Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time

Juan Peypouquet
Departamento de Matemática, Universidad Técnica Federico Santa María
Av. España 1680, Valparaíso, Chile
juan.peypouquet@usm.cl
Sylvain Sorin
Equipe Combinatoire et Optimisation, CNRS FRE 3232, Faculté de Mathématiques,
Université P. et M. Curie - Paris 6, 175 Rue du Chevaleret, 75013 Paris
and Laboratoire d'Econométrie, Ecole Polytechnique, France
sorin@poly.polytechnique.fr

May 28, 2018

Abstract

This survey is devoted to the asymptotic behavior of solutions of evolution equations generated by maximal monotone operators in Hilbert spaces. The emphasis is in the comparison of the continuous time trajectories to sequences generated by implicit or explicit discrete time schemes. The analysis covers weak convergence for the average process, for the process itself and strong convergence and aims at highlighting the main ideas and unifying the proofs. We further make the connection with the analysis in terms of almost orbits that allows for a broader scope.

Contents

Introduction

Discrete and continuous dynamical systems governed by maximal monotone operators have a great number of applications in optimization, equilibrium, fixed-point theory, partial differential equations, among others.

We are specially concerned about the connection between continuous and discrete models. This connection occurs at two levels:

1. On a compact interval, one approximates a continuous-time trajectories by interpolation of some sequences computed via discretization. By considering vanishing step size this construction is used to prove existence results and to approximate the trajectories numerically.

2. Another approximation is in the long term, were we compare asymptotic properties of a continuous trajectory to similar asymptotic properties of a given path defined inductively trough a sequence of values and step sizes.

It is important to mention that some estimations (eg. Kobayashi) can be useful for both purposes.

The literature on this subject is huge but lot of the arguments turn out to be pretty much the same. Therefore, we intend to give a concise yet complete compendium of the results available, with an emphasis on the techniques and the way the enter in the proofs.
Most of the properties will be established in the framework of Hilbert spaces since our aim is to emphasize unity in terms of tools and approach. A lot of results can be extended but in most of the case under specific assumptions. With no aim for completeness, we have included several references to the corresponding results in Banach spaces that we think might be useful.

The paper is organized as follows: In section 1 we recall the basic properties of maximal monotone operators along with some examples. Section 2 deals with the associated dynamic approach. We present the existence results for the differential inclusion $\dot{u} \in -Au$ and global properties of implicit and explicit discretizations. Section 3 establishes the convergence of the value $f(u)$ in the case of an operator of the form $A = \partial f$. In section 4 we describe general results on weak convergence: tools, arguments, characterization of the weak limits. Section 5 is devoted to weak convergence in average and Section 6 is concerned with weak convergence, especially for demipositive operators. In section 7 we present the, mostly geometric, conditions ensuring that the convergence is strong. Section 8 deals with asymptotic equivalence and explains some apparently hidden relationships between certain continuous- and discrete-time dynamical systems. Finally, section 9 contains some concluding remarks.

1 Preliminaries

The purpose of this section is to introduce notations and to recall basic results.

1.1 Maximal monotone operators

Let H be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. An *operator* is a set-valued mapping $A : H \rightrightarrows H$ whose domain

$$D(A) = \{u \in H : Au \neq \emptyset\}$$

is nonempty. For convenience of notation, sometimes we will identify A with its graph by writing $[u, u^*] \in A$ for $u^* \in Au$. The operator A^{-1} is defined by its graph: $[u, u^*] \in A^{-1}$ if, and only if, $[u^*, u] \in A$.
An operator $A : H \rightrightarrows H$ is *monotone* if one has

$$\langle x^* - y^*, x - y \rangle \geq 0 \tag{1}$$

for all $[x, x^*], [y, y^*] \in A$.
A monotone operator is *maximal* if its graph is not properly contained in the graph of any other monotone operator. Observe that if A is monotone (resp. maximal monotone) then so are A^{-1} and λA if $\lambda > 0$.

Lemma 1 *Let A be a maximal monotone operator. A point $[x, x^*] \in H \times H$ belongs to the graph of A if, and only if,*

$$\langle x^* - u^*, x - u \rangle \geq 0 \qquad \text{for all} \quad [u, u^*] \in A.$$

Proof. If $[x, x^*] \in A$ the inequality holds by monotonicity. Conversely, if $[x, x^*] \notin A$, then the set $A \cup \{[x, x^*]\}$ is the graph of a monotone operator that extends A, which contradicts maximality. ■

An operator $A : H \rightrightarrows H$ is *nonexpansive* if one has

$$\|x^* - y^*\| \leq \|x - y\| \tag{2}$$

for all $[x, x^*], [y, y^*] \in A$. Observe that a nonexpansive operator is single-valued on its domain.

Let I be the identity mapping on H. For $\lambda > 0$, the *resolvent* of A is the operator

$$J_\lambda^A = (I + \lambda A)^{-1}.$$

Theorem 2 *Let $A : X \rightrightarrows X$. Then*

i) A is monotone if, and only if, J_λ^A is nonexpansive for each $\lambda > 0$.

ii) A monotone operator A is maximal if, and only if, $I + \lambda A$ is surjective for each $\lambda > 0$.

Proof.
$i)$ Let A be monotone, $[x,x^*],[y,y^*]\in A$ and $\lambda>0$.
Inequality (1) implies
$$\|x-y\|\le\|x-y+\lambda(x^*-y^*)\|,\quad\forall\lambda\ge 0 \tag{3}$$
which is the non expansiveness of J_λ^A.
Conversely, (3) leads to
$$\lambda\langle x^*-y^*,x-y\rangle+\lambda^2\|x^*-y^*\|^2\ge 0$$
hence implies (1) by dividing by λ and letting $\lambda\to 0$.
$ii)$ It is enough to prove the result for $\lambda=1$. Given $z_0\in H$, we will find $x_0\in H$ such that $\langle y-(z_0-x_0),x-x_0\rangle\ge 0$ for all $[x,y]\in A$ so that maximality of A implies $z_0-x_0\in Ax_0$. For $[x,y]\in A$, define the weakly compact set $C_{x,y}$ by
$$C_{x,y}=\{x_0\in H:\langle y+x_0-z_0,x-x_0\rangle\ge 0\}.$$
It suffices to show that the family $\{C_{x,y}\}_{[x,y]\in A}$ has the finite intersection property. To this end take $[x_i,y_i]\in A$ for $i=1,\dots,n$. Let $\Delta=\{(\lambda_1,\dots,\lambda_n):\lambda_i\ge 0;\sum_{i=1}^n\lambda_i=1\}$ denote the n-dimensional simplex and consider the function $f:\Delta\times\Delta\to\mathbf{R}$ given by
$$f(\lambda,\mu)=\textstyle\sum_{i=1}^n\mu_i\langle y_i+x(\lambda)-z_0,x(\lambda)-x_i\rangle$$
with $x(\lambda)=\sum_{i=1}^n\lambda_i x_i$. Clearly $f(\cdot,\mu)$ is convex and continuous while $f(\lambda,\cdot)$ is linear. The Min-max Theorem (see, for instance, Theorem 1.1 in [19, Brézis]) implies the existence of $\lambda_0\in\Delta$ such that
$$\max_{\mu\in\Delta}f(\lambda_0,\mu)=\max_{\mu\in\Delta}\min_{\lambda\in\Delta}f(\lambda,\mu)\le\max_{\mu\in\Delta}f(\mu,\mu).$$
Now monotonicity of A implies
$$\begin{aligned}f(\mu,\mu)&=\textstyle\sum_{i=1}^n\mu_i\langle y_i,x(\mu)-x_i\rangle+\langle x(\mu)-z_0,x(\mu)-x(\mu)\rangle\\&=\textstyle\sum_{i,j=1}^n\mu_i\mu_j\langle y_i,x_j-x_i\rangle\\&=\tfrac12\textstyle\sum_{i,j=1}^n\mu_i\mu_j\langle y_i-y_j,x_j-x_i\rangle\le 0\end{aligned}$$
so that $f(\lambda_0,\mu)\le 0$ for all $\mu\in\Delta$, and taking for μ the extreme points we get $\langle y_i+x(\lambda_0)-z_0,x(\lambda_0)-x_i\rangle\le 0$ for all i, which is $x(\lambda_0)\in\bigcap_{i=1}^n C_{x_i,y_i}$.
Conversely, take $[u,u^*]\in H\times H$ such that $\langle u^*-v^*,u-v\rangle\ge 0$ for all $[v,v^*]\in A$. We shall prove that $[u,u^*]\in A$. Since $I+A$ is surjective, there is $[\overline{v},\overline{v}^*]\in A$ such that $\overline{v}+\overline{v}^*=u+u^*$. Then $\langle u^*-\overline{v}^*,u-\overline{v}\rangle=-\|u-\overline{v}\|^2\ge 0$ which implies $u=\overline{v}$, $u^*=\overline{v}^*$ and $[u,u^*]\in A$. ■

Comments
The study of monotone operators started in [43, Minty]. See also [35, Kato] for part $i)$ in Banach spaces. The *if* part in $ii)$ holds in Banach spaces, but not the *only if* part (see [34, Hirsh]). □

1.2 Examples and properties

Example 1 Let $C\subset H$ and let $T:C\to H$ be nonexpansive. The operator $A=I-T$ is monotone because
$$\begin{aligned}\langle Ax-Ay,x-y\rangle&=\|x-y\|^2-\langle Tx-Ty,x-y\rangle\\&\ge\|x-y\|\Big[\|x-y\|-\|Tx-Ty\|\Big]\\&\ge 0.\end{aligned}$$
Maximality depends on whether T can be extended to a nonexpansive function on a set that contains C properly (for example if C is closed and convex). □

Example 2 Let $\Gamma_0(H)$ denote the set of all proper, lower-semicontinuous convex functions $f : H \to \mathbf{R} \cup \{+\infty\}$. For $f \in \Gamma_0(H)$, the *subdifferential of* f is the operator $\partial f : H \rightrightarrows H$ defined by

$$\partial f(x) = \{x^* \in H : f(z) \geq f(x) + \langle x^*, z - x\rangle \text{ for all } z \in H\}.$$

To see that it is monotone, take $x^* \in \partial f(x)$ and $y^* \in \partial f(y)$. Thus

$$\begin{aligned} f(y) &\geq f(x) + \langle x^*, y - x\rangle \\ f(x) &\geq f(y) + \langle y^*, x - y\rangle. \end{aligned}$$

and adding these two inequalities we obtain $\langle x^* - y^*, x - y\rangle \geq 0$. For maximality, according to Theorem 2 it suffices to prove that for each $y \in H$ and each $\lambda > 0$ there is $x_\lambda \in D(\partial f)$ such that $y \in x_\lambda + \lambda \partial f(x_\lambda)$. Indeed, consider the *Moreau-Yosida approximation* of f at y, which is the function f_λ defined by

$$f_\lambda(x) = f(x) + \frac{1}{2\lambda}\|x - y\|^2. \tag{4}$$

It is proper, lower-semicontinuous, strongly convex and coercive (due to the quadratic term and the fact that f has a affine minorant). Its unique minimizer x_λ satisfies

$$0 \in \partial f_\lambda(x_\lambda) = \partial f(x_\lambda) + \frac{1}{\lambda}(x_\lambda - y).$$

That is, $y \in x_\lambda + \lambda \partial f(x_\lambda)$. □

The *solution set* of A is $\mathcal{S} = A^{-1}0 = \{x \in H; 0 \in Ax\}$. This set is relevant in optimization and fixed-point theory:

- If $A = I - T$, where T is a nonexpansive mapping, then $\mathcal{S}$ is the set of fixed points of T.
- If $A = \partial f$, where f is a proper lower-semicontinuous convex function then $\mathcal{S}$ is the set of minimizers of f.

Let us describe some topological consequences of maximal monotonicity.

Proposition 3 *Let A be a maximal monotone operator. Then A is sequentially weak-strong and strong-weak closed.*

Proof. Take sequences $\{x_n\}$ and $\{x_n^*\}$ in H such that $[x_n, x_n^*] \in A$ for each $n \in \mathbf{N}$ and suppose that $x_n \to x$ and $x_n^* \rightharpoonup x^*$, as $n \to \infty$ (consider A^{-1} for the other case). To prove that $[x, x^*] \in A$, recall that by monotonicity, for all $[u, u^*] \in A$ and all $n \in \mathbf{N}$, $\langle x_n^* - u^*, x_n - u\rangle \geq 0$. Letting $n \to \infty$ the convergence assumptions imply that $\langle x^* - u^*, x - u\rangle \geq 0$ for all $[u, u^*] \in A$. Hence $[x, x^*] \in A$ by Lemma 1. ■

Corollary 4 *Let A be maximal monotone. For each $x \in D(A)$ the set Ax is closed and convex. In particular, $\mathcal{S}$ is closed and convex.*

Proof. Proposition 3 implies Ax is closed for each $x \in D(A)$. To see that Ax is convex, take $x^*, y^* \in Ax$, $[u, u^*] \in A$ and $\lambda \in (0, 1)$. Then $\langle \lambda x^* + (1-\lambda)y^* - u^*, x - u\rangle = \lambda\langle x^* - u^*, x - u\rangle + (1-\lambda)\langle y^* - u^*, x - u\rangle \geq 0$. As before, we conclude that $\lambda x^* + (1 - \lambda)y^* \in Ax$ by Lemma 1. Finally, since A^{-1} is maximal monotone and $\mathcal{S} = A^{-1}0$, the set $\mathcal{S}$ is closed and convex. ■

2 Dynamic approach

The following sections address, among others, the issue of finding zeroes of a maximal monotone operator A. The strategy is the following: we shall consider some continuous and discrete dynamical systems whose trajectories may converge, in some sense and under some conditions, to points in $\mathcal{S} = A^{-1}0$. In this section we present these systems along with some relevant properties.

From now on we assume that A is a maximal monotone operator.

2.1 Differential inclusion

In this section we consider the following differential inclusion:

$$\left\{ \begin{array}{rcl} \dot{u}(t) & \in & -Au(t) \quad \text{a.e. on } (0,\infty) \\ u(0) = x & \in & D(A). \end{array} \right. \tag{5}$$

A *solution* of (5) is an absolutely continuous function u from $\mathbf{R}^+$ to H satisfying these two conditions. Monotonicity implies the following dissipative property:

Lemma 5 *Let u_1 and u_2 be absolutely continuous functions satisfying $\dot{u}_i(t) \in -Au_i(t)$ almost everywhere on $(0,T)$. Then the function $t \mapsto \|u_1(t) - u_2(t)\|$ is decreasing on $(0,T)$.*

Proof. For $t \in (0,T)$ define $\theta(t) = \frac{1}{2}\|u_1(t) - u_2(t)\|^2$. The hypotheses give $\dot{\theta}(t) = \langle \dot{u}_1(t) - \dot{u}_2(t), u_1(t) - u_2(t)\rangle \leq 0$ for almost every t. ■

Immediate consequences are the following:

Corollary 6 *Let $y \in \mathcal{S}$ and u be a solution of* (5)*. Then $\lim\limits_{t\to\infty} \|u(t) - y\|$ exists.*

Corollary 7 *There is at most one solution of* (5)*.*

Another aspect of dissipativity is the following:

Proposition 8 *$\|\dot{u}(t)\|$ is decreasing.*

Proof. Lemma 5 implies that for any $h > 0$ and $s < t$

$$\|u(t+h) - u(t)\| \ \leq \|u(s+h) - u(s)\|$$

hence the result by dividing by h and taking the limit as $h \to 0$. ■

We shall present two approaches for the existence of a solution of (5). The first one uses the *Yosida approximation* and is the best-known in the theory of optimization in Hilbert spaces. The second one uses *proximal sequences* to approximate the function u. It is popular in the field of partial differential equations since it works naturally in arbitrary Banach spaces.

But before doing so, and assuming for a moment that the differential inclusion (5) *does* have a solution, observe that by Lemma 5, for each $t \geq 0$ the mapping $x \mapsto u(t)$ defines a non expansive function from $D(A)$ to itself that can be continuously extended to a map S_t from $\overline{D(A)}$ to itself. The family $\{S_t\}_{t\geq 0}$ is the *semi-group* generated by A and satisfies:

i) $S_0 = I$ and $S_t \circ S_r = S_{t+r}$;

ii) $\|S_t x - S_t y\| \leq \|x - y\|$;

iii) $\lim_{t\to 0} \|x - S_t x\| = 0$.

Reciprocally, given a *continuous semi-group of contractions* i.e. satisfying i), ii) and iii), from a closed convex subset C to itself, there exists a *generator*, namely a maximal operator A with $C = \overline{D(A)}$ such that $S_t x$ coincides with $u(t)$ for $x \in D(A)$, see [19, Brézis].

We will use hereafter both notations $u(t)$ and $S_t x$.

2.2 Approach through the Yosida approximation.

2.2.1 The Yosida Approximation

Recall that the resolvent is J_λ^A. The *Yosida approximation* of A is the single-valued maximal monotone operator A_λ, $\lambda > 0$, defined by

$$A_\lambda = \frac{1}{\lambda}(I - J_\lambda^A).$$

Since J_λ^A is nonexpansive and everywhere defined, A_λ is monotone (see example 1 above) and maximal (using Lemma 1). It is also clear that A_λ is Lipschitz-continuous with constant $2/\lambda$. Observe that $\mathcal{S} = A^{-1}0 = A_\lambda^{-1}0$ for all $\lambda > 0$.

For a closed convex set $C \subset H$ and a point $x \in H$ we denote by $P_C x$ the orthogonal projection of x onto C. The *minimal section* of A is the operator A^0 defined by $A^0 x = P_{Ax}0$, which is clearly monotone but not necessarily maximal.

The following results summarize the main properties of the resolvent and the Yosida approximation. They can be found in [19, Brézis] (see also [13, Barbu] for Banach spaces).

