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

ANÁLISIS ASINTÓTICO DE SISTEMAS DE EVOLUCIÓN Y APLICACIONES EN OPTIMIZACIÓN

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

JUAN PEYPOUQUET URBANEJA

PROFESORES GUÍA:
FELIPE ÁLVAREZ DAZIANO
SYLVAIN SORIN.

MIEMBROS DE LA COMISIÓN:
HÉDY ATTOUCH.
MICHEL BENAÏM.
JEAN-BERNARD BAILLON.
ROBERTO COMINETTI COTTI-COMETTI
PASCAL FREY.
VÍCTOR AYALA BRAVO

SANTIAGO DE CHILE
DICIEMBRE 2007

Resumen

En esta tesis doctoral se estudian diversas propiedades asintóticas de algunos sistemas de evolución. La motivación y la mayor parte de las aplicaciones vienen de la optimización, pero las técnicas desarrolladas en este trabajo pueden ser emplearse, por ejemplo, teoría de puntos fijos, análisis numérico, teoría de juegos y sistemas dinámicos en un sentido amplio.

La primera parte es una monografía titulada *Ecuaciones de evolución: discretización, perturbación y análisis asintótico*. Se trata de una recopilación autocontenida pero bastante concisa sobre inclusiones diferenciales definidas por operadores acretivos en espacios de Banach y algunas de sus discretizaciones. Se demuestran los resultados clásicos de existencia y comportamiento asintótico de las soluciones, cubriendo una parte importante de la inmensa literatura sobre el tema.

La segunda parte se titula *Análisis asintótico de sistemas de evolución no autónomos* y contiene los resultados originales de esta tesis. Se estudia el comportamiento global y asintótico de algunos sistemas dinámicos con posible dependencia en el tiempo. Se destacan las siguientes líneas de trabajo:

1. El análisis de la convergencia en valor para un esquema prox-diagonal bajo hipótesis mínimas.

2. Estimaciones globales para discretizaciones de ciertas inclusiones diferenciales no autónomas. Esto además premite deducir propiedades de continuidad de sus trajectorias.

3. Desarrollo de un criterio que vincula las propiedades asintóticas de dos sistemas de evolución abstractos. Así, se logran comparar sistemas de distinta índole (algoritmos, inclusiones diferenciales, etc.).

Acknowledgements

I would like to thank all those people and institutions that helped this project come to an end.

My first words are for my advisors: *Felipe Alvarez* from Universidad de Chile (UCH) −an outstanding teacher, thesis director and human being−, who drove my attention to optimization, then nonlinear and asymptotic analysis and guided my first steps into research; and *Sylvain Sorin* from Université Pierre et Marie Curie (Paris VI), interacting with him was a remarkable experience. It would be an honor and a pleasure to continue working in collaboration with them.

Roberto Cominetti and *Jerôme Bolte* also had a major influence in my work through enlightening discussions.

I also thank *Hédy Attouch* and *Michel Benaïm* for their reports on the manuscript; as well as *Jean-Bernard Baillon*, *Víctor Ayala* and *Pascal Frey* for attending the presentation and giving helpful comments.

The development of this project would not have been possible without the financial support of the Chilean and French governments through *MECESUP grant UCH0009*, the *ECOS* project, the *French Embassy in Santiago* and the *CROUS de Paris*. The *Center for Mathematical Modelling* and the *Instituto Milenio* project supported my attendance to several schools, workshops and conferences in Chile and abroad, which gave me the opportunity to present my work to a broader community and receive useful feedback. I also thank the Department of Mathematical Engineering at the UCH, the Equipe Combinatoire et Optimisation at Paris VI and their staff for their valuable help.

Finally, I wish to thank all the people who contributed to the human environment in which I carried out my work: *Carlos Quintero* and *Eric Arnold* for making my stay in Paris a thrilling experience; my friends at UCH and Paris VI *Loredana Smaranda*, *Marcelo Sobottka*, *Daniel Coronel*, *Adriana Piazza*, *Flavio Guíñez*, *Miguel Carrasco*, *Alberto Mercado* and *Guillaume Vigeral* for being there; and *Alexis Briceño* for infinite patience and unconditional support.

Dedicated to the memory of my grandparents

Contents

Introduction

Foreword. This dissertation is divided in two parts. In the first part we provide a compendium of most classical results concerning the asymptotic behavior of some evolution systems that arise in optimization and fixed-point theory. More precisely, we discuss on differential inclusions governed by accretive operators in Banach spaces along with two well-known discretizations. The number of (now classical) works in this area is immense. Our research within the vast bibliography led us to the strong conviction that it would be useful to gather a large number of results, techniques and references in a short self-contained handbook. We provide and comment on existence results for the sake of completeness but the main emphasis is on the conditions ensuring convergence (in a wide sense). We also mention some pioneer works on nonautonomous systems, especially those having implications in optimization theory via perturbations, penalization, approximation, etc. This was done in collaboration with Dr. Sylvain Sorin at Pierre and Marie Curie University, in Paris. The second part contains new results concerning the asymptotic behavior of some nonautonomous evolution systems. We provide some abstract techniques for their analysis (in particular, the development of the theory of asymptotic almost-equivalence) and answers to selected mathematical problems that we considered both challenging and relevant. In what follows we present the main results of the second part. The first section contains new results on the convergence of a "diagonal" proximal point algorithm, which will also serve as a motivation for the theory of asymptotic almost-equivalence presented in the next section. Finally we describe our extension of a well-known inequality of Yoshikazu Kobayashi to the nonautonomous setting.

Approximation and diagonal PROX.

Let $\{f_n\}$ be a family of proper, lower-semicontinuous convex functions on a real Hilbert space H with a common (convex) domain D. Consider an objective function f with domain D_f. The functions f_n are meant to approximate f in some sense. Denote $f^* = \inf_H f = \inf_D f$.

Take a starting point $x_0 \in H$ and two sequences of real parameters $\{\lambda_k\} \subset (0, \Lambda]$ and $\{\varepsilon_k\} \subset [0, \infty)$. A sequence $\{x_k\} \subset D$ is said to be an *inexact diagonal proximal sequence* generated by $(x_0, \{\lambda_k\}, \{f_k\}, \{\varepsilon_k\})$ if

$$y_k := \frac{x_{k-1} - x_k}{\lambda_k} \in \partial_{\varepsilon_k} f_k(x_k)$$

for all $k \geq 1$, where the approximate $\varepsilon-$subdifferential ∂_ε is defined by

$$\partial_\varepsilon g(u) = \{ \ u^* \in H \mid g(v) \geq g(u) + \langle u^*, v - u\rangle - \varepsilon \ \ \forall v \in H \ \} \qquad \varepsilon \geq 0.$$

Observe that in order to construct such a sequence it is necessary to solve an optimization problem at each iteration. This is usually done by means of another algorithm (e.g. bundle methods) so the use of the approximate $\varepsilon-$subdifferential is very important for practical purposes.

If $f_n \equiv f$ and $\varepsilon_n \equiv 0$, this is the standard *proximal point algorithm*, which was first used in [85] and [81]. This algorithm has been used extensively in convex minimization (see [56] and [96]) in the search for the value f^* and, if there are any, the points at which it is attained. The term *diagonal* refers to the fact that the objective function is updated at each iteration.

Set $\sigma_0 = 0$ and for $n \geq 1$, $\sigma_n = \sum_{k=1}^n \lambda_k$. We assume $\lim_{n\to\infty} \sigma_n = \infty$. For a sequence $\{z_n\}$ denote by $\overline{z_n} = \frac{1}{\sigma_n} \sum_{k=1}^n \lambda_k z_k$ the sequence of its averages weighted by the stepsizes $\{\lambda_k\}$. Notice that $\lim_{n\to\infty} z_n = L$ implies $\lim_{n\to\infty} \overline{z_n} = L$. Finally, let $\{p_n\}$ and $\{\varrho_n\}$ be two sequences of positive numbers. We shall write $p_n = \mathcal{O}(\varrho_n)$ if the ratio p_n/ϱ_n is bounded. Observe that if $p_n = \mathcal{O}(\varrho_n)$ then $\overline{p_n} = \mathcal{O}(\overline{\varrho_n})$. If $\lim_{n\to\infty} p_n/\varrho_n = 0$ we write $p_n = o(\varrho_n)$. Set $f_n^* = \inf f_n$. Let $\mathcal{S} = \text{Argmin} f$. For $r > 0$ define $\mathcal{S}_r = \{\ u \in D \cap D_f \mid f(u) \leq f^* + r\ \}$ whenever $f^* > -\infty$ and $\mathcal{S}_r = \{\ u \in D \cap D_f \mid f(u) \leq -r\ \}$ otherwise.

Convergence of the averaged values. Consider the following:

Hypothesis $\mathbf{H_1}$: There is $r > 0$ such that $\limsup_{n\to\infty} \overline{f_n(u)} \leq f(u)$ for all u satisfying $f(u) \leq f^* + r$.

Theorem A *Under* $\mathbf{H_1}$, *if* $\lim_{n\to\infty} \overline{\varepsilon_n} = 0$, *we have:*

i) $\limsup_{n\to\infty} \overline{f_n(x_n)} \leq f^*$;

ii) If $f^* \leq \liminf_{n\to\infty} \overline{f_n^*}$, *then* $\lim_{n\to\infty} \overline{f_n(x_n)} = f^*$; *and*

iii) If $f^* \leq \liminf_{n\to\infty} f_n^*$, *then* $\liminf_{n\to\infty} f_n(x_n) = f^*$.

This result, similar to Theorem 2.1 in [13], but with different "degrees" in the conclusion, helps understand more clearly the effect of averaging. Notice that $\overline{f_n(x_n)}$ approximates the optimal value under fairly weak hypotheses on the approximation.

In the special case where there is $x^* \in \mathcal{S} \cap D$ we have

$$\overline{f_n(x_n)} - f^* \leq \frac{\|x^* - x_0\|^2}{2\sigma_n} + \left[\overline{f_n(x^*)} - f^*\right] + \overline{\varepsilon_n}.$$

Suppose for simplicity that $f^* \leq \overline{f_n^*}$ so that the left-hand side is nonnegative. The rate of convergence of $\overline{f_n(x_n)}$ to f^* depends on three parameters, namely:

- σ_n, which seems to be intrinsic to the proximal iterations;
- ε_n, which has to do with computational precision;
- the behavior of f_n on the minimizing set $\mathcal{S}$.

If $\sum \lambda_k \varepsilon_k < \infty$ and $\overline{f_n(x^*)} - f^* = \mathcal{O}(\frac{1}{\sigma_n})$ for some $x^* \in \mathcal{S}$, then $\overline{f_n(x_n)} - f^* = \mathcal{O}(\frac{1}{\sigma_n})$. The same speed of convergence is provided by Theorem 2.1 in [56] when $f_n \equiv f$ and $\varepsilon_n \equiv 0$.

Convergence in value and cluster points. In order for the values $\{f_n(x_n)\}$ to converge we need an additional assumption, namely:

Hypothesis $\mathbf{H_2}$: There exist a set $K \subseteq H$ containing the trajectory $\{x_n\}$ and a nonnegative sequence $\{a_n\}$ such that $f_n(x) \leq f_{n-1}(x) + a_{n-1}$ for all $n \geq 2$ and all $x \in K$.

This condition holds trivially if the sequence $\{f_n\}$ is decreasing or if it converges uniformly in some set. In general, since the estimation is somewhat uniform, it is difficult to establish in the whole domain without any further assumptions. However, the fact that the set K is chosen according to $\{x_n\}$ plays a key role in practice.

Convergence of the values $f_n(x_n)$ had been proved in a monotone setting (see Theorem 3.1 in [1] and Corollary 3.1 in [73]) or assuming that $\mathcal{S} \neq \emptyset$, that $f_n \to f$ in the sense of Mosco[1] and that $\{\lambda_n\}$ is bounded away from zero (see Theorems 3.2 and 3.3 in [1] and Theorem 2.2 in [13]).

