm for Minimizing the Sum of Three Functions with a Linear Operator. *Journal of Scientific Computing 76*, 3 (sep 2018), 1698–1717.

[74] YOSIDA, K. On the differentiability and the representation of one-parameter semi-group of linear operators. *J. Math. Soc. Japan 1* (1948), 15–21.