Proposition 9 *With the notation introduced above we have the following:*

1. $A_\lambda x \in AJ_\lambda^A x$
2. $\|A_\lambda x\| \leq \|A^0 x\|$, $\|A_\lambda x\|$ *is nonincreasing in* λ *and* $\lim_{\lambda\to 0} \|A_\lambda x\| \to \|A^0 x\|$.
3. $\lim_{\lambda\to 0} J_\lambda^A x = x$.
4. *If* $x_\lambda \to x$ *and* $A_\lambda x_\lambda$ *remains bounded as* $\lambda \to 0$, *then* $x \in D(A)$. *Moreover, if* y *is a cluster point of* $A_\lambda x_\lambda$ *as* $\lambda \to 0$, *then* $y \in Ax$.
5. A^0 *characterizes* A *in the following sense: If* A *and* B *are maximal monotone with common domain and* $A^0 = B^0$, *then* $A = B$.
6. $\lim_{\lambda\to 0} A_\lambda x = A^0 x$ *and* $\overline{D(A)}$, *the (strong) closure of* $D(A)$, *is convex.*

2.2.2 The existence result

The main result is the following:

Theorem 10 *There exists a unique absolutely continuous function* $u : [0, +\infty) \to H$ *satisfying* (5). *Moreover,*

1. $\dot{u} \in L^\infty(0, \infty; H)$ *with* $\|\dot{u}(t)\| \leq \|A^0 x\|$ *almost everywhere.*
2. $u(t) \in D(A)$ *for all* $t \geq 0$ *and* $\|A^0 u(t)\|$ *decreases.*
3. $A^0 u(t)$ *is continuous from the right and* $u(t)$ *admits a right-hand derivative for all* $t \geq 0$; *namely* $\dot{u}(t^+) = -A^0 u(t)$ *(lazy behavior)*

The problem of finding a trajectory satisfying (5) was first posed and studied in [38, Komura] and [29, Crandall and Pazy]. The classical proof can be found in [19, Brézis]. The idea is to consider the differential inclusion (5) with $A = A_\lambda$, which has a solution u_λ by virtue of the Cauchy-Lipschitz-Picard Theorem. Then one proves first that, as $\lambda \to 0$, u_λ converges uniformly on compact intervals to some u, then that u satisfies (5) for the original A. The following estimation plays a crucial role in the proof and is interesting on its own:

$$\|u_\lambda(t) - u(t)\| \leq 2\|A^0(u_0)\|\sqrt{\lambda t}. \tag{6}$$

Finally u is proved to have the properties enumerated in Theorem 10.

Comments
The same method can be extended to Banach spaces X such that X and X^* are uniformly convex (see [35, Kato]). □

2.3 Approach through proximal sequences.

2.3.1 Proximal sequences

Let $\{\lambda_n\}$ be a sequence of positive numbers or *stepsizes*. $\{x_n\}$ is a *proximal sequence* if it satisfies

$$\begin{cases} \dfrac{x_n - x_{n-1}}{\lambda_n} \in -Ax_n & \text{for all } n \geq 1 \\ \qquad x_0 \in D(A). \end{cases} \tag{7}$$

In other words,

$$x_n = (I + \lambda_n A)^{-1} x_{n-1} = J^A_{\lambda_n} x_{n-1}. \tag{8}$$

The existence of such a sequence follows from Theorem 2. Observe that the first inclusion in (7) can be seen as an implicit discretization of the differential inclusion (5), called also a backward scheme. The *velocity* at stage n is

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

Comments
The notion of proximal sequences and the term *proximal* were introduced in [45, Moreau] for $A = \partial f$. In that case, finding x_n corresponds to minimizing the Moreau-Yosida approximation of f at x_{n-1} (see (4)), namely

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

□

Monotonicity implies the following properties:

Lemma 11 *The sequence $\|y_n\|$ is decreasing.*

Proof. The inequality $\langle y_n - y_{n-1}, x_n - x_{n-1}\rangle \leq 0$ implies $\langle y_n - y_{n-1}, y_n\rangle \leq 0$ and therefore $\|y_n\| \leq \|y_{n-1}\|$. ■

This is the counterpart of $\|\dot{u}(t)\|$ decreasing, Proposition 8.

Lemma 12 *Let $x \in \mathcal{S}$. Then $\|x_n - x\|^2 + \lambda_n^2\|y_n\|^2 \leq \|x_{n-1} - x\|^2$.*

Proof. Simply observe that

$$\begin{aligned}\|x_{n-1}-x\|^2 &= \|x_{n-1}-x_n\|^2+\|x_n-x\|^2+2\langle x_{n-1}-x_n, x_n-x\rangle \\ &\geq \lambda_n^2\|y_n\|^2+\|x_n-x\|^2\end{aligned}$$

since $\langle x_{n-1}-x_n, x_n-x\rangle \geq 0$ by monotonicity when x is in $\mathcal{S}$. ■

An immediate consequence is the following:

Corollary 13 *Let $x \in \mathcal{S}$. The sequence $\|x_n-x\|^2$ is decreasing, thus convergent.*

Notice the similarity with Corollary 6.

2.3.2 Kobayashi inequality

The following inequality, due to Kobayashi [36], provides an estimation for the distance between two proximal sequences $\{x_k\}$ and $\{\widehat{x}_l\}$, with stepsizes $\{\lambda_k\}$ and $\{\widehat{\lambda}_l\}$, respectively.
We use the following notation throughout the paper:

$$\sigma_k=\sum_{i=1}^{k}\lambda_i \qquad \text{and} \qquad \tau_k=\sum_{i=1}^{k}\lambda_i^2$$

(similarily for $\widehat{\sigma}_l$ and $\widehat{\tau}_l$).

Proposition 14 (Kobayashi inequality) *Let $\{x_k\}$ and $\{\widehat{x}_l\}$ be two proximal sequences. If $u \in D(A)$, then*

$$\|x_k-\widehat{x}_l\| \leq \|x_0-u\|+\|\widehat{x}_0-u\|+\|A^0u\|\sqrt{(\sigma_k-\widehat{\sigma}_l)^2+\tau_k+\widehat{\tau}_l}. \tag{9}$$

We first prove the following auxiliary result:

Lemma 15 *Let $[u_1,\ v_1],\ [u_2,\ v_2] \in A$ and $\lambda,\ \mu>0$, then*

$$(\lambda+\mu)\|u_1-u_2\| \leq \lambda\|u_2+\mu v_2-u_1\|+\mu\|u_1+\lambda v_1-u_2\|.$$

Proof. Write $u=u_1-u_2$. Then

$$\begin{aligned}(\lambda+\mu)\|u_1-u_2\|^2 &= \lambda\langle u_2-u_1,-u\rangle+\mu\langle u_1-u_2,u\rangle \\ &= \lambda\langle u_2+\mu v_2-u_1,-u\rangle+\mu\langle u_1+\lambda v_1-u_2,u\rangle \\ &\qquad +\lambda\mu\langle v_2-v_1,u_1-u_2\rangle \\ &\leq [\lambda\|u_2+\mu v_2-u_1\|+\mu\|u_1+\lambda v_1-u_2\|]\,\|u_1-u_2\|\end{aligned}$$

by monotonicity. ■

Proof of Proposition 14: To simplify notation set

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

The proof will use induction on the pair (k,l).
First, let us establish inequality (9) for the pair $(k,0)$ with $k \geq 0$. Monotonicity implies, using (3) that

$$\|x_1-u\| \leq \|x_0-u-\lambda_1 A^0 u\|$$

and

$$\|x_1 - u\| \le \|x_0 - u\| + \lambda_1 \|A^0 u\|.$$

Inductively we obtain

$$\|x_k - u\| \le \|x_0 - u\| + \sigma_k \|A^0 u\|.$$

thus

$$\begin{aligned} \|x_k - \widehat{x}_0\| &\le \|x_k - u\| + \|u - \widehat{x}_0\| \\ &\le \|x_0 - u\| + \sigma_k \|A^0 u\| + \|\widehat{x}_0 - u\| \\ &\le \|x_0 - u\| + \|\widehat{x}_0 - u\| + c_{k,0} \|A^0 u\| \end{aligned}$$

because $\sigma_k \le c_{k,0}$. In a similar fashion we prove the inequality for $(0, l)$ with $l \ge 0$.

Now suppose (9) holds for $(k-1, l)$ and $(k, l-1)$. According to Lemma 15,

$$(\lambda_k + \widehat{\lambda}_l)\|x_k - \widehat{x}_l\| \le \lambda_k \|\widehat{x}_l + \widehat{\lambda}_l \widehat{y}_l - x_k\| + \widehat{\lambda}_l \|x_k + \lambda_k y_k - \widehat{x}_l\|.$$

Setting $\alpha_{k,l} = \dfrac{\widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l}$ and $\beta_{k,l} = 1 - \alpha_{k,l} = \dfrac{\lambda_k}{\lambda_k + \widehat{\lambda}_l}$ we have

$$\begin{aligned} \|x_k - \widehat{x}_l\| &\le \alpha_{k,l}\|x_{k-1} - \widehat{x}_l\| + \beta_{k,l}\|\widehat{x}_{l-1} - x_k\| \\ &\le \alpha_{k,l}\left[\|x_0 - u\| + \|\widehat{x}_0 - u\| + c_{k-1,l}\|A^0 u\|\right] \\ &\quad + \beta_{k,l}\left[\|x_0 - u\| + \|\widehat{x}_0 - u\| + c_{k,l-1}\|A^0 u\|\right] \\ &= \|x_0 - u\| + \|\widehat{x}_0 - u\| + \left[\alpha_{k,l} c_{k-1,l} + \beta_{k,l} c_{k,l-1}\right]\|A^0 u\|. \end{aligned} \tag{10}$$

It only remains to verify that

$$\alpha_{k,l} c_{k-1,l} + \beta_{k,l} c_{k,l-1} \le c_{k,l}. \tag{11}$$

Cauchy-Schwartz Inequality implies

$$\begin{aligned} \alpha_{k,l} c_{k-1,l} + \beta_{k,l} c_{k,l-1} &= \alpha_{k,l}^{1/2}(\alpha_{k,l}^{1/2} c_{k-1,l}) + \beta_{k,l}^{1/2}(\beta_{k,l}^{1/2} c_{k,l-1}) \\ &\le (\alpha_{k,l} + \beta_{k,l})^{1/2}(\alpha_{k,l} c_{k-1,l}^2 + \beta_{k,l} c_{k,l-1}^2)^{1/2} \\ &= (\alpha_{k,l} c_{k-1,l}^2 + \beta_{k,l} c_{k,l-1}^2)^{1/2}. \end{aligned}$$

On the other hand, notice that $c_{k-1,l}^2 = c_{k,l}^2 - 2\lambda_k(\sigma_k - \widehat{\sigma}_l)$, while $c_{k,l-1}^2 = c_{k,l}^2 + 2\widehat{\lambda}_l(\sigma_k - \widehat{\sigma}_l)$. Hence,

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

Inequalities (10) and (11) give (9). ■

Comments
Kobayashi's original inequality also accounts for possible errors in the determination of the proximal sequence, see [36]. Nonautonomous version of the inequality can be found in [**?**, Kobayasi, Kobayashi and Oharu] [2, Alvarez and Peypouquet]. □

2.3.3 The existence result

In general Banach spaces, existence and uniqueness can also be derived by the method in [28, Crandall and Liggett], based on the resolvent, which we now present:

Set $t \in [0,T]$, $m \in \mathbf{N}$ and consider a proximal sequence with constant stepsizes $\lambda_k \equiv t/m$. The m-th iteration defines a function

$$u_m(t) = \left(I + \frac{t}{m}A\right)^{-m} x.$$

Repeat the procedure for each m to obtain a sequence $\{u_m(t)\}$ of functions from $[0,T]$ to H. The following result was proved in [28, Crandall and Liggett]:

Theorem 16 *The sequence $\{u_m(t)\}$ defined above converges to some $u(t)$ uniformly on every compact interval $[0,T]$. Moreover, the function $t \mapsto u(t)$ satisfies* (5).

Proof. Instead of the original proof we present an easier one using Kobayashi's inequality (9)[1]. Fix $N, M \in \mathbf{N}$ and $t, s \in [0,T]$ with $T > 0$. Consider two proximal sequences with $\lambda_k = t/N$ and $\widehat{\lambda}_l = s/M$ for all k, l. Initialize x_k and $\widehat{x}_l$ both at x. Note that $x_N = u_N(t)$ and $\widehat{x}_M = u_M(s)$ hence

$$\|u_N(t) - u_M(s)\| \leq \|A^0x\| \sqrt{(t-s)^2 + \tfrac{T^2}{N} + \tfrac{T^2}{M}}.$$

Thus the sequence $\{u_n\}$ converges uniformly on $[0,T]$ to a function u, which is uniformly Lipschitz-continuous with constant $\|A^0x\|$.
In order to prove that the function u satisfies (5) it suffices to verify that it is an *integral solution* in the sense of Bénilan [17], which means that for all $[x,y] \in A$ and $t > s \geq 0$ we have

$$\frac{1}{2}\left[\|u(t) - x\|^2 - \|u(s) - x\|^2\right] \leq \int_s^t \langle y, x - u(\tau)\rangle \, d\tau. \tag{12}$$

Since u is absolutely continuous, (12) implies $\dot{u}(t) \in -Au(t)$ almost everywhere on $[0,T]$.
Monotonicity of A implies that for any proximal sequence $\{x_k\}$: $\langle x_{k-1} - x_k - \lambda_k y, x_k - x\rangle \geq 0$. But $\|x_k - x\|^2 - \|x_{k-1} - x\|^2 \leq 2\langle x_{k-1} - x_k, x - x_k\rangle$ and so

$$\|x_k - x\|^2 - \|x_{k-1} - x\|^2 \leq 2\lambda_k \langle y, x - x_k\rangle.$$

Summing up for $k = m+1, \ldots n$ we obtain

$$\|x_n - x\|^2 - \|x_0 - x\|^2 \leq 2\sum_{k=1}^{n} \lambda_k \langle y, x - x_k\rangle.$$

Setting $x_0 = u(s)$ and passing to the limit appropriately we finally get (12). Notice that $u(t) \in D(A)$ by maximality. ■

A consequence of Proposition 14 and Theorem 16 is the following

Corollary 17 *The following statements hold:*

i) For each $z \in D(A)$ we have

$$\|x_n - u(t)\| \leq \|x_0 - z\| + \|u(0) - z\| + \|A^0z\| \sqrt{(\sigma_n - t)^2 + \tau_n}.$$

[1] In fact, Kobayashi's proof is based on a simplification of Crandall and Liggett's method.

ii) For trajectories u and v we get

$$\|v(s) - u(t)\| \le \|v(0) - z\| + \|u(0) - z\| + \|A^0 z\| \ |s - t|.$$

iii) The unique function u satisfying (5) is Lipschitz-continuous with

$$\|u(s) - u(t)\| \le \|A^0 x\| \ |s - t|.$$

iv) $\dot{u} \in L^\infty(0, \infty; H)$ with $\|\dot{u}(t)\| \le \|A^0 x\|$ almost everywhere.

Proposition 14 was used to construct a continuous trajectory by considering finer and finer discretizations on a compact interval. By controlling the distance between two discrete schemes it is possible to obtain bounds for the distance between a limit trajectory and a discrete scheme. As a consequence, one can estimate the distance between two trajectories as well.

2.4 Euler sequences

Assume A maps $D(A)$ into itself. (Notice that this is a strong assumption, so the range of applications of this discretization method is limited compared to the proximal sequences). Let $\{\lambda_n\}$ be a sequence of positive numbers or *stepsizes*. Define an *Euler sequence* $\{z_n\}$ recursively by

$$\left\{ \begin{array}{rcll} \dfrac{z_n - z_{n-1}}{\lambda_{n-1}} & \in & -Az_{n-1} & \quad \text{for all } n \ge 1 \\ z_0 & \in & D(A) & \end{array} \right. \tag{13}$$

A remarkable feature of this scheme is that the terms of the sequence can be computed explicitly (forward scheme).

Observe that if $A = I - T$ with T nonexpansive 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. Some of their results will be presented in the forthcoming sections.
Note also that in this framework a Kobayashi-type inequality holds too, namely

$$\|z_k - \widehat{z}_l\| \le \|z_0 - u\| + \|\widehat{z}_0 - u\| + \|u - T(u)\| \sqrt{(\sigma_k - \widehat{\sigma}_l)^2 + \tau_k + \widehat{\tau}_l}, \tag{14}$$

where u is any point in H. This fact was recently pointed out by [60, Vigeral].

Let us define the velocity at stage n as $w_n = \dfrac{z_{n+1} - z_n}{\lambda_n} \in -Az_n$.

Lemma 18 *If $y \in \mathcal{S}$ then $\|z_{n+1} - y\|^2 \le \|z_n - y\|^2 + \lambda_n^2 \|w_n\|^2$.*

Proof. For any $y \in H$ one has

$$\|z_{n+1} - y\|^2 = \|z_n - y\|^2 + 2\lambda_n \langle w_n, z_n - y\rangle + \lambda_n^2 \|w_n\|^2. \tag{15}$$

The desired inequality follows from monotonicity if $0 \in Ay$. ■

Observe the similarity and the difference with (5) and (7). The dissipativity condition in Lemma 18 is much weaker than the corresponding ones in Lemmas 5 and 12.

An immediate consequence is the following:

Corollary 19 *Assume $\sum \|z_{n+1} - z_n\|^2 < \infty$. For each $y \in \mathcal{S}$ the sequence $\|z_n - y\|$ is convergent.*

Proof. It suffices to observe from Lemma 18 that the sequence $\|z_n - y\|^2 + \sum_{m=n}^{+\infty} \|z_{m+1} - z_m\|^2$ is decreasing. ■

Comments
The hypothesis in the previous result holds if $\{\lambda_n\} \in \ell^2$ and $\{w_n\}$ bounded. □

Notice the similarity with Corollaries 6 and 13.

The main drawback of Euler sequences is that they can be quite unstable. Most convergence results need regularity assumptions such as $\{\lambda_n\} \in \ell^2$ and the boundedness of the sequence $\{w_n\}$, or at least that $\sum \|z_{n+1} - z_n\|^2 < \infty$.