Hypotheses $\mathbf{H_1}$ and $\mathbf{H_2}$ guarantee convergence in value. This is not a very restrictive setting, as it may seem at a first glance. One also gets information about the ω-limit set under epiconvergence:

Theorem B *Assume hypotheses* $\mathbf{H_1}$ *and* $\mathbf{H_2}$ *hold. If* $f^* \leq \liminf_{n\to\infty} f_n^*$ *and the sequences* $\{a_n\}$ *and* $\{\varepsilon_n\}$ *are in* ℓ^1 *then* $\lim_{n\to\infty} f_n(x_n) = f^*$*. Suppose, in addition, that for all* $x \in D_f$ *and for all sequences* $\{z_n\}$ *in* D *converging to* x *for the topology* τ *we have* $f(x) \leq \liminf_{n\to\infty} f_n(z_n)$*. Then every* $\tau-$*cluster point of the sequence* $\{x_n\}$ *is a minimizer of* f*.*

Rates of convergence. It is natural to expect the rate of convergence of $f_n(x_n)$ to depend on the way the sequence f_n converges to f. When the function f has minimizers we can give precise estimates on the rate of convergence of the values in terms of the rate of pointwise convergence of the sequence $\{f_n\}$ on the optimal set. The main drawback of this approach is that we require the set $\mathcal{S} \cap D$ to be nonempty. This is the case for exterior penalties and for the barrier $\theta(u) = -\sqrt{-u}$, but not for the log- or the inverse barriers. In what follows we assume that $\mathbf{H_1}$ and $\mathbf{H_2}$ hold and that $f^* \leq \liminf_{n\to\infty} f_n^*$. Denote by ϱ_n the order $\mathcal{O}(\varrho_n) = |\overline{f_n(x^*)} - f(x^*)|$. If $\sum_{k=1}^{n} \sigma_k(a_k + \varepsilon_k) < \infty$ then $|f_n(x_n) - f^*| = \mathcal{O}\left(\max\left\{\varrho_n, \frac{1}{\sigma_n}\right\}\right)$. In particular, if $|\overline{f_n(x^*)} - f^*| = \mathcal{O}(\frac{1}{\sigma_n})$, then $|f_n(x_n) - f^*| = \mathcal{O}(\frac{1}{\sigma_n})$.

The effect of strong convergence. Under certain conditions, this estimation can be improved if the sequence $\{x_n\}$ converges strongly to a minimizer of f. First assume the sequence $\{f_n\}$ is uniformly bounded from below by f^* and set $\zeta_n(u) = f_n(u) - f^*$ for $u \in D$.

Theorem C *Assume that* x_n *converges strongly to some* $x^* \in \mathcal{S} \cap D$ *such that* $\zeta_n(x^*) = o(\frac{1}{\sigma_n})$*. If* $\|y_n\| = \mathcal{O}(\frac{1}{\sigma_n})$ *and* $\varepsilon_n = o(\frac{1}{\sigma_n})$*, then* $|f_n(x_n) - f^*| = o(\frac{1}{\sigma_n})$*. In particular, if* $f_n \equiv f$*,* $\varepsilon_n \equiv 0$ *(cf [56]) and* x_n *converges strongly to some* $x^* \in \mathcal{S}$*, then* $f(x_n) - f^* = o(\frac{1}{\sigma_n})$*.*

In general, it is difficult to prove that $\|y_n\| = \mathcal{O}(\frac{1}{\sigma_n})$ because the speed depends strongly on the evolution of the sequence $\{f_n\}$ so one has to study each particular case.

The previous result reveals the importance of knowing *a priori* that the sequence is convergent. Keeping this in mind we propose a method for deriving some qualitative information on the asymptotic behavior of a dynamical system by studying a different one.

Deducing qualitative properties indirectly.

Let X be a Banach space. An *evolution scheme in discrete (resp. continuous) time* is a rule that determines a sequence (resp. trajectory) in X, which we call an *orbit*, departing from certain initial data (which may include starting time, point, velocity, etc.). Researchers from many areas of mathematics have been concerned with different aspects of these kinds of schemes, namely: existence, exact or approximate computation, regularity or long-term behavior of the orbits. Our work deals with some of the asymptotic properties that are common to systems that can be considered similar in some sense.

In the study of a particular evolution scheme, say S_1, it is sometimes useful to consider another scheme S_2 for several reasons: it is not possible to determine the orbits of S_1 exactly due to restricted computational precision; more powerful or more familiar techniques are

[1] Epi-convergence both for the weak and the strong topologies (see below).

available for studying S_2; scheme S_2 has already been studied or just looks simpler; one is a particular case of the other; etc.

Once accomplished the study of S_2, and if the answers are satisfactory, it remains to verify whether the results can be somehow translated back to S_1. We have developed a criterion that guarantees this translation. But first, let us show, through an example, the kinds of properties we wish to translate.

A model example. 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:

$$u'(t) = R(u(t)). \tag{1}$$

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 $t \to \infty$.

Let $\{\lambda_n\}$ be a sequence of positive numbers and denote $\sigma_n = \sum_{k=1}^{n} \lambda_k$. Among the possible discrete approximations of equation (1) we have:

<u>Euler's explicit discretization:</u> From a given initial point z_0 define $z_{n+1} = z_n + \lambda_n R(z_n)$ for $n \geq 0$. An advantage of this scheme is that the sequence is easily computed. Write $z_n = r_n(\cos\theta_n, \sin\theta_n)$. We have $r_{n+1}^2 = \prod_{k=1}^{n}(1 + \lambda_k^2)$ and $\theta_n = \theta_0 + \sum_{k=1}^{n} \arctan(\lambda_k)$. The sequence r_n is increasing. It stays bounded if, and only if, $\lambda_n \in \ell^2$. The argument θ_n is also increasing. It converges if $\lambda_n \in \ell^1$ and diverges otherwise.

<u>The proximal point algorithm:</u> Take x_0 and define $x_{n+1} = (I + \lambda_n R)^{-1} x_n$ for $n \geq 0$. Here $r_{n+1}^2 = \prod_{k=1}^{n}(1 + \lambda_k^2)^{-1}$ and $\theta_n = \theta_0 + \sum_{k=1}^{n} \arctan(\lambda_k)$. In this case the sequence r_n is decreasing. The limit is positive if, and only if, $\lambda_n \in \ell^2$. The argument θ_n is increasing as before. It converges if $\lambda_n \in \ell^1$ and diverges otherwise.

Let us remark that if $\{\lambda_n\} \in \ell^2 \setminus \ell^1$, both x_n and z_n are bounded but not convergent. It is also possible to prove that they converge to 0 in average. This is precisely the case for the continuous trajectory $u(t)$.

Roughly speaking, our main result states that if the evolution schemes S_1 and S_2 are *asymptotically almost-equivalent* (in a sense to be defined), then their orbits have the same asymptotic behavior in terms of boundedness, convergence, convergence in average, etc. In this example, if $\{\lambda_n\} \in \ell^2 \setminus \ell^1$ the discretizations described above can be proved to be asymptotically almost-equivalent to the continuous-time scheme. Definitions, main results and examples are commented below.

Evolution systems and almost-orbits. Let C be a subset of a Banach space X. An *evolution system* on C is a family $\{V(t, s) \ : \ t \geq s \geq 0\}$ of maps from X into itself satisfying:

i) $V(t, t) = I$, the identity operator in X;

ii) $V(t, s)V(s, r) = V(t, r)$; and

If an evolution system satisfies

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

for some $M > 0$, we say it is a *M-Lipschitz evolution system* (*M-LES*, for short). A *contracting evolution system* (*CES*) is a *1-LES*.

$V(t, s)x$ represents the point where the orbit that started from x at time s will be at instant t. If $V(t, s)$ does not depend on s and t independently, but only on the time that has elapsed (i.e. $t-s$), we say it is *autonomous*. In this case $V(t,s) = V(t-s,0) = T(t-s)$ for some semigroup T, so we make no distinction between autonomous evolution systems and semigroups.

The most obvious example is the set of solutions of a differential equation. Suppose that for each $s \geq 0$ the differential equation

$$\begin{cases} \dot{x}(t) &= F(t, x(t)) \\ x(s) &= x \end{cases}$$

has a unique solution $u_{s,x}$ defined for all $t \geq s$. Then $U(t,s)x = u_{s,x}(t)$ defines an evolution system.

But this notion can be used to deal with discrete-time systems as well. Take a strictly increasing sequence of positive numbers σ_n tending to ∞. Suppose an algorithm generates a sequence $\{x_n\}$ from each given x_0 in X. It is possible to build a continuous-time trajectory by piecewise constant interpolation of the points in the sequence. More precisely, let $\nu(t) = \max\{k \in \mathbf{N} | \sigma_k \leq t\}$ and define $u(t) = x_{\nu(t)}$ on the interval $[\sigma_{\nu(t)}, \sigma_{\nu(t)+1})$. Clearly, $u(t)$ converges as $t \to \infty$ if, and only if, x_n converges as $n \to \infty$. Keeping this in mind one can think of algorithms as yielding continuous-time trajectories and treat them as evolution systems. This way, we make no distinction between evolution systems arising, for instance, from differential equations, or from algorithms.

Let V be an evolution system on C and consider a locally bounded function $u : \mathbf{R}_+ \to C$. Following [84], we shall say u is an *almost-orbit* of V if

$$\lim_{t\to\infty} \left[\sup_{h\geq 0} \|u(t+h) - V(t+h,t)u(t)\| \right] = 0.$$

Clearly the orbits of V are almost-orbits as well. Let U and V be two evolution systems. If every orbit of U is an almost-orbit of V and viceversa we say they are *asymptotically equivalent*.

Classical results on asymptotic equivalence. The first work that relates the convergence properties of two evolution systems (in any Banach space) seems to be [90]. The author first proves that if U and V are equivalent (in a sense that is a special case of the definition above), then the orbits of U converge strongly or weakly if, and only if, the orbits of V do. Next he proves that the systems defined by the proximal point algorithm $x_{n-1} - x_n \in \lambda_n A x_n$ and the differential inclusion $-\dot{u} \in Au$ *are* equivalent in two special cases:

i) The operator A is m-accretive, single-valued and Lipschitz and $\{\lambda_n\} \in \ell^2 \setminus \ell^1$; or

ii) A is m-accretive and the sequence $\{\lambda_n\}$ satisfies a more sophisticated summability condition that implies $\{\lambda_n\} \in \ell^2 \setminus \ell^1$.

A few years later, those systems are proved to be equivalent for any m-accretive operator A whenever $\{\lambda_n\} \in \ell^2 \setminus \ell^1$ (see [70]).

Finally in [56] the author proves that if A is the subdifferential of a proper, lower-semicontinuous convex function in Hilbert space then the condition $\{\lambda_n\} \notin \ell^1$ suffices to guarantee the equivalence of the systems. A remarkable consequence of this particular result is the existence of a proper, lower-semicontinuous convex funtion such that the proximal point algorithm converges weakly but not strongly, thus giving a complete answer to a question posed earlier by T Rockafellar in [97]. This was done by using Baillon's example in [17] for the continuous-time scheme and translating the result to the discretization by means of the equivalence theorem.

Now let $v : [0,\infty) \to X$ be a locally integrable function and set

$$\overline{v}_h(s) = \tfrac{1}{s}\int_0^s v(h+\xi)\ d\xi.$$

We say v is strongly (resp. weakly) *almost-convergent* (in the sense of Lorentz) if there exists $y \in X$ such that $\overline{v}_h(s)$ converges strongly (resp. weakly) to y as $s \to \infty$ uniformly in $h \geq 0$. This notion of convergence is important for two reasons: in the first place, a trajectory $v(s)$ is convergent if, and only if, it is almost-convergent and *asymptotically regular* (the difference $v(s+h) - v(s)$ converges to zero as $s \to \infty$ for each $h \geq 0$) for the corresponding topology (see [79]). This fact (or method of proof) is used, for instance, in [20]. Second, some dynamical systems yield trajectories that almost-converge (but do not converge) to a limit having particular properties. This is useful for computational purposes.

In [84], the authors prove the following (under additional hypotheses on the space X): Let T be a strongly continuous semigroup of contractions having fixed points and suppose for each x the trajectory $t \mapsto T(t)x$ is strongly (resp. weakly) almost-convergent. Then so is every almost-orbit.

In [83], the author proves the same result under weaker conditions, namely: no assumptions are made on the Banach space, the notion of almost-orbit is weakened, and finally the semigroup is only assumed to be *asymptotically nonexpansive*, in a sense the author defines. This is the first result that holds for evolution systems that are not contracting. However, no results were found in the literature without assumptions on the behavior with respect to the space variable.