An important result involving an operator A of the form $I - T$ is the following, see [19, Brézis]:

Proposition 20 (Chernoff's estimate) *Let T be non-expansive from H to itself and $\lambda > 0$. If v satifies*

$$\dot{v}(t) = -\frac{1}{\lambda}(I - T)v(t)$$

with $v(0) = v_0$ then

$$\|v(t) - T^n v_0\| \leq \|\dot{v}(0)\| \sqrt{\lambda t + (n\lambda - t)^2}.$$

Proof. It is enough to consider the case $\lambda = 1$.
Define $\phi_n(t) = \|v(t) - T^n v_0\|$ and $\gamma_n(t) = \|\dot{v}(0)\| \sqrt{t + [n - t]^2}$. We shall prove inductively that $\phi_n(t) \leq \gamma_n(t)$. For $n = 0$ simply observe that

$$\|v(t) - v_0\| \leq \int_0^t \|\dot{v}(s)\| \, ds \leq \|\dot{v}(0)\| t \leq \gamma_0(t)$$

using point 4 in Theorem 10.
Now let us assume $\phi_{n-1} \leq \gamma_{n-1}$ and prove $\phi_n \leq \gamma_n$. Multiplying $\dot{v}(t) + v(t) = Tv(t)$ by e^t and integrating we obtain $v(t) = v_0 e^{-t} + \int_0^t e^{(s-t)} Tv(s) \, ds$ so that

$$\begin{aligned} \phi_n(t) &= \left\| e^{-t}(v_0 - T^n v_0) + \int_0^t e^{(s-t)} [Tv(s) - T^n v_0] \, ds \right\| \\ &\leq e^{-t} \|v_0 - T^n v_0\| + \int_0^t e^{(s-t)} \phi_{n-1}(s) \, ds. \end{aligned}$$

Noting that $\|v_0 - T^n v_0\| \leq \sum_{i=1}^n \|T^{i-1} v_0 - T^i v_0\| \leq n\|v_0 - Tv_0\| = n\|\dot{v}(0)\|$ and using the induction hypothesis we deduce

$$\phi_n(t) \leq e^{-t} \left[n\|\dot{v}(0)\| + \int_0^t e^{s/\lambda} \gamma_{n-1}(s) \, ds \right].$$

Hence it suffices to establish the inequality

$$n + \int_0^t e^s \sqrt{s + [(n-1) - s]^2} \, ds \leq e^t \sqrt{t + [n - t]^2}.$$

Since this holds trivially for $t = 0$, it suffices to prove the inequality for the derivatives

$$e^{t/\lambda} \sqrt{t + [(n-1) - t]^2} \leq e^t \left[\sqrt{t + [n - t]^2} + \frac{1 - 2[n - t]}{2\sqrt{t + [n - t]^2}} \right].$$

This easily verified by squaring both sides. ■

In particular if T is the resolvent J_λ^A, v is u_λ and using (6), we deduce that

$$\|(I+\lambda A)^{-n}x - u(t)\| \le \|A^0(u_0)\| \left(2\sqrt{\lambda t} + \sqrt{\lambda t + (n\lambda - t)^2}\right) \tag{16}$$

hence taking $\lambda = t/n$ we obtain an exponential approximation

$$\left\|\left(I + \frac{t}{n}A\right)^{-n} x - u(t)\right\| \le 3\frac{\|A^0(u_0)\|t}{\sqrt{n}}. \tag{17}$$

2.5 Discrete to continuous

Given a sequence $\{x_n\}$ in X along with a strictly increasing sequence $\{\sigma_n\}$ of positive numbers with $\sigma_0 = 0$ and $\sigma_n \to \infty$ as $n \to \infty$, one can construct a "continuous-time" trajectory x by interpolation: for $t \in [\sigma_n, \sigma_{n+1}]$, take $x(t)$ anywhere on the segment $[x_n, x_{n+1}]$. It is easy to see that any trajectory defined this way converges to some $\bar{x}$ if, and only if, the sequence $\{x_n\}$ converges to $\bar{x}$.
Observe that if the interpolation is chosen to be piecewise constant in each subinterval $[\sigma_n, \sigma_{n+1})$, then

$$\frac{1}{t}\int_0^t x(\xi)\ d\xi = \frac{1}{\sigma_n}\sum_{k=1}^n \lambda_k x_k,$$

where $\lambda_k = \sigma_k - \sigma_{k-1}$. The sum on the right-hand side of the previous equality represents an average of the points $\{x_n\}$ that is *weighted* by the sequence $\{\lambda_n\}$ and will be denoted by $\bar{x}_n$.

From now on we will consider only proximal or Euler sequences with stepsizes $\{\lambda_n\} \notin \ell^1$.

The next sections are devoted to the asymptotic analysis. We start by considering the sequences of values in the case $A = \partial f$ in Section 3. The rest deals with the behavior of trajectories and sequences themselves. Section 4 presents general tools related to weak convergence and properties of weak limit points. These last properties are easier to satisfy for the averages and are studied in Section 5. In Section 6 we present weak convergence, in particular in the framework of demipositive operators. Section 7 introduces different geometrical conditions that are sufficient for strong convergence. Section 8 is devoted to almost orbits and describes equivalence classes that allow to recover previous results with a new perspective and extend to non autonomous processes.

3 Convex optimization and convergence of the values

This section is devoted to the case $A = \partial f$ where we evaluate f on trajectories.

3.1 Continuous dynamics

When $A = \partial f$ with $f \in \Gamma_0(H)$ the differential inclusion (5) is a generalization of the gradient method, for nondifferentiable functions. In what follows let $u : [0, \infty) \to H$ be the solution of the differential inclusion

$$\dot{u}(t) \in -\partial f(u(t)), \tag{18}$$

whose existence is given in Theorem 10. Let

$$f^* = \inf_{x \in H} f(x) \in \mathbf{R} \cup \{-\infty\}.$$

The following result and its proof are essentially from [19, Brézis] (see [32, Güler]).

Proposition 21 *The function* $t \mapsto f(u(t))$ *is decreasing and* $\lim_{t\to\infty} f(u(t)) = f^*$.

Proof. The subdifferential inequality is

$$f(u(t)) - f(u(s)) \leq -\langle \dot{u}(t), u(t) - u(s)\rangle.$$

Thus

$$\limsup_{s\to t^-} \frac{f(u(t)) - f(u(s))}{t-s} \leq -\|\dot{u}(t)\|^2$$

and so the function $t \mapsto f(u(t))$ is decreasing. For each $z \in H$ and $s \in [0,t]$ the subdifferential inequality then gives

$$f(z) \geq f(u(s)) + \langle \dot{u}(s), u(s) - z\rangle \geq f(u(t)) + \frac{1}{2}\frac{d}{ds}\|u(s) - z\|^2.$$

Integrating on $[0,t]$ we obtain that

$$tf(z) \geq tf(u(t)) + \frac{1}{2}\|u(t) - z\|^2 - \frac{1}{2}\|u(0) - z\|^2$$

and so

$$f(u(t)) + \frac{\|u(t) - z\|^2}{2t} \leq f(z) + \frac{\|u(0) - z\|^2}{2t} \tag{19}$$

for every $z \in H$. ■

Comments
Inequality (19) shows that if $\mathcal{S} \neq \emptyset$ then $f(u(t))$ converges to f^* at a rate of $O(1/t)$. However, if the trajectory $u(t)$ is known to have a strong limit, then the rate drops to $o(1/t)$ (see [32, Güler]).

□

3.2 Proximal sequences

Let $\{x_n\}$ be a proximal sequence associated to $A = \partial f$. The following result is due to [31, Güler]:

Proposition 22 *The sequence* $f(x_n)$ *is decreasing and* $\lim_{n\to\infty} f(x_n) = f^*$.

Proof. The subdifferential inequality implies $f(x_{n-1}) - f(x_n) \geq \lambda_n\|y_n\|^2$ so that $f(x_n)$ is decreasing. Convergence of $f(x_n)$ to f^* follows from Lemma 23 below since $\sigma_n \to \infty$. ■

Lemma 23 *Let* $u \in dom f$, *then*

$$f(x_n) - f(u) \leq \frac{\|u - x_0\|^2}{2\sigma_n} - \frac{\|u - x_n\|^2}{2\sigma_n} - \frac{\sigma_n}{2}\|y_n\|^2.$$

Proof. The subdifferential inequality is

$$f(u) - f(x_n) \geq \langle u - x_n, -y_n\rangle = \frac{\langle u - x_n, x_{n-1} - x_n\rangle}{\lambda_n}$$

for all u in the domain of f. Thus

$$2\lambda_n(f(u) - f(x_n)) \geq \|u - x_n\|^2 + \lambda_n^2\|y_n\|^2 - \|u - x_{n-1}\|^2.$$

Summing up from 1 to n leads to

$$2\sigma_n f(u) - 2\sum_{k=1}^{n} \lambda_k f(x_k) \geq \|u - x_n\|^2 + \sum_{k=1}^{n} \lambda_k^2\|y_k\|^2 - \|u - x_0\|^2. \tag{20}$$

On the other hand the subdifferential inequality implies $f(x_{n-1}) - f(x_n) \geq \lambda_n\|y_n\|^2$. Multiplying by σ_{n-1} and rearranging we get

$$\sigma_{n-1}f(x_{n-1}) - \sigma_n f(x_n) + \lambda_n f(x_n) \geq \lambda_n\sigma_{n-1}\|y_n\|^2,$$

from which we derive

$$-\sigma_n f(x_n) + \sum_{k=1}^{n} \lambda_k f(x_k) \geq \sum_{k=1}^{n} \lambda_k\sigma_{k-1}\|y_k\|^2$$

by summation. Adding twice this inequality to (20) we obtain

$$2\sigma_n(f(u) - f(x_n)) \geq \|u - x_n\|^2 - \|u - x_0\|^2 + \sum_{k=1}^{n} \lambda_k^2\|y_k\|^2 + 2\sum_{k=1}^{n} \lambda_k\sigma_{k-1}\|y_k\|^2.$$

Recall from Lemma 11 that $\|y_n\|$ is decreasing. We get

$$\|y_n\|^2\sigma_n^2 = \|y_n\|^2 \sum_{k=1}^{n}(\lambda_k^2 + 2\lambda_k\sigma_{k-1}) \leq \sum_{k=1}^{n}(\lambda_k^2 + 2\lambda_k\sigma_{k-1})\|y_k\|^2$$

and the result follows at once by rearranging the terms. ■

Comments
If $\mathcal{S} \neq \emptyset$, Lemma 23 gives

$$\|y_n\| \leq \frac{d(x_0, \mathcal{S})}{\sigma_n}. \tag{21}$$

A similar estimation had been proved in [20, Brézis and Lions] but the right-hand side is $\sqrt{2}$ times larger.

The fact that $f(x_n) \to f^*$ had first been proved in [42, Martinet] when f is coercive and $\lambda_n \equiv \lambda$.

By Lemma 23, if $\mathcal{S} \neq \emptyset$ the rate of convergence can be estimated at $O(1/\sigma_n)$. Moreover, (21) and the subdifferential inequality together give

$$f(x_n) - f^* \leq \langle x^* - x_n, -y_n\rangle \leq \|x^* - x_n\| \; \|y_n\| \leq \frac{d(x_0, \mathcal{S})\|x^* - x_n\|}{\sigma_n}$$

for all $x^* \in \mathcal{S}$. Therefore, if the sequence $\{x_n\}$ is known to converge strongly, then $|f(x_n) - f^*| = o(1/\sigma_n)$. This was proved in [31, Güler] using a clever but unnecessarily sophisticated argument instead of inequality (21). □

3.3 Euler sequences

In this case the sequence $f(z_n)$ need not be decreasing. However, we have the following:

Lemma 24 *If either* $\sum \|z_{n+1} - z_n\|^2 < \infty$ *or* $\lim_{n\to\infty} \lambda_n\|w_n\|^2 = 0$, *then* $\liminf_{n\to\infty} f(z_n) = f^*$.

Proof. Since $-w_n \in \partial f(z_n)$, the subdifferential inequality and (15) together imply

$$\|z_{n+1} - y\|^2 \leq \|z_n - y\|^2 - 2\lambda_n(f(y) - f(z_n)) + \lambda_n^2\|w_n\|^2 \tag{22}$$

for each $y \in H$. If $\sum \|z_{n+1} - z_n\|^2 < \infty$ then

$$\sum \lambda_n(f(z_n) - f(y)) < \infty$$

(possibly $-\infty$). Since $\{\lambda_n\} \notin \ell^1$ one must have $\liminf_{n\to\infty} f(z_n) \leq f(y)$ for each $y \in H$.

On the other hand, inequality (22) can be rewritten as

$$\lambda_n \left[2(f(z_n) - f(y)) - \lambda_n\|w_n\|^2\right] \leq \|z_n - y\|^2 - \|z_{n+1} - y\|^2$$

so that

$$\sum \lambda_n(f(z_n) - f(y) - \lambda_n\|w_n\|^2) < \infty$$

and $\liminf_{n\to\infty} f(z_n) \leq f(y)$ for each $y \in H$. ■

A complementary result is the following from [59, Shor]:

Proposition 25 *Let dim$(H) < \infty$ and assume $\mathcal{S}$ is nonempty and compact. If $\lim_{n\to\infty} \lambda_n = 0$ and the sequence w_n is bounded then $\lim_{n\to\infty} f(z_n) = f^*$.*

Proof. By continuity, it suffices to prove that $\mathrm{dist}(z_n, \mathcal{S}) = \inf_{y\in\mathcal{S}} \|z_n - y\|$ tends to 0 as $n \to \infty$. For $\gamma > f^*$ define $L_\gamma = \{x : f(x) = \gamma\}$. Take $\varepsilon > 0$ and define

$$\delta(\varepsilon) = \min_{y\in\mathcal{S}} \mathrm{dist}(y, L_{f^*+\varepsilon}) \qquad \text{and} \qquad d(\varepsilon) = \max_{y\in\mathcal{S}} \mathrm{dist}(y, L_{f^*+\varepsilon}).$$

Observe that $0 < \delta(\varepsilon) \leq d(\varepsilon) \to 0$ as $\varepsilon \to 0$. By hypothesis and Lemma 24 there is $N \in \mathbf{N}$ such that $f(z_N) \leq f^* + \varepsilon$ and $\lambda_n\|w_n\| \leq \delta(\varepsilon)$ for all $n \geq N$. We shall prove that $\mathrm{dist}(z_n, \mathcal{S}) \leq 2d(\varepsilon)$ for all $n \geq N$. Since $\varepsilon > 0$ is arbitrary this shows that $\lim_{n\to\infty} \mathrm{dist}(z_n, \mathcal{S}) = 0$.

Indeed, if $f(z_n) \leq f^*+\varepsilon$ (this holds for $n = N$) then $\mathrm{dist}(z_n, \mathcal{S}) \leq d(\varepsilon)$ and $\mathrm{dist}(z_{n+1}, \mathcal{S}) \leq d(\varepsilon)+\delta(\varepsilon) \leq 2d(\varepsilon)$. On the other hand, if $f(z_n) > f^* + \varepsilon$ then $\mathrm{dist}(z_{n+1}, \mathcal{S}) \leq \mathrm{dist}(z_n, \mathcal{S})$. To see this, notice that if $y \in \mathcal{S}$ then $\langle \frac{w_n}{\|w_n\|}, y - z_n\rangle$ is the distance from y to the hyperplane $\Pi_n = \{x : \langle w_n, z_n - x\rangle\}$, so that

$$\langle w_n, y - z_n\rangle \geq \|w_n\|\mathrm{dist}(\mathcal{S}, \Pi_n) \geq \|w_n\|\mathrm{dist}(\mathcal{S}, L_{f(z_n)}) \geq \|w_n\|\delta(\varepsilon),$$

where the second inequality follows from convexity and the last one is true whenever $f(z_n) > f^* + \varepsilon$. Using (15) and recalling that $\lambda_n\|w_n\| \leq \delta(\varepsilon)$ we deduce that

$$\mathrm{dist}(z_{n+1}, \mathcal{S})^2 \leq \mathrm{dist}(z_n, \mathcal{S})^2 - \lambda_n\|w_n\|\delta(\varepsilon),$$

proving that $\mathrm{dist}(z_{n+1}, \mathcal{S}) \leq \mathrm{dist}(z_n, \mathcal{S})$. ■

Observe that this result does not require the stabilizing summability condition but it is necessary to make a very strong assumption on the set $\mathcal{S}$.

4 General tools for weak convergence

We denote by $\Omega[u(t)]$ (resp. $\Omega[x_n]$) the set of weak cluster points of a trajectory $u(t)$ as $t \to \infty$ (resp. of a sequence $\{x_n\}$ as $n \to \infty$).

Given a trajectory $u(t)$ we define

$$\bar{u}(t) = \frac{1}{t} \int_0^t u(\xi)\, d\xi$$

Similarly, given a sequence $\{x_n\}$ in H along with stepsizes $\{\lambda_n\} \notin \ell^1$, we introduce

$$\bar{x}_n = \frac{1}{\sigma_n} \sum_{k=1}^{n} \lambda_k x_k.$$

4.1 Existence of the limit

Most of the results on weak convergence that exist in the literature rely on the combination of two types of properties involving a subset $F \subset H$:

The first one is a kind of "Lyapounov condition" on the sequence or the trajectory like

(a1) $\|x_n - u\|$ converges to some $\ell(u)$ for each $u \in F$, or

(a2) $P_F(x_n)$ converges strongly (in all that follows F will be closed and convex).

These properties imply that the sequence is somehow "anchored" to the set F.

The second one is a global one, concerning the set of weak cluster points of the sequence or trajectory:

(b) $\Omega[x_n] \subset F$.

However, it is sometimes available only for the averages:

(b') $\Omega[\bar{x}_n] \subset F$.

The following result is a very useful tool for proving weak convergence of a sequence on the basis of **(a1)** and **(b)** above. It is known, especially in Hilbert spaces, as *Opial's lemma* [47].

Lemma 26 (Opial's Lemma) *Let $\{x_n\}$ be a sequence in H and let $F \subset X$. Assume*

1. *$\|x_n - u\|$ has a limit as $n \to \infty$ for each $u \in F$; and*
2. *$\Omega[x_n] \subset F$.*

Then x_n converges weakly to some $x^* \in F$.

Proof. Since $\{x_n\}$ is bounded it suffices to prove that it has only one weak cluster point. Let $x, y \in \Omega[x_n] \subset F$ so that $\|x_n - x\|$ converges to $\ell(x)$ and similarly for y. From

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

one deduces by choosing appropriate subsequences

$$\ell(y) = \ell(x) + \|x - y\|^2 \qquad (x_{\phi(n)} \rightharpoonup x)$$

and

$$\ell(y) = \ell(x) - \|x - y\|^2 \qquad (x_{\psi(n)} \rightharpoonup y)$$

hence $x = y$. ■

Comments
A Banach space X satisfies *Opial's condition* if it is reflexive and

$$\limsup_{n\to\infty} \|x_n - x\| < \limsup_{n\to\infty} \|x_n - y\| \quad \text{whenever} \quad x_n \rightharpoonup x \neq y. \tag{23}$$

holds. Any uniformly convex Banach space having a weakly continuous duality mapping (in particular, any Hilbert space) satisfies Opial's condition (see [47, Opial]). Opial's Lemma holds in any Banach space satisfying Opial's condition. □

Following [48, Passty], one obtains a more general result:

Lemma 27 *Let $\{x_n\}$ be a sequence in H with stepsizes $\{\lambda_n\}$ and let $F \subset X$. Assume* **(a1)** *: the sequence $\|x_n - u\|$ has a limit as $n \to \infty$ for each $u \in F$. Then the sets $\Omega[x_n] \cap F$ and $\Omega[\bar{x}_n] \cap F$ each contains at most one point. In particular if $\Omega[x_n] \subset F$ (resp. $\Omega[\bar{x}_n]$), then x_n (resp. $\bar{x}_n$) converges weakly as $n \to \infty$. A similar result holds for trajectories.*

Proof. Write
$$\|x_n - y\|^2 = \|x_n - x\|^2 + \|x - y\|^2 + 2\langle x_n - x, x - y\rangle$$
So that $\langle x_n, x - y\rangle$ converges to some $m(x, y)$ for any $x, y \in F$. If u and v belong to $\Omega[x_n] \cap F$ one obtains $\langle u, u - v\rangle = \langle v, u - v\rangle$ hence $u = v$. Similarly $\langle \bar{x}_n, x - y\rangle$ converges to $m(x, y)$. Thus both $\Omega[x_n] \cap F$ and $\Omega[\bar{x}_n] \cap F$ contain at most one point. ■

An alternative proof using **(a2)** and either **(b)** or **(b')** is as follows:

Lemma 28 *Let $\{x_n\}$ be a bounded sequence in H with stepsizes $\{\lambda_n\}$ and let $F \subset X$ be closed and convex. Assume* **(a2)***: $P_F x_n \to \zeta$ as $n \to \infty$. Then*

$$\Omega[x_n] \cap F = \Omega[\bar{x}_n] \cap F = \{\zeta\}.$$

In particular, if $\Omega[x_n] \subset F$ (resp. $\Omega[\bar{x}_n]$), then x_n (resp. $\bar{x}_n$) converges weakly to ζ. A similar result is true for trajectories.

Proof. By definition of the projection, for each $u \in F$ one has

$$\langle x_n - P_F x_n, u - P_F x_n\rangle \le 0.$$

Since x_n is bounded we deduce that
$$\langle x_n - \zeta, u - \zeta\rangle \le \rho_n$$
with $\lim_{n\to\infty} \rho_n = 0$. This implies $\Omega[x_n] \cap F = \{\zeta\}$ (if $v \in \Omega[x_n] \cap F$, take $u = v$). Similarly

$$\langle \bar{x}_n - \zeta, u - \zeta\rangle \le \bar{\rho}_n,$$

which gives $\Omega[\bar{x}_n] \cap F = \{\zeta\}$. ■

In our case the set F will always be $\mathcal{S}$, which is closed and convex.

4.2 Characterization of the limit: the asymptotic center

We show here that moreover the weak limit can be characterized.

Given a bounded sequence $\{x_n\}$ let

$$G(y) = \limsup_{n\to\infty} \|x_n - y\|^2$$

(for a trajectory $u(t)$ define $G(y) = \limsup_{t\to\infty} \|u(t)-y\|^2$). The function $G(y)$ is continuous, strictly convex and coercive. Its unique minimizer is called the *asymptotic center* (see [30]) of the sequence (resp. trajectory) and is denoted by $AC\{x_n\}$ (resp. $AC\{u(t)\}$).

Observe that, by virtue of Opial's condition (23), if $x_n \rightharpoonup x$ then $x = AC\{x_n\}$.

The weak limit of the average is still the asymptotic center, under some assumptions.

Proposition 29 *Assume* **(a1)**. *If $\bar{x}_n \rightharpoonup x \in F$, then $x = AC\{x_n\}$.*
The same property holds for trajectories.

Proof. For each $y \in H$ we have

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

Hence

$$\frac{1}{\sigma_n}\sum_{m=1}^{n} \lambda_m \|x_m - x\|^2 = \frac{1}{\sigma_n}\sum_{m=1}^{n} \lambda_m \|x_m - y\|^2 + 2\langle \bar{x}_n - y, y - x\rangle + \|y - x\|^2.$$

If $x_n \rightharpoonup x$ and $x \in F$ then $\ell(x) = \lim_{n\to\infty} \|x_n - x\|$ exists. Therefore,

$$G(x) = \ell(x)^2 \le \limsup_{n\to\infty} \left[\frac{1}{\sigma_n}\sum_{m=1}^{n} \lambda_m \|x_m - y\|^2\right] - \|x-y\|^2 \le G(y) - \|x-y\|^2$$

for each $y \in H$ so that $x = AC\{x_n\}$. ■

4.3 Characterization of the weak convergence

In this section we use the fact that the trajectories or sequences are generated through a maximal monotone operator.

Let us consider first the case $A = I - T$, where T is non expansive. The following result is in [51, Pazy]:

Proposition 30 *The sequence $T^n x$ converges weakly if, and only if, $\mathcal{S} \neq \emptyset$ and $\Omega[T^n x] \subset \mathcal{S}$.*

Proof. Assume $\mathcal{S} \neq \emptyset$. Given $u \in \mathcal{S}$, the sequence $\|T^n x - u\|$ is decreasing and so $T^n x$ is bounded. By Lemma 27, the fact that $\Omega[T^n x] \subset \mathcal{S}$ implies that $T^n x$ converges weakly. Conversely, since the sequence $\{T^n x\}$ is bounded, the argument in the proof of Theorem 42 shows that the weak limit of $T^n x$ must be in $\mathcal{S}$. ■

An alternative proof relies on the following result, which is interest in its own right:

Lemma 31 *Assume the sequence $U_n x = \frac{1}{n}(z + Tz + ... + T^{n-1}z)$ is bounded. Then $\emptyset \neq \Omega[U_n x] \subset \mathcal{S}$.*

Proof. For any $y \in H$ one has

$$\begin{aligned} 0 &\leq \|T^k x - y\|^2 - \|T^{k+1}x - Ty\|^2 \\ &= \|T^k x - Ty\|^2 - \|T^{k+1}x - Ty\|^2 + \|Ty - y\|^2 + 2\langle T^k x - Ty, Ty - y\rangle. \end{aligned}$$

By taking the average we obtain

$$0 \leq \frac{1}{n}\|x - Ty\|^2 + \|Ty - y\|^2 + 2\langle U_n x - Ty, Ty - y\rangle.$$

Therefore, if $p \in \Omega[U_n x]$, we can let $n \to \infty$ to deduce that

$$0 \leq \|Ty - y\|^2 + 2\langle p - Ty, Ty - y\rangle.$$

In particular, if $y = p$ we conclude that $\|Tp - p\|^2 \leq 0$ and so $p \in \mathcal{S}$. ■

Assuming that $\mathcal{S}$ is nonempty we can give a direct proof:

Lemma 32 *Assume $\mathcal{S} \neq \emptyset$. Then $T^n x \rightharpoonup p$ implies $p \in \mathcal{S}$.*

Proof. For any $y \in H$ and $u \in \mathcal{S}$

$$\begin{aligned} 0 &\leq \|T^k x - y\|^2 - \|T^{k+1}x - Ty\|^2 \\ &= \|T^k x - u\|^2 - \|T^{k+1}x - u\|^2 + \|u - y\|^2 - \|u - Ty\|^2 \\ &\quad +2\langle T^k x - u, u - y\rangle - 2\langle T^{k+1}x - u, u - Ty\rangle. \end{aligned}$$

Take $y = p$ and let $k \to \infty$. Since $\lim\limits_{k\to\infty} \|T^k x - p\|$ exists we get

$$0 \leq \|u - p\|^2 + 2\langle p - u, u - p\rangle - \|u - Tp\|^2 - 2\langle p - u, u - Tp\rangle.$$

which is precisely $\|Tp - p\|^2 \leq 0$ and implies $p \in \mathcal{S}$. ■

Following [52, Pazy], one obtains the continuous counterpart of Proposition 30:

Proposition 33 *The trajectory $S_t x$ converges weakly if, and only if, $\mathcal{S} \neq \emptyset$ and $\Omega[S_t x] \subset \mathcal{S}$.*

Proof. Assume $\mathcal{S} \neq \emptyset$. By Corollary 6 and Lemma 27, $\Omega[S_t x] \subset \mathcal{S}$ implies $S_t x$ converges weakly. It remains to prove that if $S_t x \rightharpoonup y$ then $y \in \mathcal{S}$. To see this, take any $[u, w] \in A$. We have

$$\begin{aligned} \|S_t x - u\|^2 - \|x - u\|^2 &\leq 2\int_0^t \langle w, u - S_s x\rangle ds \\ &= 2t\langle w, u - y\rangle + 2\int_0^t \langle w, y - S_s x\rangle \; ds. \end{aligned}$$

It suffices to divide by t and let $t \to \infty$ to obtain

$$0 \leq \langle w, u - y\rangle$$

so that $y \in \mathcal{S}$ by maximality. ■

Note that the proof uses the generator A (compare to the proof of the previous Proposition 30).

A last result, due to [24, Bruck], shows that if $\mathcal{S} \neq \emptyset$, then weak convergence is equivalent to weak asymptotic regularity. We follow [53, Pazy].

Proposition 34 *Assume $\mathcal{S} \neq \emptyset$. The trajectory $S_t x$ converges weakly if, and only if,*

$$S_{t+h}x - S_t x \rightharpoonup 0 \quad as \quad t \to \infty$$

for each $h \geq 0$. A similar result holds for the sequence $T^n x$.

Proof. For $u \in \mathcal{S}$ and $t > s$ we have

$$2\langle S_{s+h}x - u, S_s x - u\rangle - 2\langle S_{t+h}x - u, S_t x - u\rangle \leq \|S_{s+h}x - u\|^2 - \|S_{t+h}x - u\|^2 + \|u - S_s x\|^2 - \|u - S_t x\|^2.$$

Let $w \in \Omega[S_t x]$ and $h_k \to \infty$ with $S_{t+h_k} \rightharpoonup w$. Then $S_{s+h_k} \rightharpoonup w$ as well by weak asymptotic regularity. Thus we obtain

$$2\langle w - u, S_s x - S_t x\rangle \leq \|u - S_s x\|^2 - \|u - S_t x\|^2.$$

so that by **(a1)**, $\langle w - u, S_t x\rangle$ has a limit $L(w)$. In particular $w' \in \Omega[S_t x]$ implies $\langle w - u, w'\rangle = L(w)$ so that $\langle w - u, w' - w\rangle = 0$. Hence by symmetry $\langle w' - u, w - w'\rangle = 0$, thus $w = w'$ and $\Omega[S_t x]$ is reduced to one point. ■

5 Weak convergence in average

A trajectory $u(t)$ *converges in average* if

$$\bar{u}(t) = \frac{1}{t}\int_0^t u(\xi)\, d\xi \qquad \text{converges as} \quad t \to \infty.$$

Similarly, consider a sequence $\{x_n\}$ in H along with stepsizes $\{\lambda_n\}$, then $\{x_n\}$ *converges in average* if

$$\bar{x}_n = \frac{1}{\sigma_n}\sum_{k=1}^{n} \lambda_k x_k \qquad \text{converges as} \quad n \to \infty.$$

5.1 Continuous dynamics

Consider $x \in \overline{D(A)}$. In order to use the semigroup notation, let us introduce

$$\sigma_t x = \frac{1}{t}\int_0^t S_s x\, ds.^2$$

In order to prove that $\sigma_t x$ converges weakly as $t \to \infty$ we follow the ideas in [12, Baillon and Brézis]. We first prove that the projection $P_{\mathcal{S}} S_t x$ converges strongly to some v **(a2)**, next that weak cluster points of $\sigma_t x$ are in $\mathcal{S}$ **(b')**, and finally use Lemma 28 to conclude that $\sigma_t x$ converges weakly to v.

Lemma 35 *Assume $\mathcal{S} \neq \emptyset$. Then $P_{\mathcal{S}} S_t x$ converges strongly.*

Proof. Let $v(t) = P_{\mathcal{S}} S_t x$ and observe that the function $\psi(t) = \|v(t) - S_t x\|$ is decreasing:

$$\psi(t+h) \leq \|v(t) - S_{t+h}x\| = \|S_h v(t) - S_h S_t x\| \leq \psi(t).$$

Therefore, it has a limit as $t \to \infty$. On the other hand, the parallelogram equality gives

$$\|v(t+h) - v(t)\|^2 + 4\left\|\tfrac{v(t+h)+v(t)}{2} - S_{t+h}x\right\|^2 = 2\|v(t+h) - S_{t+h}x\|^2 + 2\|v(t) - S_{t+h}x\|^2.$$

[2]More generally $\sigma_n x = \int_0^\infty S_s x\, a_n(s)\, ds$ where a_n is the density of a positive probability measure on $\mathbf{R}^+$, which is assumed to be of bounded variation with $\int_0^\infty |da_n| \to 0$.

$\mathcal{S}$ convex implies $\left\| \frac{v(t+h)+v(t)}{2} - S_{t+h}x \right\| \geq \psi(t+h)$. We finally get

$$\|v(t+h) - v(t)\|^2 \leq 2\left[\psi(t)^2 - \psi(t+h)^2\right]$$

and conclude that $v(t)$ has a strong limit v as $t \to \infty$. ■

Lemma 36 *$\Omega[\sigma_t x] \subset \mathcal{S}$.*

Proof. Assume $\sigma_{t_k} x \rightharpoonup u$ as $k \to \infty$ and recall that $u(t) = S_t x$. For any $v \in D(A)$ we have

$$2\int_0^{t_k} \langle u(t) - v, \dot{u}(t)\rangle \ dt = \|u(t_k) - v\|^2 - \|x - v\|^2.$$

Now take $w \in Av$, so that $\langle u(t) - v, -w\rangle \geq \langle u(t) - v, \dot{u}(t)\rangle$. This gives

$$2\int_0^{t_k} \langle w, v - S_t x\rangle \ dt \geq \|S_{t_k} x - v\|^2 - \|x - v\|^2 \geq -\|x - v\|^2.$$

Divide by t_k and take the weak limit as $k \to \infty$. We get $\langle w, v - u\rangle \geq 0$ for any $[v, w] \in A$, so $0 \in Au$ by maximality. ■

Comments
Lemma 36 implies that if $\mathcal{S} = \emptyset$ then $\|\sigma_t x\| \to \infty$ for every $x \in \overline{D(A)}$ as $t \to \infty$. On the other hand, if $\mathcal{S} \neq \emptyset$ then every trajectory $S_t x$ is bounded, so $\sigma_t x$ is bounded for all $x \in \overline{D(A)}$. □

Using Lemma 28, Lemma 35 and Lemma 36 we finally obtain

Theorem 37 *If $\mathcal{S} \neq \emptyset$, then $\sigma_t x$ converges weakly to $v = \lim\limits_{t\to\infty} P_{\mathcal{S}} S_t x$.*

As a consequence of Proposition 29 one has

Proposition 38 *If $\mathcal{S} \neq \emptyset$, the limit $w - \lim\limits_{t\to\infty} \sigma_t x$ is the asymptotic center $AC\{S_t x\}$.*

Comments
Weak convergence in average is still true in uniformly convex Banach space with Fréchet-differentiable norm (see [55, Reich]) or satisfying Opial's condition (see [33, Hirano]). □

If $A = \partial f$ with $f \in \Gamma_0(H)$, convergence in average guarantees the convergence of the trajectory (see [22, Bruck]):

Proposition 39 *If $A = \partial f$ then $\lim\limits_{t\to\infty} \left\| u(t) - \frac{1}{t}\int_0^t u(s)\ ds \right\| = 0$.*

Proof. Integration by parts gives $u(t) - \frac{1}{t}\int_0^t u(s)\ ds = \frac{1}{t}\int_0^t s\dot{u}(s)\ ds$. $\|\dot{u}(t)\|$ being decreasing by Proposition 8, one has

$$\int_{t/2}^t s\|\dot{u}(s)\|^2\ ds \geq \|\dot{u}(t)\|^2 \int_{t/2}^t s\ ds = \frac{3}{8}t^2\|\dot{u}(t))\|^2.$$

But in the case $A = \partial f$, the function $t \mapsto t\|\dot{u}(t)\|^2$ is in $L^1(0, \infty)$ (see [18, Brézis]) which implies $\lim\limits_{t\to\infty} t\|\dot{u}(t)\| = 0$ and the result follows. ■

It is known that both the trajectory and the average converge weakly (Theorems 37 and 47). The preceding result implies, in particular, that the average cannot converge strongly unless the trajectory itself does.

5.2 Proximal sequences

Consider a proximal sequence $\{x_n\}$ in H along with stepsizes $\{\lambda_n\}$, and recall that $\bar{x}_n = \frac{1}{\sigma_n}\sum_{k=1}^{n}\lambda_k x_k$. The next result was presented in [40, Lions]:

Theorem 40 *Let $\mathcal{S} \neq \emptyset$. Then $\{x_n\}$ converges weakly in average to a point in $\mathcal{S}$.*

Proof. The case $\{\lambda_n\} \notin \ell^2$ will follow from Theorem 48, which states that $\{x_n\}$ converges weakly under this condition. Therefore we assume $\{\lambda_n\} \in \ell^2$ and check the conditions of Lemma 27 with $F = \mathcal{S}$: **(a1)** follows from Corollary 13, while **(b')** follows from Lemma 41 below. ■

Lemma 41 *Assume $\{\lambda_n\} \in \ell^2$, then $\Omega[\overline{x}_n] \subset \mathcal{S}$.*

Take $[u, v] \in A$ and use (3) so that

$$\|u - x_{n+1}\|^2 \leq \|u - x_n + \lambda_n v\|^2 = \|u - x_n\|^2 + \lambda_n^2\|v\|^2 - 2\langle v, \lambda_n x_n - \lambda_n u\rangle.$$

Summing up for $k = 1,\ 2,\ \ldots\ n$ and dividing by σ_n we obtain

$$2\langle v, \overline{x}_n - u\rangle \leq \frac{1}{\sigma_n}\|x_0 - u\|^2 + \frac{\tau_n}{\sigma_n}\|v\|^2.$$

If $\overline{x}_n \rightharpoonup \overline{x}$, then $\langle v, u - \overline{x}\rangle \geq 0$, hence $\overline{x} \in \mathcal{S}$ by maximality. ■

This is the counterpart of Lemma 36.