General evolution systems and μ_t-convergence. Let $\{\mu_t\}_{t\geq 0}$ be a family of probability measures on $[0,\infty)$. A locally bounded function $v : \mathbf{R}_+ \to X$ is $\{\mu_t\}$*-regular* if

$$\overline{v}_h(t) = \int_0^\infty v(h+\xi)\ d\mu_t(\xi) \tag{2}$$

exists for all $t, h \geq 0$. We shall say that a $\{\mu_t\}$-regular function v *converges strongly (weakly) along the family* $\{\mu_t\}$ to $y \in X$ if $\overline{v}_h(t)$ converges strongly (weakly) to y as $t \to \infty$ uniformly in $h \geq 0$. This definition unifies different notions of convergence so we can state and prove our results in a more concise way.

Example 1 *If μ_t is the Dirac mass at t, then $\overline{v}_h(t) = v(t+h)$. In this case, convergence along the family $\{\mu_t\}$ is just convergence in the traditional sense of the term.*

Now suppose μ_t has density $d\mu_t(\xi) = \frac{1}{t}\chi_{[0,t]}(\xi)$, where χ_B is the indicator function of the set B. Then $\overline{v}_h(t) = \frac{1}{t}\int_0^t v(h+\xi)\ d\xi$. In this case, convergence along the family $\{\mu_t\}$ is precisely almost-convergence*, as defined in the preceding paragraph.*

Consider the following hypothesis on the family $\{\mu_t\}$:

Hypothesis $\mathbf{H_3}$: For every $\{\mu_t\}$-integrable function v, every $\varepsilon > 0$ and $H > 0$, there exists $T > 0$ such that for all $t \geq T$ and $h \in [0,H]$ one has $\|\mu_t(v_h) - \mu_t(v)\| < \varepsilon$.

This condition guarantees that only asymptotic properties are taken into account since it implies that $\mu_t(B) \to 0$ as $t \to \infty$ for every bounded measurable set B. Notice also that the families described in Example 1 satisfy Hypothesis $\mathbf{H}$.

It turns out that orbits and almost-orbits share this kind of convergence.

Theorem D *Let V be an evolution system and assume the family $\{\mu_t\}$ satisfies Hypothesis* **H**. *If $V(t,s)x$ converges strongly (resp. weakly) along the family $\{\mu_t\}$ for all x and s, then so does every $\{\mu_t\}$-regular almost-orbit. For the claim concerning the weak topology, the space is assumed to be weakly complete.*

In particular, we have the following generalization of all existing abstract results that we could find in the literature:

1. If $V(t,s)x$ converges weakly (resp. strongly) as $t \to \infty$ for all x and s, then so does every almost-orbit.

2. If $V(t,s)x$ almost-converges weakly (resp. strongly) as $t \to \infty$ for all x and s, then so does every almost-orbit.

These results can be improved if the evolution systems is M-Lipschitz: the completeness hypothesis for the weak topology is unnecessary; similar properties hold without the uniformity condition in the definition of almost-convergence; if one almost orbit is bounded, all of them are; one can get *a priori* information on the ω-limit sets of almost-orbits; etc.

We do not know whether the property of "being an almost-orbit of" is transitive. However, notice that if the orbits of U are almost-orbits of V and the orbits of V are almost-orbits of W, then the asymptotic properties mentioned above are inherited from the orbits of W to those of V, and then from those of V to those of U. Therefore, for practical purposes, we can think of that relationship as being transitive in some sense.

Applications. We just mention here a few applications of the notion of asymptotic almost-equivalence:

1. No asymptotic equivalence results were found in the literature for explicit discretizations of equation $-\dot{u} \in Au$, where A is an accretive operator. Let $\{\lambda_n\}$ be a sequence of positive numbers bounded by 1. Starting from $z_0 \in X$ define the sequence $\{z_n\}$ by

$$z_{n+1} \in z_n - \lambda_n A z_n$$

for $n \geq 0$. Now suppose $A = I - T$, where T is a nonexpansive mapping. Let U be the evolution system given by the solutions of the differential inclusion $u + u' = Tu$ with initial condition $u(s) = x \in X$ and let W be the one defined by Euler's scheme by piecewise constant interpolation. If $\{\lambda_n\} \in \ell^2 \setminus \ell^1$, then every bounded orbit of U is an almost-orbit of W and viceversa. Therefore, $u_0(t)$ converges weakly (or strongly) as $t \to \infty$ for every initial condition $x \in X$ if, and only if, z_n converges weakly (or strongly) as $n \to \infty$ for all $z_0 \in X$.

2. Tikhonov's regularization: Let $A(t)$ be a family of maximal monotone operators on a Hilbert space H and $\varepsilon \in L^1(0,\infty;\mathbf{R})$. Consider the differential inclusions

$$u'(t) + A(t)u(t) \ni 0 \qquad \text{and} \qquad v'(t) + A(t)v(t) + \varepsilon(t)v(t) \ni 0.$$

Assume that for each initial condition in H both problems have solutions and they are bounded. Let U and V be the corresponding evolution systems. Assume also that for every $R > 0$ there exists $M > 0$ such that $\|x\| \leq R$ implies $\|U(t,s)x\| \leq M$ (this holds in most applications). Then every orbit of V is an almost-orbit of U.

3. Lipschitz dynamics and diagonal PROX. Let $A(t)$ be a family of (m-)accretive maps on a Banach space X satisfying

$$\begin{aligned} \|A(t)x - A(s)y\| &\leq \alpha\|x-y\| + \Theta(x,y)|t-s| \qquad \text{and} \\ \|A(t)x\| &\leq \Phi(x), \end{aligned}$$

where Θ and Φ are locally bounded functions. Given $x \in X$ and $s \in \mathbf{R}$, the equation

$$\begin{cases} u'(t) &= -A(t)u(t) \\ u(s) &= x \end{cases}$$

has a unique solution $U(t,s)x$. The two-parameter family $\{U(t,s)\}$ is a *CES* and we shall assume there is a locally bounded function $\Psi : X \longrightarrow \mathbf{R}_+$ such that

$$\|U(t,t_0)z - U(s,s_0)z\| \leq \Psi(z)\left[|t-s| + |t_0 - s_0|\right].$$

Take $x_0 \in X$ and $\{\lambda_n\} \in (0,\Lambda]$ consider $x_n = (I + \lambda_n A(\sigma_n))^{-1}x_{n-1}$, where $\sigma_n = \sum_{k=1}^n \lambda_k$. As usual, let $\nu(t) = \max\{\ k \in \mathbf{N} \mid \sigma_k \leq t\ \}$ and define a *CES* by $V(t,s) = \prod_{k=\nu(s)+1}^{\nu(t)}(I + \lambda_k A(\sigma_k))^{-1}$. If $\sum_{k=1}^\infty \lambda_k^2 < \infty$ then every bounded orbit of V is an almost-orbit of U and viceversa.

4. A second-order system: The nonlinear oscillator with damping. In [2], the author studies the problem

$$u''(t) + \gamma u'(t) + \nabla\Phi(u(t)) = 0, \tag{3}$$

in Hilbert space H, where Φ is a $\mathcal{C}^1$ convex function which is bounded from below. He proves that if Argmin$(\Phi) \neq \emptyset$, then $u(t)$ converges weakly to a minimizer of Φ as $t \to \infty$. Strong convergence occurs if $\nabla\Phi$ is strongly monotone, if Φ is even, or if Argmin(Φ) has nonempty interior.

Later, in [7] the authors study a similar differential equation:

$$u''(t) + \gamma u'(t) + \nabla\Phi(u(t)) + \varepsilon(t)u(t) = 0.$$

The authors establish convergence results which depend on whether ε is in L^1 or not. For $\varepsilon \notin L^1$, the trajectory always converges strongly to the least norm element of Argmin(Φ). For $\varepsilon \in L^1$, they obtain the same conclusions as with $\varepsilon \equiv 0$ in most (but not all) of the cases, which is not surprising because one can prove that the corresponding evolution systems are asymptotically almost-equivalent in this case. The asymptotic almost-equivalence theory fills the gaps.

As we mentioned before, the theory of asymptotic equivalence is useful, for instance, if one has a large number of systems for which the asymptotic behavior has been studied extensively. This is the case for those dynamical systems governed by accretive operators, which are often used in optimization, fixed-point and game theory, etc. Chapter 3 contains several results on this topic.

Differential inclusions and products of resolvents.

We wish to underscore here a specific method of discretization and approximation, along with a powerful estimation, that have been used both for establishing existence of solutions to some differential inclusions and for verifying asymptotic equivalence.

Motivated by either the existence or the algorithmic approximation of solutions to a differential inclusion problem of the type

$$\dot{x} + A(t)x \ni 0, \tag{4}$$

where $A(t)$ is a family of accretive operators with domain in a Banach space, several authors have considered some special implicit discretization schemes.

In the autonomous case where $A(t) \equiv A$, Crandall and Liggett introduced in [47] the following limit:

$$S(t)x_0 = \lim_{n\to\infty} \left(I + \tfrac{t}{n}A\right)^{-n} x_0 = \lim_{n\to\infty} (J^A_{t/n})^n x_0, \tag{5}$$

where $J^A_\lambda = (I + \lambda A)^{-1}$ is the resolvent of A. Under some closedness assumptions on the operator A, they proved that this limit exists and defines a strongly continuous semigroup $\{S(t)\}_{t\geq 0}$ on X such that $x(t) := S(t-t_0)x_0$ is the strong solution to $\dot{x} + Ax \ni 0$ that satisfies $x(t_0) = x_0$. They also provided some estimates on $\|(J^A_\lambda)^n x_0 - (J^A_\mu)^m x_0\|$, and established the Lipschitz continuity of the solution.

Kobayashi's inequality. Later Kobayashi recovered in [66] similar existence results with fundamental improvements concerning certain estimates and some continuity properties. He also constructed sequences of approximate solutions which converge in an appropriate sense to a solution to the differential inclusion. The key argument is an inequality that provides an estimate for the distance between arbitrary points of two independent sequences generated by the proximal iterations. More precisely, in the case where $x_k = J^A_{\lambda_k} x_{k-1}$ and $\widehat{x}_l = J^A_{\widehat{\lambda}_l} \widehat{x}_{l-1}$ with (possibly nonconstant) stepsizes $\{\lambda_k\} \subset (0, \Lambda]$ and $\{\widehat{\lambda}_l\} \subset (0, \widehat{\Lambda}]$, Kobayashi's inequality establishes that:

$$\|x_k - \widehat{x}_l\| \leq \|x_0 - u\| + \|\widehat{x}_0 - u\| + \|\!|Au|\!\| \sqrt{(\sigma_k - \widehat{\sigma}_l)^2 + \Lambda\sigma_k + \widehat{\Lambda}\widehat{\sigma}_l},$$

where $\sigma_k = \sum_{i=1}^k \lambda_i$ (similar for $\widehat{\sigma}_l$), $u \in D(A)$ and $\|\!|Au|\!\| = \inf_{[u,v]\in A} \|v\|$.

Further implications of Kobayashi's inequality. Some continuity properties of a limit as (5) follow easily from Kobayashi's inequality. A double limiting process gives explicit Lipschitz constants in terms of the data. Moreover, passing to the limit in only one of the sequences, it is possible to compare the continuous and discrete trajectories, namely, we have an estimate of the type

$$\|x_k - x(t)\| \leq \|x_0 - u\| + \|\!|Au|\!\| \sqrt{(\sigma_k - t)^2 + \Lambda\sigma_k}.$$

In [84], and still for the autonomous case, Miyadera and Kobayasi introduced the notion of *almost-orbit* (special case of the one given above) of $\dot{x} + Ax \ni 0$. They used Kobayashi's inequality to prove that the continuous path constructed by linear interpolation of some proximal iterations is indeed an almost-orbit for the semigroup generated by the operator A. A converse result is given in [70]. It is known that several asymptotic properties of the orbits are inherited by the almost-orbits (see [90], [67] and Chapter 7). Güler's more recent result on the equivalence for the subdifferential case also relies on Kobayashi's inequality.

The nonautonomous case. Crandall and Pazy provided in [49] a suitable generalization of the limiting formula (5). On the other hand, under some additional conditions on the operator-valued function $t \mapsto A(t)$, Kobayasi *et al.* gave in [67, Lemma 3.4] a first nonautonomous version of the original Kobayashi inequality. Then, they obtained important properties of the corresponding continuous dynamics by passing to the limit in an appropriate manner. However, their nonautonomous Kabayashi-type inequality and the resulting estimates in [67] are rather involved, and based on some extrapolations and not on optimal bounds. A better estimation can be proved to hold.

Discrete proximal scheme. Let D be a nonempty subset of X and define

$$\mathcal{M}(D) = \{A : X \rightrightarrows X \mid A \text{ is } m\text{-accretive and } D(A) = D\}.$$

A collection of four sequences $(\{x_k\}, \{\lambda_k\}, \{A_k\}, \{\varepsilon_k\})$ with $\{A_k\} \subset \mathcal{M}(D)$ is said to be a *discrete proximal scheme* if $\lambda_k > 0$, and

$$(x_k - x_{k-1})/\lambda_k + A_k x_k \ni \varepsilon_k,$$

for all $k \geq 1$. The corresponding sequence $\{x_k\}$ is said to be a *discrete proximal trajectory* starting from the point $x_0 \in X$ and generated by $(\{\lambda_k\}, \{A_k\}, \{\varepsilon_k\})$.

Next, given points $x \neq \widehat{x}$ and $v, \widehat{v}$ in X, define

$$\Delta([x,v],[\widehat{x},\widehat{v}]) = \inf_{f \in \mathcal{J}(x-\widehat{x})} \frac{\langle \widehat{v} - v, f\rangle}{\|x - \widehat{x}\|}, \tag{6}$$

where $\mathcal{J}$ is the (unnormalized) duality mapping. If $x = \widehat{x}$ then we set $\Delta([x,v],[\widehat{x},\widehat{v}]) = 0$. Notice that $\Delta([x,v],[\widehat{x},\widehat{v}]) \leq \|v - \widehat{v}\|$. If $[x,v],[\widehat{x},\widehat{v}] \in A$ for some $A \in \mathcal{M}(D)$ then, by monotonicity, we have that $\Delta([x,v],[\widehat{x},\widehat{v}]) \leq 0$.

Let us consider two sequences $\{A_k\}$ and $\{\widehat{A}_l\}$ in $\mathcal{M}(D)$. As in previous works in this subject, we shall assume that the following condition holds:

$$\forall k,l \geq 1,\ \exists \Theta_{k,l} \geq 0,\ \forall\,[x,v] \in A_k,\ \forall\,[\widehat{x},\widehat{v}] \in \widehat{A}_l,\ \Delta([x,v],[\widehat{x},\widehat{v}]) \leq \Theta_{k,l}. \tag{7}$$

Products of resolvents in a nonautonomous setting. Let $A, \widehat{A} : [0,\infty) \longrightarrow \mathcal{M}(D)$. For $m \in \mathbf{N}$ and $t > t_0 \geq 0$, consider the *finite* discrete proximal trajectories $\{x_k\}_{k=0}^m$ and $\{\widehat{x}_l\}_{l=0}^m$ defined by

$$x_0 = u \text{ and } x_k = \left(I - \tfrac{t-t_0}{m} A(t_0 + \tfrac{k(t-t_0)}{m})\right)^{-1} x_{k-1}, \quad \text{for } k = 1,\dots,m;$$

$$\widehat{x}_0 = u \text{ and } \widehat{x}_l = \left(I - \tfrac{t-t_0}{m} \widehat{A}(t_0 + \tfrac{l(t-t_0)}{m})\right)^{-1} \widehat{x}_{l-1}, \quad \text{for } l = 1,\dots,m.$$

From now on, we assume that x_m and $\widehat{x}_m$ converge in X as $m \to \infty$, and we respectively denote by $U(t,t_0)u$ and $\widehat{U}(t,t_0)u$ their limits, that is, for all $u \in D$ and $t > t_0 \geq 0$ we set

$$U(t,t_0)u = \lim_{m\to\infty} x_m = \lim_{m\to\infty} \prod_{k=1}^{m} \left(I - \tfrac{t-t_0}{m} A(t_0 + \tfrac{k(t-t_0)}{m})\right)^{-1} u$$

$$\widehat{U}(t,t_0)u = \lim_{m\to\infty} \widehat{x}_m = \lim_{m\to\infty} \prod_{l=1}^{m} \left(I - \tfrac{t-t_0}{m} \widehat{A}\left(t_0 + \tfrac{l(t-t_0)}{m}\right)\right)^{-1} u.$$

We set $U(t_0,t_0)u = \widehat{U}(t_0,t_0)u = u$. In addition, we will assume that for each $u \in D$, the functions

$$t \mapsto \|A(t)u\| \text{ and } t \mapsto \|\widehat{A}(t)u\| \text{ are locally bounded and Riemann-integrable.} \tag{8}$$

Sufficient conditions on $\{A(t)\}_{t\in[0,\infty)}$ ensuring that $U(t,t_0)u$ is well defined are given in [49]. The function $U(\cdot,t_0)u$ is said to be a *weak* or *DS-limit solution* (DS for "discrete scheme") of inclusion (4). In a time-independent domain these generalized solutions happen to coincide with *integral solutions* in the sense of Bénilan ([27]) under hypothesis (9) below. The trajectory $t \mapsto U(t,t_0)u$ can also be proved to satisfy (4) under supplementary assumptions. We shall not go further on this matter here but only mention that such conditions imply in particular the continuity of $[0,\infty) \ni t \mapsto \|A(t)u\|$, hence (8).

Finally, we assume the continuous version of (7): there exists a bounded and locally Riemann-integrable function $\Theta : [0,\infty) \times [0,\infty) \to [0,\infty)$ satisfying

$$\forall t,s \in [0,\infty),\ \forall\,[x,v] \in A(t),\ \forall\,[\widehat{x},\widehat{v}] \in \widehat{A}(s),\ \Delta([x,v],[\widehat{x},\widehat{v}]) \leq \Theta(t,s), \tag{9}$$

for $\Delta(\cdot,\cdot)$ given by (6).

An estimation for the continuous trajectory. A new Kobayashi-type estimation for the distance between proximal sequences x_n and $\widehat{x}_k$ allows us to prove the following, via a limiting process:

Theorem E *Let $u \in D$ and suppose $t - t_0 \leq s - s_0$. We have*

$$\|U(s,s_0)u - \widehat{U}(t,t_0)u\| \leq \sqrt{2\left[\mathcal{A}_u(s,s_0) - \widehat{\mathcal{A}}_u(t,t_0)\right]^2 - \left[\mathcal{A}_u(t_0+s-t,s_0)\right]^2} + \int_0^{\tau} \Theta(s-\xi, t-\xi)\ d\xi.$$

Here $\mathcal{A}_u(t,t_0) := \int_{t_0}^{t} \|A(\xi)u\|\ d\xi$ and $\widehat{\mathcal{A}}_u(t,t_0) := \int_{t_0}^{t} \|\widehat{A}(\xi)u\|\ d\xi$. Observe that the inequality above is an equality if one takes $\widehat{A}(t)x = A(t)x \equiv c \in X$. Notice also that this estimation shows that the function U automatically inherits continuity properties from the function Θ. For instance, if $\Theta(t,s)$ is locally bounded by the difference $|t-s|$, the function U is locally Lipschitz-continuous in the pair (t_0,t). This method also yields, as a subproduct, a continuous-discrete inequality that can be used to estimate the distance between the discrete and continuous trajectories.

Outline

This dissertation is organized as follows:

PART I: Evolution Equations: Discretization, perturbation and asymptotic analysis.

Chapter 1: In this preliminary chapter we set the notation, discuss on the basic definitions of monotone/accretive operators and give examples. We also recall some facts on Banach space theory.

Chapter 2: We describe some dynamical systems that appear recurrently in the study of optimization problems and comment on existence and basic properties.

Chapter 3: Here we state and prove some of the most relevant results concerning the asymptotic behavior of the systems described in the preceding chapter.

Chapter 4: Finally, we mention some of the now classical results on nonautonomous systems. We also present some new results found in collaboration with Dr. Roberto Cominetti at the University of Chile.

PART II: Asymptotic analysis of nonautonomous evolution systems.

Chapter 5: This chapter deals with a diagonal proximal point algorithm applied to an approximation $\{f_n\}$ for a given objective function f. We provide conditions that guarantee convergence and convergence in average of the sequence $\{f_n(x_n)\}$ of values and estimate the rate of convergence. We also prove that the convergence of the values is faster when the proximal sequence $\{x_n\}$ converges strongly. Numerical tests are presented in order to illustrate these results. Convergence in a finite number of steps is also discussed. These results are contained in a research paper *Asymptotic convergence to the optimal value of diagonal proximal iterations in convex minimization* [92], submitted for publication at Journal of Convex Analysis.

Chapter 6: We provide nonautonomous Kobayashi-type estimates for discrete- and continuous-time evolution systems governed by families of accretive operators in Banach spaces. The continuous-continuous estimate is used in Chapter 8 to prove asymptotic equivalence of the systems defined by two distinct differencial inclusions. The results in this chapter were obtained in collaboration with Dr. Felipe Álvarez at the University of Chile, in Santiago, and are contained in the research paper *Asymptotic equivalence and Kobayashi-type estimates for nonautonomous monotone operators in Banach spaces* [4], submitted for publication at Discrete and Continuous Dynamical Systems.

Chapter 7: Here we develop a general unified theory for the asymptotic equivalence of (nonautonomous) evolution systems in Banach spaces. More precise results are given in the Lipschitz case, which generalize some of the existing ones. These results, which are the most relevant part of this dissertation, were obtained in collaboration with Dr. Felipe Álvarez at the University of Chile, in Santiago, and are contained in the research paper *Asymptotic almost-equivalence of abstract evolution systems* [5], submitted for publication at Proceedings of the American Mathematical Society.

Chapter 8: Several different applications to optimization and fixed-point theory are provided by analyzing both autonomous and nonautonomous problems. This is work in progress with Dr. Felipe Álvarez at the University of Chile, in Santiago.

Perspectives and open questions

Evolution systems are a very powerful tool in optimization and a large amount of research concerning their applications has been carried out in the last decades. However, several questions remain unresolved. We mention only a few among the ones that we consider relevant, challenging and closely related to our work.

Interior barriers. The convergence rates described at the beginning of the Introduction hold when $\mathcal{S} \cap D \neq \emptyset$. However, for several penalization schemes using barriers with an infinite jump (for instance, the log- or the inverse barriers) it is important to consider the case where $\mathcal{S} \cap D = \emptyset$, but $\mathcal{S} \cap \overline{D} \neq \emptyset$. One way to address this problem is to use our estimations in nearby interior points which are not minimizers but give values close to f^* and then try to obtain estimations via a limiting process.

Discrete to continuous. It is useful to compare discrete- and continuous-time evolution systems. For instance, one can obtain important qualitative information on the behavior of algorithms. On the other hand, it is possible to derive, via discretization, existence and global estimations on the solutions of a differential inclusion:

1. A continuous-time analogue to Corollary 5.9 (when $f_n \equiv f$) is proved in [58]. More precisely, let $x(t)$ satisfy $-\dot{x}(t) \in \partial f(x(t))$ almost everywhere on $(0, \infty)$ and suppose $\lim_{t\to\infty} x(t) = x^* \in \mathcal{S}$. Then $f(x(t)) - f^* = o(1/t)$. The necessary estimations are obtained by arguing that the trajectory $x(t)$ can be approximated by a family of proximal sequences, and then translating the discrete-time inequalities into continuous-time versions. It is possible to develop a similar analysis in the nonautonomous setting $-\dot{x}(t) \in \partial f_t(x(t))$ and we conjecture that a continuous-time analogue of Theorem C must hold true. The approximation of the continuous-time trajectory by a family of diagonal proximal sequences needs the more sophisticated tools developed in [49] but follows easily once the conditions on the evolution of the family $\{f_t\}$ are properly stated. However, the approximation of the speed is a complicated issue that remains unresolved.
2. Concerning the Kobayashi-type estimates, we only worked out an equivalence result based on the continuous-continuous inequality. However, the estimation on the distance between the proximal sequence and the continuous trajectory would give an equivalence result in a non-Lipschitz setting that would complement the analysis sketched in Application 3 above. Moreover, it would not be difficult to implement for certain approximation schemes $f_n \to f$.
3. In [2] the author shows that the sequences defined by a second-order algorithm based on the differential equation (3) converge weakly to a minimizer of the function Φ. We have the strong conviction that the same conclusion can be drawn by means of the asymptotic almost-equivalence theory described above. If it were the case, one could also translate the results on strong convergence that are known for the solutions of (3).

More general settings. There are several applications one can think of, where the notions of evolution system and almost-orbit given here might still be too restrictive. We state here some possible extensions that might be useful to study:

1. Set-valued evolution systems: One may also consider a set-valued version of the evolution systems defined before. Let $C \subset X$. A *set-valued evolution system* on C is a two-parameter family $\{\mathcal{V}(t,s)\}_{t\geq s\geq 0}$ of subsets of C such that
   (a) $\mathcal{V}(s,s)x = \{x\}$ for all s and $x \in C$, and
   (b) $\mathcal{V}(t,r)\mathcal{V}(r,s)x = \mathcal{V}(t,s)x$ for all $t \geq r \geq s$ and $x \in C$.