The extension to the sum of two operators is in [48, Passty].

5.3 Euler sequences

For nonexpansive mappings, weak convergence in average of the discrete iterates was established in [7, Baillon]. The proof is again of the form **(a2)** and **(b')** but note that the property $\mathcal{S} \neq \emptyset$ is not assumed but obtained during the proof.

Theorem 42 *Let T be a nonexpansive mapping on a bounded closed convex subset C of H. For every $z \in C$ the sequence $z_n = T^n z$ converges weakly in average to a fixed point of T, which is the strong limit of the sequence $P_{\mathcal{S}} T^n z$.*

Proof. Note that for any a and $a^i, i = 0, ..., n-1$, in H, the quantity

$$\left\| a - \frac{1}{n}\sum_{i=0}^{n-1} a^i \right\|^2 - \frac{1}{n}\sum_{i=0}^{n-1}\|a - a^i\|^2$$

is independent of a. Hence with $U_n z = \frac{1}{n}(z + Tz + ... + T^{n-1}z)$ one has

$$\begin{aligned}\|TU_n z - U_n z\|^2 &= \frac{1}{n}\sum_{i=0}^{n-1}\|TU_n z - T^i z\|^2 - \frac{1}{n}\sum_{i=0}^{n-1}\|U_n z - T^i z\|^2 \\ &\leq \frac{1}{n}\left(\|TU_n z - z\|^2 - \|U_n z - T^{n-1}z\|^2\right)\end{aligned}$$

so that

$$\|TU_nz - U_nz\| \leq \frac{1}{\sqrt{n}}\|TU_nz - z\|.$$

Thus $TU_nz - U_nz \to 0$ and if $U_nz \rightharpoonup u$ then $Tu = u$ by Proposition 3. It follows that $\Omega[U_nz] \subset \mathcal{S}$, which is **(b')** and $\mathcal{S} \neq \emptyset$. Since, for $u \in \mathcal{S}$, $\|T^nz - u\|$ decreases, then letting $V_nz = P_{\mathcal{S}}T^nz$, $\|T^nz - V_nz\|$ decreases as well, hence V_nz converges to some V (like in the proof of Lemma 35) which implies that $\Omega[U_nz] = \{V\}$ by Lemma 28. ■

Comments
The conclusion of Theorem 42 holds also if X is uniformly convex with Fréchet-differentiable norm and $\lambda_n \to 1$ or if X is superreflexive ([55, Reich]). □

By following an idea of Konishi (see [11, Baillon]) one can prove that the ergodic theorem for non expensive mappings implies in fact the analogous results for the semi-group:

Proposition 43 *Theorem 42 implies Theorem 37.*

Proof. Let $0 < h < t$ and $n = [t/h]$ the integer part of t/h and set $T_h = S_h$ and $U_nx = \frac{1}{n}\sum_{m=0}^{n-1} T^m x$. One has

$$t\sigma_t x = \int_0^h S_s x ds + ... + \int_{(n-1)h}^{nh} S_s x ds + \int_{nh}^t S_s x ds$$

and

$$\left\|\int_0^h S_s x ds - hx\right\| \leq \int_0^h \|S_s x - x\| ds.$$

Similarly

$$\left\|\int_{mh}^{(m+1)h} S_s x ds - hT_h^m x\right\| \leq \int_{mh}^{(m+1)h} \|S_s x - S_{mh}x\| ds \leq \int_0^h \|S_s x - x\| ds$$

hence

$$\|t\sigma_t x - nhU_n x\| \leq n\int_0^h \|S_s x - x\| ds + Mh,$$

where $\|S_s x\| \leq M$. Thus

$$\|\sigma_t x - U_n x\| \leq \frac{1}{h}\int_0^h \|S_s x - x\| ds + \frac{2M}{n}.$$

But as $t \to +\infty$, $U_n x$ converges weakly to a fixed point u_h of T_h by Theorem 42.
Let us now prove that u_h is a Cauchy net as $h \to 0$. Given $0 < h, h' < t$, $n = [t/h], n' = [t/h']$ one has

$$\|U_n x - U_{n'} x\| \leq \frac{1}{h}\int_0^h \|S_s x - x\| ds + \frac{2M}{n} + \frac{1}{h'}\int_0^{h'} \|S_s x - x\| ds + \frac{2M}{n'}.$$

Hence as $t \to +\infty$

$$\|u_h - u_{h'}\| \leq \frac{1}{h}\int_0^h \|S_s x - x\| ds + \frac{1}{h'}\int_0^{h'} \|S_s x - x\| ds,$$

thus u_h is a Cauchy net that converges to some u, since $\|S_s x - x\| \to 0$ as $s \to 0$. But $S_{mh}u_h = u_h$, so that given s and $h = s/m$ one has $S_s u_h = u_h$. As $m \to +\infty$ this implies $S_s u = u$, thus $u \in \mathcal{S}$. Now write, given $y \in H$

$$|\langle \sigma_t x - u, y\rangle| \leq |\langle \sigma_t x - U_n x, y\rangle| + |\langle U_n x - u_h, y\rangle| + \|u_h - u\|\|y\|$$

hence

$$|\langle\sigma_t x-u,y\rangle|\leq\left(\frac{1}{h}\int_0^h\|S_s x-x\|ds+\frac{2M}{n}\right)\|y\|+|\langle U_n x-u_h,y\rangle|+\|u_h-u\|\|y\|.$$

It follows that

$$\limsup_{t\to+\infty}|\langle\sigma_t x-u,y\rangle|\leq\left(\frac{1}{h}\int_0^h\|S_s x-x\|ds\right)\|y\|+\|u_h-u\|\|y\|$$

for all $h>0$. Letting $h\to 0$ we obtain $\sigma_t x\rightharpoonup u$. ■

Set $\bar{z}_n=\frac{1}{\sigma_n}\sum_{k=1}^n\lambda_k z_k$, where z_n is given in (13). A general result on convergence in average is the following from [23, Bruck]:

Theorem 44 *Assume $\sum\|z_n-z_{n-1}\|^2<\infty$. If $\mathcal{S}\neq\emptyset$, then z_n converges weakly in average to $w=\lim\limits_{n\to\infty}P_{\mathcal{S}}z_n$. Otherwise $\lim\limits_{n\to\infty}\|\bar{z}_n\|=\infty$.*

Proof. We first prove that $\Omega[\bar{z}_n]\subset\mathcal{S}$ which is **(b')**. Then we show, if $\mathcal{S}$ is non empty, that the sequence of projections $\zeta_n=P_{\mathcal{S}}z_n$ converge strongly to some $\zeta\in\mathcal{S}$ which is **(a2)** and finally that ζ is the only weak cluster point of the bounded sequence $\{\bar{z}_n\}$.

First, let $[u,v]\in A$ and set $w_n=(z_n-z_{n+1})/\lambda_n\in Az_n$. We have

$$\begin{aligned}\|z_{n+1}-u\|^2 &= \|z_n-\lambda_n w_n-u\|^2\\ &= \|z_n-u\|^2+\|\lambda_n w_n\|^2+2\lambda_n\langle w_n,u-z_n\rangle\\ &\leq \|z_n-u\|^2+\|\lambda_n w_n\|^2+2\lambda_n\langle v,u-z_n\rangle.\end{aligned}\tag{24}$$

Summing up, neglecting the positive term of the telescopic sum on the left-hand side and dividing by σ_n we get

$$0\leq\frac{\|z_1-u\|^2}{\sigma_n}+\frac{1}{\sigma_n}\sum_{k=1}^n\|z_k-z_{k-1}\|^2+2\langle v,u-\bar{z}_n\rangle.$$

Therefore $\liminf_{n\to\infty}\langle v,u-\bar{z}_n\rangle\geq 0$ and every weak cluster point of $\{\bar{z}_n\}$ lies in $\mathcal{S}$, by maximality.
Note that this is **(b')**, hence the counterpart of Lemma 36 and Lemma 41.

Next, take $u\in\mathcal{S}$. From equation (24) we get

$$\|z_{n+1}-u\|^2\leq\|z_n-u\|^2+\|\lambda_n w_n\|^2.\tag{25}$$

This implies the convergence of $\|z_{n+1}-u\|^2$ hence **(a1)** which ends the proof by using Lemma 27.
So the use of the following alternative is to identify the limit. This proof of **(a2)** parallels Lemma 35.
Using the parallelogram identity and the convexity of $\mathcal{S}$ we obtain

$$\begin{aligned}\|\zeta_{n+1}-\zeta_n\|^2 &= 2\|z_{n+1}-\zeta_n\|^2+2\|z_{n+1}-\zeta_{n+1}\|^2-4\|z_{n+1}-\tfrac{1}{2}(\zeta_n+\zeta_{n+1})\|^2\\ &\leq 2\|z_{n+1}-\zeta_n\|^2-2\|z_{n+1}-\zeta_{n+1}\|^2.\end{aligned}$$

Inequality (25) with $u=\zeta_n$ gives

$$0\leq\|\zeta_{n+1}-\zeta_n\|^2\leq 2\|\lambda_n w_n\|^2+2\|z_n-\zeta_n\|^2-2\|z_{n+1}-\zeta_{n+1}\|^2.$$

This implies that the sequence $\{\|z_n-\zeta_n\|^2+\rho_n\}$ decreases, where $\rho_n=\sum_{k\geq n}\|\lambda_k w_k\|^2$, which tends to 0 as $n\to\infty$. Since

$$0\leq\|\zeta_{n+p}-\zeta_n\|^2\leq 2\rho_n+2\|z_n-\zeta_n\|^2-2\|z_{n+p}-\zeta_{n+p}\|^2,$$

the sequence $\{\zeta_n\}$ is Cauchy and converges as well to some ζ in $\mathcal{S}$. The results now follows from Lemma 28. ■

Observe that the same structure of proof could be applied to proximal sequences.

For a similar proof with two operators and forward-backward procedure see [48, Passty].

The following result due to [53, Pazy] of **(b')** leads to a unified proof of weak convergence in average for contractions in the discrete (Theorem 42) or continuous case (Theorem 37). Note that the first step assumes $\mathcal{S} \neq \emptyset$ and then one uses **(a1)** to achieve the result.

Proposition 45 *Assume $\mathcal{S} \neq \emptyset$, then $\Omega[\sigma_t x] \subset \mathcal{S}$.*

Proof. For $t, h \geq 0$ we have

$$\begin{aligned} 0 &\leq \|S_t x - y\|^2 - \|S_{t+h}x - S_h y\|^2 \\ &= \|S_t x - S_h y\|^2 - \|S_{t+h}x - S_h y\|^2 + 2\langle S_t x - S_h y, S_h y - y\rangle + \|S_h y - y\|^2. \end{aligned}$$

By taking the average we deduce that

$$0 \leq \frac{1}{t}\int_0^t [\|S_s x - S_h y\|^2 - \|S_{s+h}x - S_h y\|^2]ds + 2\langle \sigma_t x - S_h y, S_h y - y\rangle + \|S_h y - y\|^2.$$

Since $\mathcal{S} \neq \emptyset$, $\|S_t x - S_h y\|$ is bounded, hence letting $t \to +\infty$, it follows that for any $p \in \Omega[\sigma_t x]$, any $h \geq 0$ and any $y \in H$

$$0 \leq 2\langle p - S_h y, S_h y - y\rangle + \|S_h y - y\|^2.$$

Finally take $y = p$ so that $p = S_h p$, which means $p \in \mathcal{S}$. ■

6 Weak convergence

Not all maximal monotone operators generate weakly convergent trajectories.

Example 3 Let $R : \mathbf{R}^2 \to \mathbf{R}^2$ be the counterclockwise $\pi/2-$rotation and consider the evolution scheme defined by the differential equation:

$$\dot{u}(t) = R(u(t)).$$

Note that $\mathcal{S} = \{0\}$. The orbit starting at time $t = 0$ from the point $u_0 = r_0(\cos(\theta_0), \sin(\theta_0))$, $r > 0$ is described by $u(t) = r_0(\cos(t - \theta_0), \sin(t - \theta_0))$, which is bounded but does not have a limit as $t \to \infty$. However, the average $\frac{1}{t}\int_0^t u(s)\, ds$ converges to 0 as $t \to \infty$, by Theorem 37.

Now let $x_n = r_n(\cos\theta_n, \sin\theta_n)$ satisfy

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

We have $r_{n+1}^2 = \prod_{k=1}^n (1 + \lambda_k^2)^{-1} r_0$ and $\theta_n = \theta_0 + \sum_{k=1}^n \arctan(\lambda_k)$. The sequence r_n is decreasing. If $\lambda_n \notin \ell^2$ then $\lim\limits_{n\to\infty} x_n = 0$; otherwise it stays bounded away from zero. On the other hand, the argument θ_n is increasing. It converges if $\lambda_n \in \ell^1$ and diverges otherwise. Observe also that x_n converges in average to 0 as $n \to \infty$, by Theorem 40.

Finally, let $z_n = \rho_n(\cos\phi_n, \sin\phi_n)$ satisfy

$$\frac{z_{n+1} - z_n}{\lambda_n} = R(z_n).$$

Here $\rho_{n+1}^2 = \prod_{k=1}^n (1+\lambda_k^2)\rho_0$ and $\phi_n = \phi_0 + \sum_{k=1}^n \arctan(\lambda_k)$. In this case the sequence r_n is increasing. It remains bounded and is convergent if, and only if, $\lambda_n \in \ell^2$. The argument θ_n is increasing as well. It converges if $\lambda_n \in \ell^1$ and diverges otherwise. As before, z_n converges in average to 0 as $n \to \infty$, by Theorem 44 . □

Tools

Assuming $\mathcal{S}$ non empty and using Lemma 27, by virtue of Corollaries 6, 13 and 19, in order to prove weak convergence of $u(t)$, it suffices to verify that its set of weak cluster points lie in $\mathcal{S}$ (condition **(b)**). The key tool is the concept of *demipositivity*, first developed in [22, Bruck].

A maximal monotone operator A is *demipositive* if there exists $w \in \mathcal{S}$ such that for every sequence $\{u_n\} \in D(A)$ converging weakly to u and every bounded sequence $\{v_n\}$ such that $v_n \in Au_n$

$$\langle v_n, u_n - w\rangle \to 0 \qquad \text{implies} \qquad u \in \mathcal{S}. \tag{26}$$

Proposition 46 *Each of the following conditions is sufficient for a maximal monotone operator A to be demipositive:*

1. *$A = \partial\phi$, where ϕ is a proper lower-semicontinuous convex function having minimizers ($\mathcal{S} \neq \emptyset$).*
2. *$A = I - T$, where T is nonexpansive and has a fixed point ($\mathcal{S} \neq \emptyset$).*
3. *The set $\mathcal{S}$ has nonempty interior.*
4. *A is odd and* firmly positive*, which means that there is $w \in \mathcal{S}$ such that $v \in Au$ and $\langle v, u - w\rangle = 0$ together imply $0 \in Ax$.*
5. *A is firmly positive and sequentially weakly closed (its graph is sequentially weak/weak closed).*
6. *$\mathcal{S} \neq \emptyset$ and A is 3-monotone, which means that $\sum_{n=1}^3 \langle y_n, x_n - x_{n-1}\rangle \geq 0$ for every set $\{[x_n, y_n] \mid 1 \leq n \leq 3\} \subset A$ ($x_0 \equiv x_N$).*

For demipositivity in Banach spaces see [26, Bruck and Reich].

Comments
We just mention another assumption that guarantees that the weak cluster points will lie in $\mathcal{S}$: Let S the semi-group generated by A. A satisfies condition (L) if

$$\lim_{t\to\infty} \|A^0 S_t x\| \leq \lim_{t\to\infty} \left(\frac{1}{h}\|S_{t+h}x - S_t x\|\right)$$

for every $h > 0$ and $x \in D(A)$. An equivalent formulation is the following: Denote by a^0 the element of minimal norm in $\overline{R(A)}$. Then A satisfies condition (L) if, and only if, for every $x \in D(A)$ one has

$$\lim_{t\to\infty} A^0 S_t x = a^0.$$

Unlike demipositivity, this does not impose *a priori* that $\mathcal{S} \neq \emptyset$. For instance, if $A = \partial f$ with $f \in \Gamma_0(H)$ or if $A = I - T$ with T nonexpansive, then A satisfies condition (L) but is not demipositive unless $\mathcal{S} \neq \emptyset$.
Condition (L) is essentially used in [50, Pazy] to prove that the weak cluster points of the trajectory $S_t x$ lie in $\mathcal{S}$. If $\mathcal{S} = \emptyset$ one immediately deduces that $\lim_{t\to\infty} \|S_t x\| = \infty$. The interested reader may find this definition and related results in [50, Pazy]. □

6.1 Continuous dynamics

The following classical result of weak convergence for demipositive operators was proved in [22, Bruck].

Theorem 47 *If A is demipositive then $u(t)$ converges weakly as $t \to \infty$ to an element of $\mathcal{S}$.*

Proof. By Corollary 6 and Opial's Lemma it suffices to prove $\Omega[u(t)] \subset \mathcal{S}$, which is **(b)**. Let $w \in \mathcal{S}$ satisfy (26) and let $u(t_n) \rightharpoonup u$ as $n \to \infty$. The sequence $\dot{u}(t_n)$ is bounded by Theorem 10. Let $\theta_w(t) = \frac{1}{2}\|u(t) - w\|^2$, thus $\dot{\theta}_w(t) = \langle \dot{u}(t_n), u(t_n) - w\rangle$. Since θ_w is bounded by Corollary 6, $\dot{\theta}_w \in L^1$ and there is a subsequence t_{n_k} of t_n such that $\dot{\theta}_w(t_{n_k}) \to 0$ as $k \to \infty$. But $u(t_{n_k}) \rightharpoonup u$, so $u \in \mathcal{S}$ by demipositivity. ∎

Comments
Theorem 47 was extended in [49, Passty] to the class of φ-*demipositive* operators. □

6.2 Proximal sequences

A first detailed study of the asymptotic behavior of the proximal sequence $\{x_n\}$ was performed in [57, Rockafellar], when the stepsizes are bounded away from zero. The author also considers an inexact version of the algorithm. The next convergence results under more general hypotheses are investigated in [20, Brézis and Lions].

Recall that $\sigma_n = \sum_{m \le n} \lambda_m$ and $\tau_n = \sum_{m \le n} \lambda_m^2$.

Theorem 48 *Assume $\mathcal{S} \neq \emptyset$. If $\{\lambda_n\} \notin \ell^2$ then x_n converges weakly to some $x^* \in \mathcal{S}$. Moreover, $\|y_n\| \le d(x_0, \mathcal{S})\tau_n^{-1/2}$.*

Proof. By Lemmas 11 and 12, we have for any $x \in \mathcal{S}$

$$\|y_n\|^2 \tau_n \le \sum_{k \le n} \lambda_k^2 \|y_k\|^2 \le \|x_0 - x\|^2.$$

$\tau_n \to \infty$ implies $\|y_n\| \to 0$. Since $-y_n \in Ax_n$, we deduce that $\Omega[x_n] \subset \mathcal{S}$, which is **(b)**, by Proposition 3. We conclude by Corollary 13 and Opial's Lemma 26. ∎

The following result, adding the demipositivity hypothesis, is also from [20, Brézis and Lions]:

Theorem 49 *If A is demipositive then x_n converges weakly to some $x^* \in \mathcal{S}$.*

Proof. As above, using Corollary 13 the result follows from Opial's Lemma 26 if $\Omega[x_n] \subset \mathcal{S}$ which is **(b)**. Let $x_{n_k} \rightharpoonup x$ and w be the element in $\mathcal{S}$ used in the definition of demipositivity (26). Using Lemma 50 below we construct another subsequence $\{x_{m_k}\}$ such that both $\|x_{m_k} - x_{n_k}\|$ and $\langle x_{m_k} - w, y_{m_k}\rangle$ tend to 0 as $k \to \infty$. Since $x_{m_k} \rightharpoonup x$ and A is demipositive, x must belong to $\mathcal{S}$. ∎

Lemma 50 *Let $\{x_n\}$ be a proximal sequence and $w \in \mathcal{S}$. For each $\varepsilon > 0$, there is N such that: for any $n \ge N$, there exists $m \in \mathbf{N}$ satisfying $N \le m \le n$, $\|x_m - x_n\| \le \varepsilon$ and $\langle -y_m, x_m - w\rangle \le \varepsilon$.*

Proof. For each $w \in \mathcal{S}$ we have $\|x_{k-1} - w\|^2 \geq \|x_k - w\|^2 + 2\lambda_k \langle -y_k, x_k - w\rangle$ and so

$$\sum_k \lambda_k \langle y_k, w - x_k\rangle < \infty \tag{27}$$

where all terms are nonnegative by monotonicity. Given $\varepsilon > 0$, define $P = \{k \in \mathbf{N} \mid \langle y_k, w - x_k\rangle \geq \varepsilon\}$ so that $\sum_{k\in P} \lambda_k < \infty$. Since $\|x_{k-1} - x_k\| = \lambda_k \|y_k\|$, Lemma 11 implies $\sum_{k\in P} \|x_{k-1} - x_k\| < \infty$.
Let N_1 so that $\sum_{k\in P, k\geq N_1} \|x_{k-1} - x_k\| < \varepsilon$. By virtue of (27), since $\{\lambda_n\} \notin \ell^1$ there is $N \geq N_1$ with $\langle y_N, w - x_N\rangle \leq \varepsilon$. Consider $n \geq N$: if $n \notin P$ we choose $m = n$. If $n \in P$, let $m = \max\{k < n \mid k \notin P\}$. Since $m \geq N_1$ and all integers between m and n are in P, we have $\|x_m - x_n\| \leq \sum_{m<k\leq n} \|x_{k-1} - x_k\| \leq \varepsilon$. ■

Comments

1. Theorem 49 is still true if the sequence satisfies $\|x_n - (I + \lambda_n A)^{-1} x_{n-1}\| \leq \varepsilon_n$ with $\sum \varepsilon_n < \infty$. This is proved in [20, Brézis and Lions] and can also be derived using asymptotic equivalence results in Section 8 (see [2, 3]).
2. In uniformly convex Banach spaces with Fréchet differentiable norm there is weak convergence in the following cases (see [56, Reich]):
   (a) $\{\lambda_n\}$ does not converge to zero, or
   (b) The modulus of convexity of the space satisfies $\delta(\varepsilon) \geq K\varepsilon^p$ for some $K > 0$ and $p \geq 2$ and $\sum \lambda_n^p = \infty$.
3. Demipositive can be replaced by φ-demipositive (see [49, Passty]). □

6.3 Euler sequences

Let $\{z_n\}$ be an Euler sequence and recall that $w_n = \frac{z_{n+1} - z_n}{\lambda_n} \in -Az_n$.

Theorem 51 *Let A be demipositive and assume $\{\lambda_n\} \in \ell^2$ and $\{w_n\}$ bounded. Then z_n converges weakly to some $z \in \mathcal{S}$.*

Proof. If $y \in \mathcal{S}$, Corollary 19 shows that the sequence $\|z_n - y\|$ is convergent. On the other hand, equality (15) implies $\sum_{n\geq 1} \lambda_n \langle w_n, y - z_n\rangle < \infty$. One concludes as in Theorem 49 using an analogue of Lemma 50. ■

Comments
The previous result from [26, Bruck and Reich] works for demipositive operators in "a few" Banach spaces, namely $X = L^{2m}$, $m \in \mathbf{N}$ or $X = \ell^p$, $p \in (1, \infty)$. □

A related result from [56, Reich] is the following (and holds in uniformly convex Banach spaces with Fréchet-differentiable norm):

Proposition 52 *Let T be non-expansive, $A = I - T$ and $\{\lambda_n\}$ satisfying $0 \leq \lambda_n \leq 1$ and $\sum \lambda_n(1 - \lambda_n) = \infty$. If $\mathcal{S} \neq \emptyset$ then $\{z_n\}$ converges weakly to a point in $\mathcal{S}$.*

If $A = \partial f$ with $f \in \Gamma_0(H)$ and $\dim(H) < \infty$ one can circumvent the difficulties of Lemma 50 and provide a simpler proof of Theorem 51. Let

$$z_{n+1} \in z_n - \lambda_n \partial f(z_n)$$

and w_n as above.

Theorem 53 *Assume $\mathcal{S} \neq \emptyset$ and dim$(H) < \infty$. If $\sum \|z_n - z_{n-1}\|^2 < \infty$ then z_n converges to a minimizer of f.*

Proof. Lemma 24 gives $\liminf_{n\to\infty} f(z_n) = f^*$. Since $\{z_n\}$ is bounded and the space is finite dimensional, there is a subsequence $\{z_{n_k}\}$ such that $\lim_{k\to\infty} f(z_{n_k}) = f^*$ and $\lim_{k\to\infty} \|z_{n_k} - z\| = 0$ for some $z \in H$. Since z must be in $\mathcal{S}$ by lower-semicontinuity, Corollary 19 implies $\lim_{n\to\infty} \|z_n - z\| = 0$, which means z_n converges to z. ∎

The preceding result from [58] was pointed out to the authors by R. Cominetti.

7 Strong convergence

Even if $A = \partial\phi$ with $\phi \in \Gamma_0(H)$ having minimizers, the trajectory $u(t)$ need not converge strongly as $t \to \infty$. This is shown by Baillon's example in [10, Baillon]: the author defines a function $\phi \in \Gamma_0(\ell^2)$ having minimizers and proves that the trajectories converge weakly but not strongly.

This also true for the proximal point algorithm. Even if $A = \partial\phi$ with $\phi \in \Gamma_0(H)$ having minimizers, a sequence satisfying (7) need not converge strongly. This was proved in [31, Güler] using Baillon's example and the equivalence techniques from [49, Passty]. A simpler example of this type can be found in [14] and can be retranslated to provide a new counterexample for strong convergence of the continuous trajectory, different from that of Baillon.

Conditions

We introduce here a series of conditions, mainly of geometric nature, that will be used to obtain strong convergence of the process in the continuous or discrete set-up.

Strong monotonicity. Let $\alpha > 0$. An operator A is *α-strongly monotone* if for all $[x, x^*], [y, y^*] \in A$ one has

$$\langle x^* - y^*, x - y\rangle \geq \alpha\|x - y\|^2.$$

Observe that if A is strongly monotone and $Ax \cap Ay \neq \emptyset$, then $x = y$. If A is α-strongly monotone then $J^A_{1/\alpha}$ is a strict contraction. Therefore it has a fixed point p and only one, say $\mathcal{S} = \{p\}$. Strongly monotone operators are demipositive.

Clearly, if A is monotone, then $A + \alpha I$ is α-strongly monotone. Also, subdifferentials of proper, lower-semicontinuous strongly convex functions are strongly monotone.

A weaker notion of strong monotonicity found for instance in [50, Pazy] is the following: A is α-strongly monotone if $\mathcal{S} \neq \emptyset$ and

$$\langle A^0x, x - P_{\mathcal{S}}x\rangle \geq \alpha\|x - P_{\mathcal{S}}x\|^2$$

for every $x \in D(A)$. In this case the set $\mathcal{S}$ need not be a singleton. Proposition 54 below also holds if A is strongly monotone in this sense but the proof is more involved.

Solution set $\mathcal{S}$ with nonempty interior. If $p \in \text{int}\mathcal{S}$ then there is $r > 0$ such that the ball $B(p, r)$ of radius r centered at p is contained in $\mathcal{S}$. Then $\langle u^*, u - p + rh\rangle \geq 0$ for all $[u, u^*] \in A$ and all $h \in H$ with $\|h\| \leq 1$. Therefore $\langle u^*, u - p\rangle \geq r\langle u^*, -h\rangle$ and

$$r\|u^*\| = r \sup_{\|h\|\leq 1} \langle u^*, -h\rangle \leq \langle u^*, u - p\rangle. \tag{28}$$

The NR convergence condition. A maximal monotone operator A on H satisfies the *NR convergence condition* if $\mathcal{S} \neq \emptyset$ and for every bounded sequence $[x_n, y_n] \in A$ one has

$$\liminf_{n\to\infty} \langle y_n, x_n - P_{\mathcal{S}}x_n\rangle = 0 \qquad \text{implies} \qquad \liminf_{n\to\infty} \|x_n - P_{\mathcal{S}}x_n\| = 0.$$

Strongly monotone operators satisfy this condition. So do operators having compact resolvent (see below) and those satisfying $\langle y, x - P_{\mathcal{S}}x\rangle > 0$ for all $[x, y] \in A$ such that $x \notin \mathcal{S}$.

The NR convergence condition can be easily stated in a Banach space X by means of the duality mapping. The results below hold when both X and X^* are uniformly convex. The interested reader can consult [46, Nevanlinna and Reich] and [26, Bruck and Reich].

Compactness. The strong ω-limit set of a trajectory $u : [0, \infty) \to H$ is the set $\omega[u(t)] = \bigcap_{t>0} \overline{\{u(s) : s \geq t\}}$. For a sequence $\{x_n\}$ it is defined by $\omega[x_n] = \bigcap_{n\in\mathbf{N}} \overline{\{x_k : k \geq n\}}$.

By virtue Lemma 27 the sets $\omega[u(t)] \cap \mathcal{S}$ and $\omega[x_n] \cap \mathcal{S}$ contain, at most, one element.

If $\mathcal{S} \neq \emptyset$ and J_1^A is a compact operator (maps bounded sets to relatively compact sets) then $\omega[u(t)] \neq \emptyset$ for every trajectory u satisfying (5) (see Theorem 11.8 in [50, Pazy]) and $\omega[x_n] \neq \emptyset$ for every sequence $\{x_n\}$ satisfying (7).

For instance, if $A = \partial f$ and the set $\{ u \in H \,|\varphi(u) + \|u\|^2 \leq M \}$ is compact for each $M \geq 0$, then J_1^A is compact. This case was first studied in [19, Brézis].

Symmetry. An operator A is odd if $D(A) = -D(A)$ and $A(-x) = -Ax$ for all $x \in D(A)$.[3] This is the case, for instance, if $A = \partial f$ and f is even. If A is odd, the semigroup generated is odd as well (see, for instance, [50, Pazy]). On the other hand, it is easy to see that J_λ^A is odd for each $\lambda > 0$ if A is odd.

Notice also that if A is odd then $\mathcal{S} \neq \emptyset$. Moreover, $0 \in \mathcal{S}$. To see this, take $x \in D(A)$ and let $[x, y], [-x, -y] \in A$. We have

$$\begin{aligned} 4\langle y - 0, x - 0\rangle &= \langle y + y, x + x\rangle \\ &= \langle y - (-y), x - (-x)\rangle \\ &\geq 0. \end{aligned}$$

Then $0 \in A0$ by Lemma 1.

Asymptotic regularity. A trajectory u is *asymptotically regular* if $\lim_{t\to\infty} \|u(t + h) - u(t)\| = 0$ for each $h \geq 0$. A sequence $\{x_n\}$ is asymptotically regular if $\lim_{n\to\infty} \|x_{n+m} - x_n\| = 0$ for each $m \in \mathbf{N}$.

Comments
Recall that the notion of weak asymptotic regularity was mentioned in Proposition 34 as a characterization of weak convergence of the trajectories satisfying (5). □

7.1 Continuous dynamics

Strong monotonicity.

Proposition 54 *If A is α-strongly monotone for some $\alpha > 0$ then $u(t)$ converges strongly to the unique $p \in \mathcal{S}$ as $t \to \infty$.*

Proof. Strong monotonicity implies

$$\frac{1}{2}\frac{d}{dt}\|u(t) - p\|^2 = \langle \dot{u}(t), u(t) - v(t)\rangle \leq -\alpha\|u(t) - p\|^2$$

[3] A weaker notion is that $A^0(-x) = -A^0x$. The results below still hold but the proofs become more technical.

and so $\|u(t) - p\| \leq e^{-2\alpha t}\|u_0 - p\|$. ■

Comments
The previous result can be extended in the following way: Let X be a Banach space such that X and X^* are uniformly convex. In [46, Nevanlinna and Reich] the authors prove that if A satisfies NR convergence condition then $u(t)$ converges strongly to a point in $\mathcal{S}$ as $t \to \infty$. If only X^* is uniformly convex, the result remains true provided Ax is proximinal and convex for every x (see [26, Bruck and Reich]). If neither X nor X^* is uniformly convex, the result is still true if the semigroup is differentiable (see [46, Nevanlinna and Reich]). □

Solution set with nonempty interior.

Proposition 55 *Assume int $\mathcal{S} \neq \emptyset$. Then $u(t)$ converges strongly as $t \to \infty$ to a point in $\mathcal{S}$.*

Proof. If $B(p, r) \subset \mathcal{S}$, inequality (28) implies

$$\begin{aligned}
r\|u(t) - u(s)\| &\leq r \int_s^t \|\dot{u}(\tau)\|\, d\tau \\
&\leq -\int_s^t \langle \dot{u}(\tau), u(\tau) - p\rangle\, d\tau \\
&= \frac{1}{2}\|u(s) - p\|^2 - \frac{1}{2}\|u(t) - p\|^2.
\end{aligned}$$

Since $\|u(t) - p\|$ is convergent by Corollary 6, $u(t)$ has the Cauchy property. ■

Comments
Theorem 4 in [46, Nevanlinna and Reich] shows that this result remains true if X and X^* are uniformly convex. In the same paper, the authors give a counterexample in $\mathcal{C}([0, 1]; \mathbf{R})$. See also [26, Bruck and Reich]. □

Compactness.

Proposition 56 *If $\omega[u(t)] \cap \mathcal{S} \neq \emptyset$ then $u(t)$ converges strongly to some $p \in \mathcal{S}$.*

Proof. If $p \in \omega[u(t)] \cap \mathcal{S}$ then $\|u(t) - p\|$ is decreasing and $\liminf_{t\to\infty} \|u(t) - p\| = 0$. Hence $u(t) \to p$ as $t \to \infty$. ■

Comments
If $\mathcal{S}$ has nonempty interior then A is demipositive and $\omega[u(t)] \neq \emptyset$ for every trajectory u satisfying (5). Every strong cluster point is also a weak cluster point, that must lie in $\mathcal{S}$ by demipositivity. Hence $\omega[u(t)] \cap \mathcal{S} \neq \emptyset$ and Proposition 55 can also be deduced from Proposition 56. □

Symmetry.

Proposition 57 *If $A = \partial f$ and $f \in \Gamma_0(H)$ is even then $u(t)$ converges strongly as $t \to \infty$ to a point in $\mathcal{S}$.*

Proof. Take $s > 0$ and define $\gamma(t) = \|u(t)\|^2 - \|u(s)\|^2 - \frac{1}{2}\|u(t) - u(s)\|^2$. For $t \in [0, s]$ one has

$$\dot{\gamma}(t) = \langle \dot{u}(t), u(t) + u(s) \rangle \leq f(-u(s)) - f(u(t)) = f(u(s)) - f(u(t)) \leq 0.$$

Therefore, $\gamma(t) \geq \gamma(s) = 0$ and so

$$\frac{1}{2}\|u(t) - u(s)\|^2 \leq \|u(t)\|^2 - \|u(s)\|^2.$$

Since $0 \in \operatorname{Argmin} f$, $\|u(t)\|$ converges as $t \to \infty$ so $u(t)$ has the Cauchy property. ■

For general A one has to assume additional hypotheses on the trajectory:

Proposition 58 *Let A be odd. If u is asymptotically regular then $u(t)$ converges strongly to some $p \in \mathcal{S}$ as $t \to \infty$.*

Proof. Let us use the semigroup notation $u(t) = S_t x$. If A is odd then $0 \in \mathcal{S}$ and

$$\begin{aligned} \|S_{t+h+s}x + S_{t+s}x\| &= \|S_{t+h+s}x - S_{t+s}(-x)\| \\ &\leq \|S_{t+h}x - S_t(-x)\| \\ &= \|S_{t+h}x + S_t x\| \end{aligned}$$

for each $h \geq 0$ so that

$$\lim_{t\to\infty} \|S_t x + S_{t+h}x\| \leq \|S_t x + S_{t+h}x\|. \tag{29}$$

Since $0 \in \mathcal{S}$ the limit $d = \lim\limits_{t\to\infty} \|S_t x\|$ exists. Moreover, the fact that $\|2S_t x\| \leq \|S_t x + S_{t+h}x\| + \|S_t x - S_{t+h}x\|$ implies

$$2d \leq \lim_{t\to\infty} \|S_t x + S_{t+h}x\| \leq \|S_t x + S_{t+h}x\|$$

for each t, h by asymptotic regularity and inequality (29). Finally,

$$\begin{aligned} \|S_{t+h}x - S_t x\|^2 &= 2\|S_t x\|^2 + 2\|S_{t+h}x\|^2 - \|S_{t+s}x + S_t x\|^2 \\ &\leq 4\|S_t x\|^2 - 4d^2 \end{aligned}$$

and so $\{S_t x\}$ has the Cauchy property. ■

Comments
Without the asymptotic regularity assumption, strong convergence holds for the averages when S is odd, as proved in (see [8, Baillon]). □

7.2 Proximal sequences

Strong monotonicity.