These evolution systems appear when the scheme does not provide a unique orbit for a given initial condition, which is the case for a large number of differential inclusions appearing in game theory. Several notions of almost-orbit can be considered for these systems. An interesting and challenging task is to determine which ones are more suitable and give full versions of the asymptotic almost-equivalence result (Theorem D) in the corresponding setting.

2. Another relevant matter is the uniformity in the limit defining the almost-orbits. Some authors have considered weaker versions of the definition of almost-orbit to obtain results that can be somehow connected with the idea of asymptotic almost-equivalence:

   (a) In [25] and [26] the authors consider $\lim_{t\to\infty} \sup_{0\le h\le H}$ for each $H > 0$ to define *asymptotic pseudotrajectories*. The limit is not uniform in $h \ge 0$ as in our definition of almost-orbit, but just *uniform on compact intervals*. They obtain information on the ω-limit set of these asymptotic pseudotrajectories under some compactness assumptions and provide applications in game theory.

   (b) In [83], the author uses the joint limit $\lim_{t,h\to\infty}$ instead of $\lim_{t\to\infty} \sup_{h\ge 0}$ in the definition of almost-orbit and succeeds in finding an equivalence result in the more restrictive setting of asymptotically nonexpansive semigroups.

3. As we pointed out before, in the Lipschitz case one can obtain equivalence results for averages of the form $\frac{1}{t}\int_0^t u(s)ds$. If the orbits converge "in average", the same is true for almost-orbits. Convergence in average is weaker than almost-convergence, but is also useful for computational purposes. We have not been able to prove that almost-orbits of general evolution systems inherit convergence in average from the orbits and this seems to be a difficult problem.

Asymptotic semigroups. If the evolution system U is equivalent to a semigroup S (this is the case when approximating an autonomous problem; e.g. penalization, regularization, etc). These are the *asymptotic semigroups* in the sense of [90]. Fixed points of S are *almost-stationary points* of U: they satisfy $\lim_{t\to\infty} \|x - U(t+h,t)x\| = 0$ uniformly in $h \ge 0$. It is not clear whether the converse is true. This fact would have important consequences in determining the properties of the limits of convergent almost-orbits of U as well as their ω-limit sets.

Part I

Evolution Equations: Discretization, perturbation and asymptotic analysis.

Preface to Part I

In this short review we provide a compendium of most classical results concerning the asymptotic behavior of some evolution systems that arise in optimization and fixed-point theory. More precisely, we discuss on differential inclusions governed by accretive operators in Banach spaces along with two well-known discretizations, which are commonly used to approximate their solutions numerically. The number of (now classical) works in this area is immense. Our research within the vast bibliography led us to the strong conviction that it would be useful to gather a large number of results, techniques and references in a short self-contained handbook. We provide and comment on existence results for the sake of completeness but the main emphasis is on the conditions ensuring convergence (in a wide sense). Some of the relationships between the differential inclusion and its discretizations become more evident with the introduction of the theory of asymptotic almost-equivalence, which we present in more depth in Part II. We also mention some pioneer works on nonautonomous systems, especially those having implications in optimization theory via perturbations, penalization, approximation, etc. This was done in collaboration with Dr. Sylvain Sorin at Pierre and Marie Curie University, in Paris.

Chapter 1

Preliminaries

1.1 Banach spaces

Let $(X, \|\cdot\|)$ be a Banach space and denote by X^* its dual. The duality product $\langle \cdot, \cdot \rangle : X \times X^* \to \mathcal{R}$ is defined by $\langle u, f \rangle = f(u)$ for all $u \in X$ and $f \in X^*$. The dual space X^* is endowed with the norm $\|f\|_* = \sup_{\|u\| \le 1} \langle u, f \rangle$. The duality mapping $\mathcal{J} : X \rightrightarrows X^*$ is defined by

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

One has the following geometric property from [63]:

Proposition 1.1 *Let $x, y \in X$. $\|x\| \le \|x + \lambda y\|$ for all $\lambda > 0$ if, and only if, there is $f \in \mathcal{J}(x)$ such that $\langle y, f \rangle \ge 0$.*

Proof. We may assume in what follows that $x \neq 0$. If $\langle y, f \rangle \ge 0$ for some $f \in \mathcal{J}(x)$ then for every $\lambda > 0$ we have

$$\|x\|^2 = \langle x, f \rangle \le \langle x + \lambda y, f \rangle \le \|x + \lambda y\| \ \|x\|.$$

Conversely, take $f_\lambda \in \mathcal{J}(x + \lambda y)$ and set $g_\lambda = f_\lambda / \|f_\lambda\|$ so that $\|g_\lambda\| = 1$. Then

$$\|x\| \le \|x + \lambda y\| = \langle x + \lambda y, g_\lambda \rangle = \langle x, g_\lambda \rangle + \lambda \langle y, g_\lambda \rangle \le \|x\| + \lambda \langle y, g_\lambda \rangle.$$

Therefore $\liminf_{\lambda \to 0} \langle x, g_\lambda \rangle \ge \|x\|$ and $\langle y, g_\lambda \rangle \ge 0$. Since the closed unit ball B_* in X^* is compact for the weak* topology, there is a sequence $\{\lambda_k\}$ of positive numbers such that $\lambda_k \to 0$ as $k \to \infty$ and g_{λ_k} converges to $g \in B_*$ in the weak* topology as $k \to \infty$. Clearly g satisfies $\langle x, g \rangle \ge \|x\|$ and $\langle y, g \rangle \ge 0$. Thus $\|g\| = 1$ and $\langle x, g \rangle = \|x\|$. Finally we set $f = \|x\| g$. We get $f \in \mathcal{J}(x)$ and $\langle y, f \rangle \ge 0$. ■

A Banach space X is *strictly convex* if the unit ball is a strictly convex set. That is, if $\|x\| = \|y\| = 1$, $x \neq y$ and $\alpha \in (0, 1)$, then $\|\alpha x + (1 - \alpha) y\| < 1$. It is easy to prove that if the dual X^* of X is strictly convex the duality mapping $\mathcal{J}$ is single-valued and if X is also reflexive then $\mathcal{J}$ is continuous (strong-weak) and the function ϕ defined by $\phi(x) = \frac{1}{2}\|x\|^2$ is Gâteaux-differentiable with $\phi'(x; h) = \langle h, \mathcal{J}(x) \rangle$.

Given a nonempty closed set $C \subset X$ and a point $x \in X$, we denote $d(x, C) = \inf_{y \in C} \|x - y\|$. The set C is *proximinal* if for each $x \in X$ there is a *nearest point* in C; that is, there is $y(x) \in C$ such that $d(x, C) = \|x - y(x)\|$. If X is reflexive, then every nonempty weakly closed set is proximinal (in particular, every closed convex set). If, in addition, C is convex and X is strictly convex, the nearest point is unique. In this case we can define a function called the *nearest-point mapping* by $P_C x = \min_{y\in C} \|x - y\|$.

A space X is *uniformly convex* if for each $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x, y \in X$ such that if $\|x\| \le 1$, $\|y\| \le 1$ and $\|x + y\| \ge 2 - \delta$, then $\|x - y\| < \varepsilon$. Every uniformly convex space is strictly convex and reflexive. Notice that if X^* is uniformly convex then $\mathcal{J}$ is uniformly continuous (strong-strong) on every bounded subset of X (see [63]). Therefore the function ϕ defined above is Fréchet-differentiable with $D\phi(x) = \mathcal{J}(x)$.

Finally, X satisfies *Opial's condition* (see [89]) if whenever $x_n \rightharpoonup x$ one must have

$$\limsup_{n\to\infty} \|x_n - x\| < \limsup_{n\to\infty} \|x_n - y\|$$

for all $y \neq x$.

1.2 Monotone and accretive operators

Given a set valued mapping $A : X \rightrightarrows X$, its graph and domain are given respectively by $Gr(A) = \{[u, v] \in X \times X \mid v \in Au\}$ and $D(A) = \{u \in X \mid Au \neq \emptyset\}$. For convenience of notation, sometimes we will identify A with its graph by writing $[u, v] \in A$ for $v \in Au$. A mapping $A : X \rightrightarrows X$ is said to be a *monotone operator* if for all $[u_1, v_1], [u_2, v_2] \in A$ there exists $f \in \mathcal{J}(u_1 - u_2)$ such that

$$\langle v_1 - v_2, f\rangle \ge 0. \tag{1.1}$$

If the previous inequality holds for all $f \in \mathcal{J}(u_1 - u_2)$, the operator is *accretive in the sense of Browder* (see, for instance [41]).

A monotone operator is said to be *maximal* if its graph is not properly contained in the graph of any other monotone operator.

Let I be the identity mapping in X. For $\lambda > 0$, the *resolvent* of A is defined as the mapping $J^A_\lambda = (I + \lambda A)^{-1}$.

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

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

This implies that J^A_λ is a single-valued nonexpansive mapping. If, in addition, the range of $I + \lambda A$ equals X for all $\lambda > 0$[1], the operator A is said to be *m-accretive*. Clearly, if $\mathcal{J}$ is single valued and A is m-accretive, then for each $x \in D(A)$ the set Ax is closed and convex. The following summarizes some results in [82] and [63].

Theorem 1.2 (Minty-Kato) *Let $A : X \rightrightarrows X$. Then*

i) A is monotone if, and only if, it is accretive;

ii) If A is m-accretive, then it is maximal monotone.

iii) If X is Hilbert space and A is maximal monotone, then it is m-accretive.

Proof.

i) The result follows from Proposition 1.1.

ii) Let $[u, u^*] \in X \times X$ such that $\langle u^* - v^*, f\rangle \ge 0$ for all $[v, v^*] \in A$ and some $f \in \mathcal{J}(u - v)$. Since $I + A$ is surjective, there is $[\overline{v}, \overline{v}^*] \in A$ such that $\overline{v} + \overline{v}^* = u + u^*$. Then $\|u - \overline{v}\|^2 \le 0$ and so $u = \overline{v}$. Finally, $u^* = \overline{v} + \overline{v}^* - u = \overline{v}^*$ and $[u, u^*] \in A$.

[1] It can be proved that this happens for all or no $\lambda > 0$.

iii) It suffices to prove that $I + A$ is surjective. Take $z_0 \in X$. We shall find $x_0 \in X$ such that $\langle y - (z_0 - x_0), x - x_0\rangle \geq 0$ for all $(x, y) \in Gr(A)$ so that maximality of A implies $z_0 - x_0 \in Ax_0$. Consider the family of weakly compact sets $\{C_{x,y}\}_{(x,y)\in G_A}$ defined as

$$C_{x,y} = \{x_0 \in H : \langle y + x_0 - z_0, x - x_0\rangle \geq 0\}.$$

It suffices to show that this family has the finite intersection property. To this end take $(x_i, y_i) \in G_A$ for $i = 1, \ldots, n$. Let $\Delta = \{(\lambda_1, \ldots, \lambda_n) : \lambda_i \geq 0; \sum_{i=1}^n \lambda_i = 1\}$ denote the n-dimensional simplex and consider the function $f : \Delta \times \Delta \to \mathcal{R}$ given by

$$f(\lambda, \mu) = \sum_{i=1}^n \mu_i \langle y_i + x(\lambda) - z_0, x(\lambda) - x_i\rangle$$

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 minimax theorem implies the existence of $\lambda_0 \in \Delta$ such that

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

Now, the monotonicity of A implies

$$\begin{aligned} f(\mu, \mu) &= \sum_{i=1}^n \mu_i \langle y_i, x(\mu) - x_i\rangle + \langle x(\mu) - z_0, x(\mu) - x(\mu)\rangle \\ &= \sum_{i,j=1}^n \mu_i\mu_j \langle y_i, x_j - x_i\rangle \\ &= \tfrac{1}{2} \sum_{i,j=1}^n \mu_i\mu_j \langle y_i - y_j, x_j - x_i\rangle \leq 0 \end{aligned}$$

so that $f(\lambda_0, \mu) \leq 0$ for all $\mu \in \Delta$, and taking for μ the canonical vectors we get $\langle y_i+x(\lambda_0)-z_0, x(\lambda_0)-x_i\rangle \leq 0$ for all i, that is to say $x(\lambda_0) \in C_{x_1,y_1} \cap \cdots \cap C_{x_n,y_n}$. ■

In general Banach spaces, the converse of ii) does not hold. See [60, Chapter V] for a counterexample.

The *Yosida approximation* of a m-accretive operator A is the single-valued maximal monotone operator $A_\lambda = \frac{1}{\lambda}(I - J_\lambda^A)$.

The *solution set* of A is $\mathcal{S} = A^{-1}0$. Observe that $A_\lambda^{-1}0 = \mathcal{S}$ for all λ.

Example 2 *In Banach space, if $A = I - T$, where T is a nonexpansive mapping, then A is m-accretive and the solution set $\mathcal{S}$ is the set of fixed points of T. In Hilbert space, the subdifferential of a proper lower-semicontinuous convex function ϕ is maximal monotone and $\mathcal{S}$ is the set of minimizers of ϕ.*

For $x \in D(A)$ we define $|\!|\!|Ax|\!|\!| = \inf\{\ \|y\| \mid y \in Ax\}$. The *minimal section* of A is the operator A^0 defined by $A^0x = \{y \in Ax \mid \|y\| = |\!|\!|Ax|\!|\!|\ \}$,[2] which is clearly accretive but not necessarily m-accretive. Observe that if both X and X^* are strictly convex and reflexive, then $A^0x = P_{Ax}0$ is a single-valued operator.

The following results summarize the main properties of accretive operators. The proofs can be found in the classical book [22]:

Proposition 1.3 *Let A be m-accretive. We have:*

1. $\|J_\lambda^A x - J_\lambda^A y\| \leq \|x - y\|$.

[2]This is not always well-defined. However, if the set Ax is proximinal for each $x \in D(A)$ (the minimal norm is attained) then $D(A^0) = D(A)$. This occurs, for instance, if X^* is strictly convex and reflexive because Ax will be a closed convex set for each $x \in D(A)$

2. A_λ *is monotone and* $2/\lambda$*-Lipschitz.*

3. $A_\lambda x \in AJ_\lambda^A x$

4. $\|A_\lambda x\| \leq |\!|\!|Ax|\!|\!|$

5. $\lim_{\lambda\to 0} J_\lambda^A x = x$

6. A *is* closed*:* $x_n \to x$*,* $y_n \to y$ *and* $[x_n, y_n] \in A$ *together imply* $y \in Ax$*.*

7. *If* $x_\lambda \to x$ *and* $A_\lambda x_\lambda \to y$ *as* $\lambda \to 0$*, then* $y \in Ax$*.*

Proposition 1.4 *Let* X^* *be uniformly convex and let* A *be m-accretive. We have the following:*

1. A *is* demiclosed*:* $x_n \to x$*,* $y_n \rightharpoonup y$ *and* $[x_n, y_n] \in A$ *together imply* $y \in Ax$*.*

2. *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* $y \in Ax$*.*

3. $\|A_\lambda x\|$ *is nonincreasing in* λ *and* $\|A_\lambda x\| \to |\!|\!|Ax|\!|\!|$*.*

4. Ax *is a closed convex subset of* X*.*

5. $\mathcal{J}(A^0 x)$ *is single-valued and* $\lim_{\lambda\to 0} \mathcal{J}(A_\lambda x) = \mathcal{J}(A^0 x)$

6. A^0 *characterizes* A *in the following sense: If* A *and* B *are m-accretive with common domain and* $A^0 = B^0$*, then* $A = B$*.*

7. *If* X *is also uniformly convex, then* $\lim_{\lambda\to 0} A_\lambda x = A^0 x$ *and* $\overline{D(A)}$ *is convex.*

1.3 Special classes of monotone operators

1.3.1 Demipositive

A maximal monotone operator A on a Hilbert space H is *demipositive* if there exists $u_0 \in \mathcal{S}$ such that for every sequence $(u_n) \in D(A)$ converging weakly to u and every bounded sequence (v_n) such that $v_n \in Au_n$ we have

$$\langle v_n, u_n - u_0\rangle \to 0 \qquad \text{implies} \qquad u \in \mathcal{S}.$$

This concept was developed in [36], where the reader will find the following examples:

1. Let $\Gamma_0(H)$ be the set of all proper lower-semicontinuous convex functions on H. If $A = \partial\phi$ with $\phi \in \Gamma_0(H)$ having minimizers, then A is demipositive;

2. $A = I - T$ with T nonexpansive and having a fixed point;

3. A maximal monotone such that $\text{int}\mathcal{S} \neq \emptyset$; and

4. A maximal monotone, odd and *firmly positive*, which means that there is $y_0 \in \mathcal{S}$ such that $v \in Ax$ and $\langle v, x - y_0\rangle = 0$ together imply $0 \in Ax$.

5. A maximal monotone, firmly positive and weakly closed.

Recall that an operator A is N-monotone if $\sum_{n=1}^{N} \langle y_n, x_n - x_{n-1}\rangle \geq 0$ for every set $\{[x_n, y_n] \mid 1 \leq n \leq N\} \subset A$ $(x_0 \equiv x_N)$. An additional example found in [91]:

6. A maximal monotone with $\mathcal{S} \neq \emptyset$ and 3-monotone.

Extensions: For demipositivity in Banach spaces see [41].

1.3.2 Strongly monotone

Let $\alpha > 0$. An operator A is *α-strongly monotone* if for all $[x, x^*], [y, y^*] \in A$ there is $f \in \mathcal{J}(x - y)$ such that

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

Observe that if A is strongly monotone and $Ax \cap Ay \neq \emptyset$, then $x = y$. In particular, the solution set is, at most, a singleton. Observe also that if A is α-strongly monotone, it is easy to see using Proposition 1.1, that $J^A_{1/\alpha}$ is a strict contraction. Therefore it has a fixed point $\bar{x}$ and only one. This implies $\mathcal{S} = \{\bar{x}\}$. Strongly monotone operators are demipositive.

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

1.3.3 The Nevanlinna-Reich convergence condition

We shall say that a m-accretive operator A on X satisfies the *Nevanlinna-Reich convergence condition* if $\mathcal{S} \neq \emptyset$ and for every bounded sequence $[x_n, y_n] \in A$ such that $\liminf_{n\to\infty}\langle y_n, \mathcal{J}(x_n - Px_n)\rangle = 0$ one has $\liminf_{n\to\infty} \|x_n - Px_n\| = 0$. Here P is a projection onto $\mathcal{S}$. Strongly monotone operators have this property. So do operators having compact resolvent and those satisfying $\langle y, \mathcal{J}(x - Px)\rangle > 0$ for all $[x, y] \in A$ such that $x \notin \mathcal{S}$.

1.4 Several notions of convergence

Let τ be the weak or the strong topology on X and let $x : [0, \infty) \to X$ be a trajectory in X. x *converges* to $\bar{x}$ for the topology τ if for each τ-neigborhood $\mathcal{V}$ of $\bar{x}$ there is $T \geq 0$ such that $x(t) \in \mathcal{V}$ for every $t \geq T$. We present some weaker notions of convergence that we will find in some of the results in Chapter 7:

The trajectory x is *almost-convergent* if there is $\bar{x}$ such that

$$\frac{1}{t}\int_0^t x(\xi + h)\, d\xi$$

converges to $\bar{x}$ as $t \to \infty$ uniformly in $h \geq 0$. A slightly weaker notion is the following: the trajectory x *converges in average* if

$$\frac{1}{t}\int_0^t x(\xi)\, d\xi$$

converges as $t \to \infty$.

Suppose now that one is 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 interpolating: for $t \in [\sigma_n, \sigma_{n+1}]$, take $x(t) \in [x_n, x_{n+1}]$, where $[a, b]$ denotes the segment joining a and b. It is easy to see that any trajectory defined this way converges to some $\bar{x}$ for the topology τ if, and only if, the sequence $\{x_n\}$ is converges to $\bar{x}$; that is if every $\tau-$neigborhood of $\bar{x}$ contains all but a finite number of the elements of $\{x_n\}$.

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 $\{\lambda_n\}$ of positive numbers, called *stepsizes*.

Given a sequence $\{x_n\}$ in X along with stepsizes $\{\lambda_n\}$, we shall say that $\{x_n\}$ is *almost-convergent* if there is $\bar{x}$ such that $\frac{1}{\sigma_{i+n}-\sigma_i}\sum_{k=1}^{n}\lambda_{i+k}x_{i+k}$ converges to $\bar{x}$ as $n \to \infty$ uniformly in $i \geq 0$. A slightly weaker notion is the following: The sequence $\{x_n\}$ *converges in average* if $\frac{1}{\sigma_n}\sum_{k=1}^{n}\lambda_k x_k$ converges as $n \to \infty$.

Remark 1.5 *The concept of almost-convergence, introduced in [79], is of particular interest due to the following fact: We shall say a sequence $\{x_n\}$ is* strongly (weakly) asymptotically regular *if the difference $x_n - x_{n+1}$ converges strongly (weakly) to 0 as $n \to \infty$. It turns out that almost-convergence (with any sequence of stepsizes) and asymptotic regularity together imply convergence of the sequence for the corresponding topology. This fact, or method of proof has been used, for instance, in [20]. For trajectories, asymptotic regularity amounts to $x(t+h) - x(t)$ converging to 0 for each $h \geq 0$ and also implies convergence of the trajectory when combined with almost-convergence.*

1.5 An important tool

The following result from [89] is a very useful tool for proving weak convergence of a sequence without any information about the limit. It is stated in Hilbert space but also holds in Banach spaces satisfying Opial's condition:

Lemma 1.6 (Opial's Lemma) *Let x_n be a sequence in a Hilbert space H and let $F \subset H$. Assume that $\|x_n - f\|$ has a limit as $n \to \infty$ for each $f \in F$ and that every weak cluster point of x_n lies in F. Then x_n converges weakly to some $x^* \in F$.*

Proof. It suffices to prove that x_n has only one weak cluster point. Assume instead that $x_{\phi(n)} \rightharpoonup x^*$ while $x_{\psi(n)} \rightharpoonup y^*$. By hypothesis, $x^*, y^* \in F$. Observe that

$$\|x_{\phi(n)} - y^*\|^2 = \|x_{\phi(n)} - x^*\|^2 + \|x^* - y^*\|^2 + 2\langle x_{\phi(n)} - x^*, x^* - y^*\rangle.$$

Therefore $\lim\limits_{n\to\infty}\|x_n - y^*\|^2 = \lim\limits_{n\to\infty}\|x_n - x^*\|^2 + \|x^* - y^*\|^2$. In a similar way we prove that $\lim\limits_{n\to\infty}\|x_n - x^*\|^2 = \lim\limits_{n\to\infty}\|x_n - y^*\|^2 + \|x^* - y^*\|^2$ and conclude that $x^* = y^*$. ■

Opial's Lemma can be generalized as follows (see [31]):

Lemma 1.7 *Let x_n, y_n be sequences in a Hilbert space H and let $F \subset H$. Denote C_m the closed convex hull of $\{x_n, n \geq m\}$. Assume*

1. *$\|x_n - f\|$ converges for each $f \in F$;*
2. *$d(y_k, C_m)$ goes to 0 as $k \to \infty$, for all m; and*
3. *every weak cluster point of y_k lies in F;*

then y_k converges weakly to a point in F.

Chapter 2

Discrete and continuous dynamical systems

In this section we describe some dynamical systems that have been studied extensively and have several different applications, namely in Optimization and Fixed-point Theory. We also give some useful global estimations. The results concerning their asymptotic behavior will be presented in the next chapter. In what follows, A is a m-accretive operator in a Banach space X.

2.1 The proximal point algorithm (PROX)

Let $\{\lambda_n\}$ be a sequence of positive numbers, which we call *stepsizes*. We shall say $\{x_n\}$ is a *proximal sequence* if it satisfies

$$\begin{cases} \frac{x_n - x_{n-1}}{\lambda_n} & \in & -Ax_n \qquad \text{for all } n \geq 1 \\ x_0 & \in & D(A). \end{cases} \tag{2.1}$$

Observe that $x_n = (I - \lambda_n A)^{-1} x_{n-1} = J^A_{\lambda_n} x_{n-1}$. The existence of such a sequence follows from the definition of m-accretivity. The *velocity* at stage n is $y_n = \frac{x_n - x_{n-1}}{\lambda_n}$.