Proposition 59 *If A is α-strongly monotone for some $\alpha > 0$ then x_n converges strongly to the unique $p \in \mathcal{S}$ as $n \to \infty$.*

Proof. Strong monotonicity implies

$$
\begin{aligned}
\alpha\lambda_n\|x_n-p\|^2 &\le \langle x_{n-1}-x_n, x_n-p\rangle \\
&= \langle x_{n-1}-p, x_n-p\rangle - \|x_n-p\|^2 \\
&\le \|x_n-p\|\left(\|x_{n-1}-p\|-\|x_n-p\|\right)
\end{aligned}
$$

so that

$$
\alpha\sum_{n=1}^{\infty}\lambda_n\|x_n-p\| \le \|x_0-p\| < \infty.
$$

Since the sequence $\|x_n-p\|$ is decreasing this implies $\lim_{n\to\infty}\|x_n-p\|=0$. ■

Solution set with nonempty interior.

Proposition 60 *Let A be maximal monotone with int $\mathcal{S}\neq\emptyset$. Then x_n converges strongly as $n\to\infty$.*

Proof. If $B(p,r)\subset\mathcal{S}$ inequality 28 gives $r\|x_{k-1}-x_k\|\le\langle x_{k-1}-x_k, x_k-p\rangle$ and so

$$
\begin{aligned}
r\|x_{k-1}-x_k\| &\le \langle x_{k-1}-p, x_k-p\rangle - \|x_k-p\|^2 \\
&\le \|x_0-p\|\Big(\|x_{k-1}-p\|-\|x_k-p\|\Big)
\end{aligned}
$$

by Corollary 13. Hence

$$
\begin{aligned}
r\ \|x_n-x_m\| &\le r\sum_{k=n+1}^{m}\|x_{k-1}-x_k\| \\
&\le \|x_0-p\|\Big(\|x_n-p\| - \|x_m-p\|\Big).
\end{aligned}
$$

Since $\|x_n-p\|$ is convergent, x_n is a Cauchy sequence. ■

The NR convergence condition.
A fairly general result is the following, from [46, Nevanlinna and Reich]:

Theorem 61 *If A satisfies the NR convergence condition then x_n converges strongly as $n\to\infty$.*

Proof. Setting $j_n = x_n - P_{\mathcal{S}}x_n$ we have

$$
\begin{aligned}
\|j_n\|^2+\lambda_n\langle y_n, j_n\rangle &= \langle x_{n-1}-P_{\mathcal{S}}x_n, j_n\rangle \\
&= \langle j_{n-1}, j_n\rangle + \langle P_{\mathcal{S}}x_{n-1}-P_{\mathcal{S}}x_n, x_n-P_{\mathcal{S}}x_n\rangle \\
&\le \|j_{n-1}\|\ \|j_n\| \\
&\le \frac{1}{2}\left[\|j_{n-1}\|^2+\|j_n\|^2\right].
\end{aligned}
$$

Thus $\|j_n\|^2+2\lambda_n\langle y_n,j_n\rangle \le \|j_{n-1}\|^2$ and $\sum_{n=1}^{\infty}\lambda_n\langle y_n,j_n\rangle<\infty$. Since $\langle y_n,j_n\rangle\ge 0$ one must have $\liminf_{n\to\infty}\langle y_n,j_n\rangle=0$. The sequences $\{x_n\}$ and $\{y_n\}$ are bounded, and the convergence condition implies $\liminf_{n\to\infty}\|x_n-P_{\mathcal{S}}x_n\|=0$. Since $\|x_n-P_{\mathcal{S}}x_n\|$ is nonincreasing, it must converge to 0. On the other hand, the sequence $\|x_n-p\|$ is nonincreasing for each $p\in\mathcal{S}$. In particular, $\|x_{n+m}-P_{\mathcal{S}}x_n\|\le\|x_n-P_{\mathcal{S}}x_n\|$ and therefore $\|x_{n+m}-x_n\|\le 2\|x_n-P_{\mathcal{S}}x_n\|$. We conclude that x_n converges strongly to some $p\in\mathcal{S}$ as $n\to\infty$. ■

Compactness.

Proposition 62 *If $\omega[x_n] \cap \mathcal{S} \neq \emptyset$ then x_n converges strongly to some $p \in \mathcal{S}$.*

Proof. If $p \in \omega[x_n] \cap \mathcal{S}$ then $\|x_n - p\|$ is decreasing and $\liminf_{n\to\infty} \|x_n - p\| = 0$. ■

Symmetry.
For even functions we have the following result from [20, Brézis and Lions]:

Proposition 63 *If A is the subdifferential of an even function in $f \in \Gamma_0(H)$ then x_n converges strongly as $n \to \infty$.*

Proof. Recall that $2\lambda_n(f(u) - f(x_n)) \geq \|u - x_n\|^2 - \|u - x_{n-1}\|^2$. Let $m \geq n$ and take $u = -x_m$. Since $n \mapsto f(x_n)$ is decreasing we have $\|x_m + x_n\| \leq \|x_m + x_{n-1}\|$ and the function $n \mapsto \|x_m + x_n\|$ is decreasing. In particular $\|x_m + x_m\| \leq \|x_m + x_n\|$, thus $4\|x_m\|^2 \leq \|x_m + x_n\|^2$. We have $2\|x_n\|^2 + 2\|x_m\|^2 = \|x_m + x_n\|^2 + \|x_m - x_n\|^2 \geq 4\|x_m\|^2 + \|x_m - x_n\|^2$, so that $\|x_m - x_n\|^2 \leq 2\|x_n\|^2 - 2\|x_m\|^2$. Since $\|x_n\|$ converges as $n \to \infty$ this proves that x_n is a Cauchy sequence. ■

As before, asymptotic regularity is required for a general A:

Proposition 64 *Let A be odd. If $\{x_n\}$ is asymptotically regular then x_n converges strongly to some $p \in \mathcal{S}$ as $n \to \infty$.*

Proof. First, one easily verifies that $0 \in \mathcal{S}$ and that the sequence $\|x_{n+k} + x_n\|$ is decreasing for each $k \in \mathbf{N}$. Finally one concludes as in the proof of Proposition 58. ■

Comments
Without asymptotic regularity on can still prove strong convergence of the averages (see [40, Lions]) for odd operators. This was first proved in [9, Baillon] in the case $\lambda_n \equiv \lambda$. □

7.3 Euler sequences

Strong monotonicity.

Proposition 65 *Let A be α-strongly monotone. If $\sum \|z_n - z_{n-1}\|^2 < \infty$ then z_n converges strongly to the unique $p \in \mathcal{S}$ as $n \to \infty$.*

Proof. Strong monotonicity implies

$$2\alpha\lambda_n\|z_n - p\|^2 + \|z_{n+1} - p\|^2 \leq \|z_n - p\|^2 + \lambda_n^2\|w_n\|^2.$$

Therefore

$$2\alpha \sum_{n=1}^{\infty} \lambda_n \|z_n - p\|^2 \leq \|z_0 - p\|^2 + \sum \lambda_n^2 \|w_n\|^2 < \infty.$$

This implies $\liminf_{n\to\infty} \|x_n - p\| = 0$. But $\|x_n - p\|$ converges by Corollary 19. ■

Solution set with nonempty interior.

Proposition 66 *Assume int $\mathcal{S} \neq \emptyset$. If $\sum \|z_n - z_{n-1}\|^2 < \infty$ then z_n converges strongly as $n \to \infty$.*

Proof. If $B(p,r) \subset \mathcal{S}$ inequalities (28) and (18) together give

$$2r\lambda_n\|w_n\| + \|z_{n+1} - p\|^2 \leq \|z_n - p\|^2 + \lambda_n^2\|w_n\|^2.$$

This implies the sequence $\lambda_n\|w_n\| = \|z_{n+1} - z_n\|$ is in ℓ^1 and so z_n converges. ■

The NR convergence condition.

Theorem 67 *Assume $\sum \|z_{n+1} - z_n\|^2 < \infty$ and w_n is bounded. If A satisfies the NR convergence condition then $\{z_n\}$ converges strongly as $n \to \infty$.*

Proof. To simplify notation write $j_n = z_n - P_S z_n$. We have

$$\|j_{n+1}\|^2 \leq \|z_{n+1} - P_{\mathcal{S}} z_n\|^2 = \|j_n + \lambda_n w_n\|^2 = \|j_n\|^2 - 2\lambda_n\langle -w_n, j_n\rangle + \lambda_n^2\|w_n\|^2.$$

By hypothesis and Corollary 19 the sequence $[z_n, -w_n]$ is bounded. Moreover,

$$\sum_{n=1}^{\infty} \lambda_n\langle -w_n, j_n\rangle < \infty.$$

But $\langle -w_n, j_n\rangle \geq 0$ and so $\liminf_{n\to\infty}\langle w_n, j_n\rangle = 0$ and the convergence condition implies $\liminf_{n\to\infty}\|j_n\| = 0$. This sequence being convergent we have $\lim_{n\to\infty} j_n = 0$. Finally, $\|z_{n+m} - z_n\| \leq 2\|j_n\|$ and so z_n converges as $n \to \infty$. ■

Comments
The previous result holds if X and X^* are uniformly convex (see [46, Nevanlinna and Reich]).

According to [26, Bruck and Reich], the convergence condition can be replaced by int $\mathcal{S} \neq \emptyset$. In that case, if X is not uniformly convex it suffices that Ax be proximinal and convex for each x. On the other hand, according to [46, Nevanlinna and Reich], the conclusion of Theorem 67 is still true, even if X and X^* are not uniformly convex, provided $\mathcal{S}$ is proximinal and A is *accretive in the sense of Browder.* □

Compactness.

Proposition 68 *Assume that $\sum \|z_{n+1} - z_n\|^2 < \infty$ and $\omega[z_n] \cap \mathcal{S} \neq \emptyset$. Then z_n converges strongly to some $p \in \mathcal{S}$.*

Proof. The argument is the same as in Proposition 56 by virtue of Corollary 19. ■

Symmetry.

The following results uses the same ideas as in Propositions 58 and 64 but is apparently new:

Proposition 69 *Let T be non-expansive, $A = I - T$ and $\lambda_n \equiv 1$ so that $z_n = T^n z_0$. If T is odd and $\sum \|z_{n+1} - z_n\|^2 < \infty$ then $\{z_n\}$ is strongly convergent.*

Proof. Since T is odd one easily deduces that the sequence $\|z_{n+k}+z_n\|$ is decreasing for each k. From the fact that $\sum \|z_{n+1}-z_n\|^2 < \infty$ we can draw two conclusions: In the first place, Corollary 19 implies $d = \lim\limits_{n\to\infty} \|z_n\|$ exists because $0 \in \mathcal{S}$. On the other hand, the sequence z_n is asymptotically regular, so $\lim\limits_{n\to\infty} \|z_n - z_{n+k}\|$ exists for each k. As a consequence, $2d \leq \|z_{n+k}+z_n\|$ for each n and k. One concludes as in the proof of Proposition 58. ■

Comments
Without any further assumptions, $T^n z$ converges strongly in average if T (see [9, Baillon]). □

8 Asymptotic equivalence

In this section we explain how to deduce qualitative information on the asymptotic behavior of the systems defined by (5), (7) and (13). We provide a comparison tool that guarantees that two evolution systems share certain asymptotic properties. For the complete abstract theory see [3, Alvarez and Peypouquet].

8.1 Evolution systems

Let C be a convex subset of a Banach space X and let I denote the identity operator in X. An *evolution system (ES)* on C is a family $\{V(t,s) \ : \ t \geq s \geq 0 \ \}$ of maps from C into itself satisfying:

i) $V(t,t) = I$; and

ii) $V(t,s)V(s,r) = V(t,r)$.

Let $L > 0$. An evolution system is *L-Lipschitz* if it satisfies

iii) $\|V(t,s)x - V(t,s)y\| \leq L\|x-y\|$

and is *contracting (CES)* if it is 1-Lipschitz.

Example 4 Let F be a (possibly multivalued) function from $[t_0,\infty)\times C$ to C. Suppose that for every $s \geq t_0$ and $x \in C$ the differential inclusion $u'(t) \in F(t,u(t))$, with initial condition $u(s) = x$, has a unique solution $u_{s,x} : [s,\infty) \mapsto C$. The family U defined by $U(t,s)x = u_{s,x}(t)$ is an evolution system on C. If X is Hilbert space and $F(t,x) = -A_t x$, where $\{A_t\}$ is a family of maximal monotone operators, then the corresponding U is a *CES*. □

Example 5 Take a strictly increasing unbounded sequence $\{\sigma_n\}$ of positive numbers and set $\nu(t) = \max\{n \in \mathbf{N} \mid \sigma_n \leq t\}$. Consider a family $\{F_n\}$ of functions from C into C and define $U(t,s) = \prod_{n=\nu(s)+1}^{\nu(t)} F_n$, the product representing composition of functions. Then U is an ES. If each F_n is M_n-Lipschitz and the product $\prod_{n=1}^{\infty} M_n$ is bounded from above by M, then U is an *M-LES*. For instance, if $F_n = (I + A_n)^{-1}$, where $\{A_n\}$ is a family of m-accretive operators on C, then the piecewise constant interpolation of infinite products of resolvents defines a *CES*. □

8.2 Almost-orbits and asymptotic equivalence

Let V be an evolution system on C. A locally bounded trajectory of the form $t \mapsto V(t,s)x$ for s and x fixed is an *orbit* of V. A locally bounded function $u : \mathbf{R}_+ \to C$ is an *almost-orbit* of V if

$$\lim_{t\to\infty} \|u(t+h) - V(t+h,t)u(t)\| = 0 \qquad \text{uniformly in} \quad h \geq 0. \tag{30}$$

Orbits and almost-orbits have, essentially, the same asymptotic behavior.
Note the relation and difference with the notion of asymptotic pseudotrajectories where the convergence is uniform on compact time intervals ([15, Benaim and Hirsch], [16, Benaim, Hofbauer and Sorin]). The current concept is more demanding but will allow for more precise results (convergence rather than properties on the set of limit points).

Theorem 70 *Let V be an evolution system. For the weak topology assume either that V is Lipschitz or X is weakly complete (weak Cauchy nets are weakly convergent[4]). If every orbit of V converges weakly (resp. strongly), then so does every almost-orbit.*

Proof. For the strong topology, let u be an almost-orbit of V and let $\varepsilon > 0$. By definition, there is $S > 0$ such that

$$\|u(t+h) - V(t+h,t)u(t)\| < \varepsilon/4$$

for all $h \geq 0$ and $t \geq S$. Define $\zeta(S) = \lim_{t\to\infty} V(t,S)u(S)$ and choose $T > S$ such that $\|V(t,S)u(S) - \zeta(S)\| < \varepsilon/4$ for all $t \geq T$. Then

$$\|u(t+h) - \zeta(S)\| \leq \|u(t+h) - V(t+h,S)u(S)\| + \|V(t+h,S)u(S) - \zeta(S)\| < \varepsilon/2$$

for all $t \geq T$ and all $h \geq 0$. Thus $\|u(t') - u(t)\| < \varepsilon$ for all $t, t' \geq T$ so that $u(t)$ is Cauchy and converges.
It is clear that this argument is valid for the weak topology if X is weakly complete. If it is not the case but V is L-Lipschitz, one defines $\zeta(s) = w - \lim_{t\to\infty} V(t,s)u(s)$ and verifies that

$$\sup_{p\geq 0} \|\zeta(s+p) - \zeta(s)\| \leq L \sup_{p\geq 0} \|u(s+p) - V(s+p,s)u(s)\|,$$

which tends to zero as $s \to \infty$ showing that $\zeta(s)$ converges strongly to some ζ. Then one easily proves that $u(t)$ converges weakly to ζ as $t \to \infty$. ∎

A special case of Theorem 70 was proved in [49, Passty], when V is defined by a semigroup of contractions or if the almost-orbits are orbits of a semigroup of contractions.

Theorem 71 *Under the hypotheses of Theorem 70, the conclusion remains valid if the word* converges *is replaced by* converges in average.

The proof of this result can be found in [3, Alvarez and Peypouquet].

Comments
A similar result holds for *almost-convergence* (see [3, Alvarez and Peypouquet]), a concept developed in [41, Lorentz] that is stronger than convergence in average. It had been proved in [44, Miyadera and Kobayasi] under supplementary assumptions: i) V is defined by a strongly continuous semigroup of contractions; ii) $\mathcal{S} \neq \emptyset$; and iii) for the weak topology, X is weakly complete. □

[4]The spaces ℓ^1 and L^1, as well as all reflexive Banach spaces, have this property. It is not the case if X contains c_0, though (see p. 88 in [39, Li and Queffélec]).

Lemma 72 *Let U and V be evolution systems and assume that for each $r > 0$*

$$\lim_{t\to\infty} \sup_{h\geq 0} \sup_{\|z\|\leq r} \|U(t+h,t)z - V(t+h,t)z\| = 0$$

then every bounded orbit of V is an almost-orbit of U and viceversa.

Proof. Let v be an orbit of V such that $\|v(t)\| \leq r$ for all t. Then

$$\begin{aligned} \|v(t+h) - U(t+h,t)v(t)\| &= \|V(t+h,t)v(t) - U(t+h,t)v(t)\| \\ &\leq \sup_{\|z\|\leq r} \|U(t+h,t)z - V(t+h,t)z\| \end{aligned}$$

and so v is an almost-orbit of U. ■

8.3 Continuous dynamics and discretizations

The following results explain why in most cases the systems defined in the preceding sections converge under the same hypotheses. The proofs are considerably simplified if one assumes boundedness of the almost-orbits by virtue of Lemma 72. We shall give them in this case along with the references for more general settings. The following proposition gathers results from [37, Sugimoto and Koizumi] and [31, Güler].