2.1.1 The origin.

The notion of proximal sequences and the term "proximal" were introduced in [85] for $A = \partial f$ in Hilbert space, where finding x_n corresponds to minimizing the *Moreau-Yosida approximation* of f:

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

Denote $\sigma_n = \sum_{k=1}^n \lambda_k$. In what follows we assume $\sigma_n \to \infty$ as $n \to \infty$ unless otherwise stated. The following result can be found in [56]:

Lemma 2.1 *For any $u \in domf$ we have*

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

Proof. From the subdifferential inequality we have

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

for all u in the domain of f. Thus

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

If we sum up from 1 to n we deduce that

$$2\sigma_n f(u) - 2\sum_{k=1}^{n} \lambda_k f(x_k) \geq \|u - x_n\|^2 + \sum_{k=1}^{n} \lambda_k^2\|y_k\|^2 - \|u - x_0\|^2. \tag{2.2}$$

On the other hand we have $f(x_{n-1}) - f(x_n) \geq \lambda_n\|y_n\|^2$. 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 (2.2) we obtain

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

and the result follows at once. ■

This gives a global estimation for speed decay, namely:

$$\|y_n\| \leq \frac{d(x_0, \mathcal{S})}{\sigma_n}. \tag{2.3}$$

A similar estimation had been proved in [33] but the right-hand side is $\sqrt{2}$ times larger. It is stated in Theorem 3.1 along with a similar estimation for general A.

The inequality above also implies $f(x_n) \to f^* = \inf f$. Moreover, if $\mathcal{S} \neq \emptyset$ the rate of convergence is estimated at $O(1/\sigma_n)$. However, if the sequence $\{x_n\}$ is known to converge strongly, then $|f(x_n) - f^*| = o(1/\sigma_n)$ (see [58]).

Convergence of the values had already been proved in [81] if f is coercive and $\lambda_n \equiv \lambda$.

2.1.2 A global estimation.

Back to general m-accretive operators in Banach spaces, the following inequality, due to Kobayashi, provides an estimation for the distance between two proximal sequences:

Proposition 2.2 (Kobayashi Inequality) *Consider stepsizes $\{\lambda_k\} \subset (0, \Lambda]$ and $\{\widehat{\lambda}_l\} \subset (0, \widehat{\Lambda}]$:*

$$\|x_k - \widehat{x}_l\| \leq \|x_0 - u\| + \|\widehat{x}_0 - u\| + \|Au\| \sqrt{(\sigma_k - \widehat{\sigma}_l)^2 + \tau_k + \widehat{\tau}_l}, \tag{2.4}$$

where $\sigma_k = \sum_{i=1}^{k} \lambda_i$ and $\tau_k = \sum_{i=1}^{k} \lambda_i^2$ (similarily for $\widehat{\sigma}_k$ and $\widehat{\tau}_l$).

We first prove the following lemma:

Lemma 2.3 *Given* $[u_1,\ v_1],\ [u_2,\ v_2] \in A$ *and* $\lambda,\ \mu > 0$, *we have*

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

Proof. For $f \in \mathcal{J}(u_1-u_2)$, we have

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

■

Proof of Proposition 2.2: 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, we shall establish inequality (2.4) for the pair $(k,0)$ with $k \ge 0$. By accretivity we have

$$\|x_k-u\| \le \|x_0-u\| + \sigma_k|\!|\!|Au|\!|\!|.$$

Thus

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

because $\sigma_k \le c_{k,0}$. In a similar fashion we prove the inequality for $(0,l)$ with $l \ge 0$. Now suppose (2.4) holds for $(k-1,l)$ and $(k,l-1)$. According to Lemma 2.3,

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

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

$$\begin{aligned}\|x_k-\widehat{x}_l\| &\le \alpha_{k,l}\|x_{k-1}-\widehat{x}_l\| + \beta_{k,l}\|\widehat{x}_{l-1}-x_k\| \\ &\le \alpha_{k,l}\left[\|x_0-u\|+\|\widehat{x}_0-u\|+c_{k-1,l}|\!|\!|Au|\!|\!|\right] \\ &\quad +\beta_{k,l}\left[\|x_0-u\|+\|\widehat{x}_0-u\|+c_{k,l-1}|\!|\!|Au|\!|\!|\right] \\ &= \|x_0-u\|+\|\widehat{x}_0-u\|+[\alpha_{k,l}c_{k-1,l}+\beta_{k,l}c_{k,l-1}]\,|\!|\!|Au|\!|\!|. \qquad (2.5)\end{aligned}$$

It only remains to verify that

$$\alpha_{k,l}c_{k-1,l}+\beta_{k,l}c_{k,l-1} \le c_{k,l}. \qquad (2.6)$$

From the Cauchy-Schwarz Inequality we have

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

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 (2.5) and (2.6) give (2.4). ■

Kobayashi's original inequality also accounts for possible errors in the determination of the proximal sequence (see [66]).

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

Let A be a m-accretive operator and take $x_0 \in D(A)$. Consider the following differential inclusion:

$$\begin{cases} \dot{x}(t) & \in & -Ax(t) \quad \text{a.e. on } (0,\infty) \\ x(0) & = & x_0. \end{cases} \tag{2.7}$$

2.2.1 Classical existence result in Hilbert space setting.

If H is Hilbert space we have:

Theorem 2.4 *There exists a unique absolutely continuous function $x : [0,\infty) \to H$ satisfying* (2.7)*. Moreover,*

1. $\|\dot{x}(t)\| \leq \|A^0 x_0\|$ *almost everywhere on* $(0,\infty)$.
2. $x(t) \in D(A)$ *for all* $t \geq 0$ *and* $\|A^0 x(t)\|$ *decreases.*
3. $A^0 x(t)$ *is continuous from the right and* $x(t)$ *admits a right derivative for all* $t \geq 0$*; namely* $\dot{x}(t^+) = -A^0 x(t)$ *(lazy behavior).*

The problem of finding a trajectory satisfying (2.7) was first posed and studied in [71] and [48]. The classical proof can be found in [29]. The idea is to consider the differential inclusion (2.7) with $A = A_\lambda$ and prove that it has a solution u_λ. Then one verifies that u_λ converges to some u satisfying (2.7) for the original A. The following estimation plays a crucial role in the proof:

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

Finally u is proved to have the properties enumerated in Theorem 2.4. Uniqueness of solution follows at once from monotonicity.

2.2.2 The theory for Banach spaces.

The same method can be extended to Banach spaces X such that X and X^* are uniformly convex (see [63]). In general Banach spaces, existence and uniqueness can also be derived by the method in [47], via an approximation of the trajectory by the PROX discretization 2.1:

Set $t \in [0,T]$, $m \in \mathbf{N}$ and run the proximal point algorithm with constant stepsizes $\lambda_k \equiv t/m$. Denote the m-th iteration by $x_m(t)$ and repeat the procedure for each m. The following result can be found in [47]:

Proposition 2.5 *The sequence $\{x_m(t)\}$ defined above converges to some $x(t)$ uniformly on every compact interval $[0,T]$. Moreover, the function $t \mapsto x(t)$ is differentiable and satisfies $-\dot{x}(t) \in Ax(t)$ almost everywhere on $[0,\infty]$.*

Proof. Instead of the original proof we present an easier one using Kobayashi's inequality (2.4)[1]. Fix $N, M \in \mathbf{N}$ and $t, s \in [0,T]$ with $T > 0$. Set $\lambda_n = t/N$ for all n and $\widehat{\lambda}_m = s/M$ for all m. Initialize $x_0(t)$ and $\widehat{x}_0(s)$ both at $u = x_0$. For $n = N$ and $m = M$ we have

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

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

We conclude that the sequence $\{x_n\}$ converges uniformly on $[0,T]$ to a function x, which is uniformly Lipschitz-continuous with constant $|\!|\!|Ax_0|\!|\!|$.

For simplicity, we present the rest of the proof in Hilbert space, but the argument works in any Banach space with slight modifications. In order to prove that the function x satisfies (2.7) it suffices to verify that it is an *integral solution* in the sense of Bénilan (see [27]), which means that for all $[u,v] \in A$ and $t > s \geq 0$ we have

$$\|x(t) - u\|^2 - \|x(s) - u\|^2 \leq -2 \int_s^t \langle v, x(\tau) - u \rangle \, d\tau. \tag{2.8}$$

Since x is absolutely continuous, (2.8) implies $-\dot{x}(t) \in Ax(t)$ almost everywhere on $[0,T]$. From the monotonicity of A, for any proximal sequence $\{x_k\}$ we have $\langle x_{k-1} - x_k - \lambda_k v, x_k - u \rangle \geq 0$. But $\|x_k - u\|^2 - \|x_{k-1} - u\|^2 \leq 2\langle x_{k-1} - x_k, x_k - u \rangle$ and so

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

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

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

Setting $x_0 = x(s)$ and passing to the limit appropriately we finally get (2.8). Notice that $x(t) \in D(A)$ by maximality. ■

In a few words this result expresses that the proximal sequence approaches the continuous-time trajectory as the mesh is refined. Back to Proposition 2.2, we have

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

Therefore, for trajectories x and y we get

$$\|y(s) - x(t)\| \leq \|y(0) - u\| + \|x(0) - u\| + |\!|\!|Au|\!|\!| \; |s - t|.$$

In particular, the solution of 2.7 is unique and

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

2.2.3 Semigroup of contractions

Let $S(t)x_0 = x(t)$ be the value at t of the solution starting from x_0. S forms a strongly continuous semigroup of contractions on $D(A)$. We have:

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

Observe that the set of fixed points of the semigroup coincides with the solution set $\mathcal{S}$ and that $\mathcal{S} \neq \emptyset$ if, and only if, $t \mapsto S(t)x$ is bounded for each x.

Let C be a closed convex subset of X and S a semigroup of nonlinear contractions on C. Then S has a generator: there exists a m-accretive operator A with domain $D(A)$ dense in C, such that for every $x \in D(A)$ the mapping $t \mapsto x(t) = S(t)x$ is the unique absolutely continuous solution of (2.7). Moreover (see [22]), one has

$$\lim_{t\to 0} \frac{x - S(t)x}{t} = A^0 x. \tag{2.9}$$

Conversely, every strongly continuous semigroup of contractions defines an accretive operator via the limit formula (2.9).

2.2.4 Further features of certain particular systems

In general, if the initial point x_0 is in the closure of $D(A)$ we can only guarantee that $x(t) \in \overline{D(A)}$. However, some kinds of operators have a *regularizing* or *smoothing effect* (see, for instance, Theorems 3.2 and 3.3 in [29]):

Proposition 2.6 *If $A = \partial\phi$ or $intD(A) \neq \emptyset$ and $x_0 \in \overline{D(A)}$ then $x(t) \in D(A)$ for all $t > 0$.*

If $A = \partial\phi$ in Hilbert space H and $\mathcal{S} \neq \emptyset$ one can get an estimation for speed decay. More precisely, Theorem 3.7 in [29] shows that

$$\left\| \frac{d^+}{dt} S(t)u_0 \right\| \leq \frac{d(u_0, \mathcal{S})}{t}$$

for all $t > 0$.

2.3 Euler's explicit discretization

Define a sequence $\{z_n\}$ recursively by

$$\begin{cases} \frac{z_n - z_{n-1}}{\lambda_n} & \in & -Az_{n-1} \qquad \text{for all } n \geq 1 \\ x_0 & \in & D(A) \end{cases} \tag{2.10}$$

Assume A maps $D(A)$ into itself. A remarkable feature of this scheme is that the terms of the sequence can be computed explicitly. Observe that if $A = I - T$ with T nonexpansive and $\lambda_n \equiv 1$ then $z_n = T^n x_0$. This particular case has been studied extensively by several authors in the search for fixed points. Some of their results will be presented in the next chapter. It is also important to say that if A is of the form $I - T$ for a nonexpansive T, then a Kobayashi-type inequality holds too. This fact was recently pointed out by [105].

Continuous vs Euler: Chernoff's Estimate

Proposition 2.7 *If T is non-expansive and v satifies*

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

with $v(0) = v_0$ then

$$\|v(t) - T^n v_0\| \leq \|v'(0)\| \sqrt{\lambda t + (n\lambda - t)^2}.$$

Remark 2.8 *Since $A = (-1/\lambda)(I - T)$ is a monotone operator, one can use Kobayashi's inequality to get*

$$\|v(t) - T^n v_0\| \leq \|Av_0\| \frac{t}{\sqrt{n}}.$$

This is precisely Chernoff's estimate with $\lambda = t/n$.