Proposition 73 *Let A be a maximal monotone operator on H and let U and V be the evolution systems defined by the differential inclusion* (5) *and the proximal algorithm* (7)*, respectively. Assume one of the following conditions holds:*

i) $\{\lambda_n\} \in \ell^2 \setminus \ell^1$*; or*

ii) $A = \partial f$ *and* $\{\lambda_n\} \notin \ell^1$.

Then every orbit of U is an almost-orbit of V and viceversa.

Proof. Define $\nu(t)$ as in Example 5. If $\{\lambda_n\} \in \ell^2 \setminus \ell^1$, part *i)* in Corollary 17 gives

$$\|U(t+s,t)z - V(t+s,t)z\|^2 \leq 3\|A^0 z\|^2 \sum_{n=\nu(t)}^{\infty} \lambda_n^2$$

and we conclude using Lemma 72. For unbounded almost-orbits, see [37, Sugimoto and Koizumi]. If $A = \partial f$ and $\{\lambda_n\} \notin \ell^1$ the proof is highly technical and can be found in [31, Güler]. It also relies on part *i)* in Corollary 17 but sharper estimations on $\|A^0 x_n\|$ and $\|A^0 u(t)\|$ are needed. ■

Proposition 74 *Let T be nonexpansive, set $A = I - T$ and let U and W be the evolution systems defined by the differential inclusion* (5) *and Euler's discretization* (13)*, respectively. Assume $\{\lambda_n\} \in \ell^2 \setminus \ell^1$. Then every orbit of U is an almost-orbit of W and viceversa.*

Proof. The argument in the proof of part *i)* in Proposition 73 can be applied here as well, by virtue of inequality (14). ■

These properties allow for a better understanding of similar asymptotic behavior of the continuous and discrete processes: in general for weak convergence in average (Section 4), for weak convergence in the case of demi-positive operators (Section 5) and for strong convergence under addtitional geometrical hypotheses (Section 6).

8.4 Quasi-autonomous systems

One of the advantage of this approach through almost-orbits is that it extends to non-autonomous systems.

8.4.1 Continuous dynamics

Recall that the solutions of the differential inclusion (5) define an evolution system U as in Example 4. Let us consider quasi-autonomous versions of (5), namely

$$-\dot{v}(t) \in Av(t) + \varphi(t) \tag{31}$$

and

$$-\dot{v}(t) \in Av(t) + \varepsilon(t)v(t). \tag{32}$$

Proposition 75 *If $\varphi \in L^1(0,\infty;X)$, then every function satisfying* (31) *is an almost-orbit of U. The same holds for every function satisfying* (32) *provided $\varepsilon \in L^1(0,\infty;\mathbf{R})$.*

Proof. For the first part we follow [44, Miyadera and Kobayasi]. If v satisfies (31) and $t \geq 0$ we have

$$\|v(t+s) - U(t+s,t)v(t)\|^2 \leq 2\int_0^s \|\varphi(t+\tau)\|\|v(t+\tau) - U(t+\tau,t)v(t)\|d\tau$$

and so

$$\|v(t+s) - U(t+s,t)v(t)\| \leq \int_0^s \|\varphi(t+\tau)\|d\tau \leq \int_t^\infty \|\varphi(\tau)\|d\tau.$$

On the other hand, let v satisfy (32). Fix t and consider as above $\psi(s) = \frac{1}{2}\|U(t+s,t)v(t) - v(t+s)\|^2$. Using $\langle \zeta, \zeta - \xi\rangle \geq -\frac{1}{4}\|\xi\|^2$ for all $\zeta, \xi \in H$, we deduce $\dot{\psi}(s) \leq \frac{1}{4}|\varepsilon(t+s)|\|U(t+s,t)v(t)\|^2$ for almost every $s > 0$. Integrating from 0 to s and observing that $\psi(0) = 0$ we obtain

$$\|U(t+s,t)v(t) - v(t+s)\|^2 \leq \frac{1}{4}\int_t^{t+s} |\varepsilon(\tau)|\ \|U(t+\tau,t)v(t)\|^2\ d\tau \leq \frac{M}{4}\int_t^\infty |\varepsilon(\tau)|\ d\tau$$

if v is bounded. ■

Comments
In [1, Alvarez], the author studies the problem

$$u''(t) + \gamma u'(t) + \nabla\Phi(u(t)) = 0, \tag{33}$$

where Φ is a $\mathcal{C}^1$ convex function. He proves that if $\mathrm{Argmin}(\Phi) \neq \emptyset$, then each solution $u(t)$ converges weakly to a minimizer of Φ as $t \to \infty$ and gives conditions for strong convergence. Later, in [6, Attouch and Czarnecki] the authors establish, among other results, that if $\varepsilon \in L^1$ the solutions of

$$u''(t) + \gamma u'(t) + \nabla\Phi(u(t)) + \varepsilon(t)u(t) = 0. \tag{34}$$

also converge weakly to minimizers of Φ. It turns out (see [4, Alvarez and Peypouquet]) that under this condition ($\varepsilon \in L^1$) the solutions of (34) are almost-orbits of the evolution system defined by (33).
This is an alternative way to prove the cited result from [6, Attouch and Czarnecki] and it shows that these tools building on almost-orbits to classify the asymptotic behavior through equivalence classes (continuous trajectories, proximal or Euler approximations, Tykhonov regularization, perturbations) can be applied to second-order systems as well. □

8.4.2 Proximal sequences

In a similar fashion one can prove any interpolation of a sequence $\{y_n\}$ satisfying

$$y_{n-1} - y_n \in \lambda_n A y_n + \phi_n \tag{35}$$

or

$$y_{n-1} - y_n \in \lambda_n A y_n + \epsilon_n y_n \tag{36}$$

is an almost-orbit of the evolution system U defined by the proximal scheme (7) as in Example 5 provided $\{\phi_n\} \in \ell^1(\mathbf{N}; X)$ and $\{\epsilon_n\} \in \ell^1(\mathbf{N}; \mathbf{R}_+)$, respectively.

For additional applications and examples see [4, Alvarez and Peypouquet].

9 Concluding remarks

It is useful to observe that there are two aspects related to the ideas of asymptotic equivalence discussed in the last section. In the first place, one can obtain sufficient conditions for a perturbed, regularized or discretized system to have the same asymptotic properties as the original one. The issue here is in terms of stability or regularity or computational purposes. On the other hand, if a given dynamics does not have some desirable asymptotic behavior, one can introduce pertubation in order to generate orbits having better properties. In this case, the tools of asymptotic equivalence give necessary condition for a perturbation to be effective.

Observe that the trajectories defined by (5) only converge weakly in average. Even in the case where $A = \partial f$, convergence is still weak and the limit depends on the initial point. One can get a better asymptotic behavior by forcing the system to stabilize in the direction of the origin. More precisely, consider a piecewise absolutely continuous function $\varepsilon : \mathbf{R}_+ \to \mathbf{R}_+$ such that $\lim_{t\to\infty} \varepsilon(t) = 0$. If $\varepsilon \in L^1(0,\infty; \mathbf{R}_+)$ the system defined by (32) will have the same asymptotic behavior as (5) by Proposition 75. If we expect the regularized system to have better properties we must consider $\varepsilon \notin L^1(0,\infty; \mathbf{R}_+)$. The following result is from [27, Cominetti, Peypouquet and Sorin]:

Proposition 76 *Suppose $v : \mathbf{R}_+ \to H$ satisfies*

$$-\dot{v}(t) \in Av(t) + \varepsilon(t)v(t).$$

with $\varepsilon \notin L^1(0,\infty; \mathbf{R}_+)$. Assume further that $A = \partial f$ or $\int_0^\infty |\dot{\varepsilon}(t)|\, dt < \infty$ (finite total variation). Then $\lim_{t\to\infty} v(t) = P_{\mathcal{S}}0$.

Special cases of the preceding result had been proved earlier in [21, Browder], [54, Reich] and [5, Attouch and Cominetti]. A similar result for the second order appears in [6, Attouch and Czarnecki].

Also, as we mentioned before, the trajectories defined by (5) need not be weakly convergent. If one applies the proximal point algorithm with stepsizes $\lambda_n \in \ell^2$, by Proposition 73, the corresponding system will have the same asymptotic properties. In other words, the approximation is too good: "the discrete approximation mirrors the behavior of the differential equation *too* well" [25, Bruck, p. 29]. If one wishes to get a better (or different) behavior, it is necessary to consider $\lambda_n \notin \ell^2$. This turns out to be fruitful because, in that case Theorem 48 guarantees weak convergence even when the operator is not demipositive (see also Example 3 in Section 6).

Acknowledgments
The authors wants to thank Roberto Cominetti for his help during the preparation of this work. In particular his (unpublished) notes on "Evolution equations and monotone maps" were very helpful.
S. Sorin acknowledges support from grant ANR-08-BLAN-0294-01 (France).

References

[1] Alvarez F, *On the minimizing property of a second order dissipative system in Hilbert spaces*, SIAM J. Control Optim., **38** (2000), 1102-1119.

[2] Alvarez F, Peypouquet J, Asymptotic equivalence and Kobayashi-type estimates for nonautonomous monotone operators in Banach spaces, to appear in *DCDS*, (2009).

[3] Alvarez F, Peypouquet J, Asymptotic almost-equivalence of abstract evolution systems, *submitted* (2008).

[4] Alvarez F, Peypouquet J, Asymptotic almost-equivalence and applications, *in preparation* (2009).

[5] Attouch H, Cominetti R, A dynamical approach to convex minimization coupling approximation with the steepest descent method, *J. Diff. Equations*, **128** (1996), 519-540.

[6] Attouch H, Czarnecki MO, *Asymptotic control and stabilization of nonlinear oscillators with non-isolated equilibria*, J. Diff. Equations, **179** (2002), 278-310.

[7] Baillon JB, Un théorème de type ergodique pour les contractions non linéaires dans un espace de Hilbert, *CRAS*, **280** (1975), 1511-1514.

[8] Baillon JB, Quelques propriétés de convergence asymptotique pour les semi-groupes de contractions impaires, *CRAS*, **283** (1976), A75-A78.

[9] Baillon JB, Quelques propriétés de convergence asymptotique pour les contractions impaires, *CRAS*, **283** (1976), A587-A590.

[10] Baillon JB, Un exemple concernant le comportement asymptotique de la solution du problème $du/dt + \partial\varphi(u) \ni 0$, *J. Funct. Anal.*, **28** (1978), 369-376.

[11] Baillon JB, "Comportement asymptotique des contractions et semi-groupes de contraction", Thèse, Université Paris 6, 1978.

[12] Baillon JB, Brézis H, Une remarque sur le comportement asymptotique des semi-groupes non linéaires, *Houston J. Math.*, **2** (1976), 5-7.

[13] Barbu V, "Nonlinear semigroups and differential equations in Banach spaces". Noordhoff, Leyden, 1976.

[14] Bauschke HH, Burke JV, Deutsch FR, Hundal HS, Vanderwerff JD, A new proximal point iteration that converges weakly but not in norm, *Proc. Amer. Math. Soc.*, **133** (2005), 1829-1835.

[15] Benaïm M, Hirsch MW, Asymptotic pseudotrajectories and chain recurrent flows, with applications, *J. Dynamical Differential Equations*, **8** (1996), 141-176.

[16] Benaïm M, Hofbauer J, Sorin S, Stochastic approximations and differential inclusions, *SIAM J. Control Optim.*, **44** (2005), 328-348.

[17] Bénilan P, "Équations d'évolution dans un espace de Banach quelconque et applications". Thèse, Orsay, 1972.

[18] Brézis H, "Monotonicity methods in Hilbert spaces and some applications to nonlinear partial differential equations". Contributions to Nonlinear Functional Analysis, ed. by H. Zarantonello, Academic Press, 1971, 101-156.

[19] Brézis H, "Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert". North Holland Publishing Company, Amsterdam, 1973.

[20] Brézis H, Lions PL, Produits infinis de résolvantes, *Israel J. Math.*, **29** (1978), 329-345.

[21] Browder FE, Nonlinear operators and nonlinear equations of evolution in Banach spaces, *Proc. Symp. Pure Math*, **18** (2) (1976), Amer. Math. Soc., Providence RI.

[22] Bruck RE, Asymptotic convergence of nonlinear contraction semigroups in Hilbert space, *J. Funct. Anal.*, **18** (1975), 15-26.

[23] Bruck RE, On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in Hilbert space, *J. Math. Anal. Appl.*, **61** (1977), 159-164.

[24] Bruck RE, On the almost-convergence of iterates of nonexpansive mappings in a Hilbert space and the structure of the ω-limit set, *Israel J. Math.*, **29** (1978), 1-17.

[25] Bruck RE, Asymptotic behavior of nonexpansive mappings, *Proc. Symp. Pure Math*, Nonlinear Functional Analysis and Applications, **45** (1986), 1-47, Amer. Math. Soc., Providence RI.

[26] Bruck RE, Reich S, A general convergence principle in nonlinear functional analysis, *Nonlinear Anal.*, **4** (1980), 939-950.

[27] Cominetti R, Peypouquet J, Sorin S, Strong asymptotic convergence of evolution equations governed by maximal monotone operators with Tikhonov regularization, *J. Diff. Equations*, **245** (2008), 3753-3763.

[28] Crandall MG, Liggett TM, Generation of semigroups of nonlinear transformations on general Banach spaces, *Am. J. Math.*, **93** (1971), 265-298.

[29] Crandall MG, Pazy A, Semi-groups of nonlinear contractions and dissipative sets, *J Funct. Anal.*, **3** (1969), 376-418.

[30] Edelstein M, The construction of an asymptotic center with a fixed-point property, *Bull. Amer. Math. Soc.*, **78** (1972), 206-208.

[31] Güler O, On the convergence of the proximal point algorithm for convex minimization, *SIAM J. Control & Opt.*, **29** (1991), 403-419.

[32] Güler O, Convergence rate estimates for the gradient differential inclusion, *Optim. Methods Softw.*, **20** (2005), 729-735.

[33] Hirano N, Nonlinear ergodic theorems and weak convergence theorems, *J. Math. Soc. Japan*, **34** (1982), 35-46.

[34] Hirsch F, "Familles résolvantes, générateurs, cogénérateurs, potentiels". Thèse, Orsay, 1971.

[35] Kato T, Nonlinear semi-groups and evolution equations, *J. Math. Soc. Japan*, **19** (1967), 508-520.

[36] Kobayashi Y, Difference approximation of Cauchy problems for quasi-dissipative operators and generation of nonlinear semigroups, *J. Math Soc. Japan*, **27** (1975), 640-665.

[37] Sugimoto T, Koizumi M, On the asymptotic behavior of a nonlinear contraction semigroup and the resolvent iteration, *Proc. Japan Acad.*, **59** (1983), 238-240.

[38] Komura Y, Nonlinear semi-groups in Hilbert space, *J. Math. Soc. Japan*, **19** (1967), 493-507.

[39] Li D, Queffélec, H "Introduction à l'étude des espaces de Banach - Analyse et probabilités". Cours specialisés **12**, Société Mathématique de France, 2004.

[40] Lions PL, Une méthode itérative de résolution d'une inéquation variationnelle, *Israel J. Math.*, **31** (1978), 204-208.

[41] Lorentz GG, A contribution to the theory of divergent sequences, *Acta Math.*, **80** (1948), 167-190.

[42] Martinet B, Régularisation d'inéquations variationnelles par approximations successives, *Rev. Française Informat. Recherche Opérationnelle* , **4** (1970), 154-158.

[43] Minty G, Monotone (nonlinear) operators in Hilbert space, *Duke Math. J.*, **29** (1962), 341-346.
On the monotonicity of the gradient of a convex function, *Pacific J. Math.*, **14** (1964), 243-247.

[44] Miyadera I, Kobayasi K, On the asymptotic behavior of almost-orbits of nonlinear contractions in Banach spaces, *Nonlinear Anal.*, **6** (1982), 349-365.

[45] Moreau JJ, Proprietés des applications "prox", *CRAS*, **256** (1963), 1069-1071.

[46] Nevanlinna 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.

[47] Opial Z, Weak Convergence of the sequence of successive approximations for nonexpansive mappings, *Bull. Amer. Math. Soc.*, **73** (1967), 591-597.

[48] Passty G, Ergodic convergence to a zero of a sum of monotone operators in Hilbert space, *J. Math. Anal. Appl.*, **72** (1979), 383-390.

[49] Passty G, Preservation of the asymptotic behavior of a nonlinear contraction semigroup by backward differencing, *Houston J. Math.*, **7** (1981), 103-110.

[50] Pazy A, "Semigroups of nonlinear contractions and their asymptotic behavior". Nonlinear Analysis and Mechanics: Heriot-Watt Symposium, vol III, 1979, Pitman, 36-134.

[51] Pazy, A. On the asymptotic behavior of iterates of nonexpansive mappings in Hilbert space, *Israel Journal of Mathematics*, **26** (1977), 197-20.

[52] Pazy, A. On the asymptotic behavior of semigroups of nonlinear contractions in Hilbert space, *Journal of Functional Analysis*, **27** (1978), 292-307.

[53] Pazy, A. Remarks on nonlinear ergodic theory in Hilbert spaces, *Nonlinear Analysis, Theory, Methods & Applications*, **3** (1979), 863-871.

[54] Reich S, Nonlinear evolution equations and nonlinear ergodic theorems, *Nonlinear Anal.*, **1** (1977), 319-330.

[55] Reich S, Nonlinear ergodic theory in Banach spaces, *Argonne National Lab.* ANL-79-76 (1979).

[56] Reich S, Weak convergence theorems for nonexpansive mappings in Banach spaces, *J. Math. Anal. & App.*, **67** (1979), 274-276.

[57] Rockafellar RT, Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.*, **14** (1976), 877-898.

[58] Shepilov MA, "On gradient and penalty methods in mathematical programming problems". PhD thesis, Moscow, 1974.

[59] Shor NZ, "Minimization methods for non-differentiable functions". Springer, Berlin, 1985.

[60] Vigeral G, Evolution equations in discrete and continuous time for nonexpansive operators in Banach spaces, to appear in *COCV*, (2009).