Chapter 3

Asymptotic behavior

In this chapter we present the main results concerning the asymptotic behavior of the systems described above in terms of different kinds of convergence. We also point out some extensions. The first three sections correspond to each of the systems described in the preceding chapter. Each section contains a subsection dealing with

1. Weak convergence;
2. Strong convergence;
3. Other kinds of convergence, where we include almost-convergence and convergence in average.

In the last section we point out some relations between the systems (2.1), (2.7) and (2.10) and show how one can derive information about the three systems by studying only one of them.

3.1 The proximal point algorithm (PROX)

In this section $\{x_n\}$ is a proximal sequence as defined in (2.1):

$$\begin{cases} y_n = \frac{x_n - x_{n-1}}{\lambda_n} & \in & -Ax_n \qquad \text{for all } n \geq 1 \\ \qquad\quad x_0 & \in & D(A). \end{cases}$$

Recall that $x_n = (I + \lambda_n A)^{-1} x_{n-1}$. Unless otherwise stated, the underlying space is Hilbert.

3.1.1 Weak Convergence

The following result can be found in [33]:

Theorem 3.1 *Let A be maximal monotone with $\mathcal{S} \neq \emptyset$. Denote $\sigma_n = \sum_{m \leq n} \lambda_m$ and $\tau_n = \sum_{m \leq n} \lambda_m^2$.*

i) If $\{\lambda_n\} \notin \ell^2$ then x_n converges weakly to some $x^* \in \mathcal{S}$ and $\|y_n\| \leq d(x_0, \mathcal{S})\tau_n^{-1/2}$;

ii) If $\{\lambda_n\} \notin \ell^1$ and A is demipositive then x_n converges weakly to some $x^* \in \mathcal{S}$.

iii) If $A = \partial f$ we also have $\|y_n\| \leq \sqrt{2} d(x_0, \mathcal{S})\sigma_n^{-1}$.

Proof. First, we have $\langle y_n - y_{n-1}, x_n - x_{n-1}\rangle \geq 0$ which implies $\langle y_n - y_{n-1}, y_n\rangle \leq 0$ and so $\|y_n\|$ is decreasing. For any $x \in \mathcal{S}$ we have

$$\|x_n - x\|^2 + \lambda_n^2\|y_n\|^2 \leq \|x_{n-1} - x\|^2$$

so that $\|x_n - x\|^2$ converges and

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

For part i), since $\tau_n \to \infty$, we have $\|y_n\| \to 0$. This implies every weak cluster point lies in $\mathcal{S}$. To see this, let $[u, v] \in A$. By monotonicity,

$$\langle v, x_n - u\rangle \leq \langle -y_n, x_n - u\rangle \leq \|y_n\| \cdot \|x_n - u\|.$$

Since $\{x_n\}$ is bounded, every weak cluster point w satisfies $\langle v, w - u\rangle \leq 0$, thus $w \in \mathcal{S}$. Finally, we conclude using Opial's lemma.

For part ii), the result can also be derived from Opial's lemma if we prove that $x_{n_k} \rightharpoonup u$ implies $u \in \mathcal{S}$. Let y be the "special" element of $\mathcal{S}$ in the definition of demipositivity. Using Lemma 3.2 below we construct another subsequence $\{x_{m_k}\}$ such that $\|x_{m_k} - x_{n_k}\|$ and $\langle x_{m_k} - y, y_{m_k}\rangle$ tend to 0 as $k \to \infty$. Since $x_{m_k} \to u$ and A is demipositive, u must belong to $\mathcal{S}$.

We shall not prove the estimation in iii) here because we already presented a better one in (2.3). ∎

Lemma 3.2 *Assume A is maximal monotone, $\{\lambda_n\} \notin \ell^1$ and $y \in \mathcal{S}$. Let $\{x_n\}$ be a proximal sequence. For each $\varepsilon > 0$ there is N such that for any $n \geq N$, we can find $m \in \mathbf{N}$ satisfying $N \leq m \leq n$, $\|x_m - x_n\| \leq \varepsilon$ and $\langle y_m, x_m - y\rangle \leq \varepsilon$.*

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

$$\sum_k \lambda_k\langle y_k, x_k - y\rangle < \infty. \tag{3.1}$$

Given $\varepsilon > 0$, define $P = \{k \in \mathbf{N} \mid \langle y_k, x_k - y\rangle \geq \varepsilon\}$ so that $\sum_{k\in P} \lambda_k < \infty$. Since $\|x_{k-1} - x_k\| = \lambda_k\|y_k\|$ and $\|y_k\|$ is decreasing one has $\sum_{k\in P} \|x_{k-1} - x_k\| < \infty$.

Let N_1 so that $\sum_{k\in P, k\geq N_1} \|x_{k-1} - x_k\| < \varepsilon$. By virtue of (3.1), since $\{\lambda_n\} \notin \ell^1$ there is $N \geq N_1$ with $\langle y_N, x_N - z\rangle \leq \varepsilon$. 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 $\|x_m - x_n\| \leq \sum_{m<k\leq n} \|x_{k-1} - x_k\| \leq \varepsilon$. ∎

Extensions:

1. Theorem 3.1 is still true if the sequence satisfies $\|x_n - (I + \lambda_n A)^{-1}x_{n-1}\| \leq \varepsilon_n$ with $\sum \varepsilon_n \leq \infty$. This is proved in the same article but can be derived using asymptotic equivalence results (see Chapters 7 and 8).

2. For any m-accretive operator on a uniformly convex Banach space with Fréchet differentiable norm there is weak convergence in the following cases (see [95]):

(a) $\{\lambda_n\}$ does not converge to zero, or

(b) The modulus of convexity of X 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, as defined in [90].

3.1.2 Strong Convergence

We have the following (see [33]):

Proposition 3.3 *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. ■

Proposition 3.4 (Strong monotonicity) *If A is α-strongly monotone for some $\alpha > 0$ then $\{x_n\}$ converges strongly to the unique $x^* \in \mathcal{S}$ as $n \to \infty$.*

Proof. Strong monotonicity implies $\alpha\|x_n - x^*\| \leq \|y_n\| \to 0$ as $n \to \infty$. ■

Proposition 3.5 (Nonempty interior) *Let A be maximal monotone with int $\mathcal{S} \neq \emptyset$. Then $\{x_n\}$ converges strongly as $n \to \infty$.*

Proof. First recall that the sequence $\{x_n\}$ is bounded and that $\|x_n - y\|$ decreases as n increases for each $y \in \mathcal{S}$. Take $x^* \in \text{int}\mathcal{S}$ so that $B(x^*, r) \subset \mathcal{S}$ for some r. For all $h \in B(0, 1)$ we have $\langle x_{n-1} - x_n, x_n - x^* - rh \rangle \geq 0$ and so $r\|x_{n-1} - x_n\| \leq \langle x_{n-1} - x_n, x_n - x^* \rangle$. Finally,

$$\begin{aligned} r\ \|x_n - x_m\| &\leq r \sum_{k=n+1}^{m} \|x_{k-1} - x_k\| \\ &\leq \sum_{k=n+1}^{m} \langle x_{k-1} - x_k, x_k - x^* \rangle - \|x_k - x^*\|^2 \\ &\leq C\Big(\|x_n - x^*\| - \|x_m - x^*\|\Big) \end{aligned}$$

for some constant C. Since $\|x_n - x^*\|$ is convergent, x_n is a Cauchy sequence. ■

A fairly general result in Banach spaces is the following, from [87]:

Theorem 3.6 *Let X, X^* be uniformly convex. Assume A is m-accretive and satisfies the Nevanlinna-Reich convergence condition (see Chapter 1, Subsection 1.3.3). If $\{\lambda_n\} \notin \ell^1$ then $\{x_n\}$ converges strongly as $n \to \infty$.*

Proof. Let $j_n = \mathcal{J}(x_n - Px_n)$, where P is a projection onto $\mathcal{S}$. We have

$$\begin{aligned}
\|x_{n+1} - Px_{n+1}\|^2 + \lambda_{n+1}\langle y_{n+1}, j_{n+1}\rangle &= \langle x_n - Px_{n+1}, j_{n+1}\rangle \\
&= \langle x_n - Px_n, j_{n+1}\rangle + \langle Px_n - Px_{n+1}, j_{n+1}\rangle \\
&\leq \|x_n - Px_n\| \, \|x_{n+1} - Px_{n+1}\| \\
&\leq \frac{1}{2}\left[\|x_n - Px_n\|^2 + \|x_{n+1} - Px_{n+1}\|^2\right].
\end{aligned}$$

Thus $\|x_{n+1}-Px_{n+1}\|^2+2\lambda_{n+1}\langle y_{n+1}, j_{n+1}\rangle \leq \|x_n-Px_n\|^2$ and $\sum_{n=1}^{\infty} \lambda_{n+1}\langle y_{n+1}, j_{n+1}\rangle < \infty$. Since $\{\lambda_n\} \notin \ell^1$ and $\langle y_{n+1}, j_{n+1}\rangle \geq 0$ 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 $\liminf_{n\to\infty}\|x_n-Px_n\| = 0$. But $\|x_n - Px_n\|$ is nonincreasing, so it must converge to 0. On the other hand, the sequence $\|x_n-p\|$ is nonincreasing for each $p \in \mathcal{S}$. In particular, $\|x_{n+m}-Px_n\| \leq \|x_n-Px_n\|$ and therefore $\|x_{n+m} - x_n\| \leq 2\|x_n - Px_n\|$. We conclude that x_n converges strongly to some $p \in \mathcal{S}$ as $n \to \infty$. ■

Extensions: In general Banach spaces the previous result remains true provided $\mathcal{S}$ is proximinal and convex and that A is accretive in the sense of Browder.

3.1.3 Almost-convergence.

The following result is from [68]:

Proposition 3.7 *Let X be a uniformly convex Banach space and assume $\mathcal{S} \neq \emptyset$.*

i) If the norm of X is Fréchet-differentiable or X satisfies Opial's condition and if $\{\lambda_n\} \in \ell^2 \setminus \ell^1$ then x_n almost-converges weakly as $n \to \infty$ to an element of $\mathcal{S}$.

ii) If $\lim\limits_{n\to\infty}\|x_n - x_{n+k}\| = \rho(k)$ exists uniformly in $k \in \mathbf{N}$ then x_n almost-converges strongly as $n \to \infty$ to an element of $\mathcal{S}$.

Remark 3.8 *If the norm of X is Fréchet-differentiable, it suffices to require $\{\lambda_n\} \notin \ell^1$ for part i), as had been proved earlier in [40].*

3.1.4 Convergence in average

The following result was presented in [78]:

Proposition 3.9 *Let $\mathcal{S} \neq \emptyset$.*

1. *If $\{\lambda_n\} \in \ell^2 \setminus \ell^1$ then x_n converges weakly in average to a point in $\mathcal{S}$; and*
2. *If A is odd and $\{\lambda_n\} \notin \ell^1$, then x_n converges strongly in average to a point in $\mathcal{S}$.*

Proof. We check the conditions of Lemma 1.7 with $F = \mathcal{S}$. For i) take $x^* \in \mathcal{S}$ and observe that

$$\|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 + \lambda_n^2 C$$

for some constant C. Since $\{\lambda_n\} \in \ell^2$, $\|x_n - x^*\|^2$ must be convergent. For ii), take $k \geq m$, denote the average by $\overline{x}_k$ and write

$$\overline{x}_k = \tfrac{1}{\sigma_k}\sum_{j=1}^{m-1}\lambda_j x_j + \left(\tfrac{\sigma_k - \sigma_m}{\sigma_k}\right)\tfrac{1}{(\sigma_k-\sigma_m)}\sum_{j=m}^{k}\lambda_j x_j.$$

The first term tends to 0 as $k \to \infty$ while the second term is $\frac{\sigma_k - \sigma_m}{\sigma_k}$ times a term in C_m for each $k \geq m$. For iii), take $[u, v] \in A$ and notice that $u = (I + \lambda_n A)^{-1}(u + \lambda_n v)$. But also $x_{n+1} = (I + \lambda_n A)^{-1} x_n$ and so

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

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

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

It is clear that if $\overline{x}_n \rightharpoonup \overline{x}$ then $\overline{x} \in \mathcal{S}$. Thus Lemma 1.7 gives the first part. ■

Strong convergence in average for A odd had already been proved in [16] in the case where $\lambda_n \equiv \lambda$.

If the sequence $\{\lambda_n\}$ of stepsizes is in ℓ^1 ($\sigma_n \to \sigma < \infty$), the trajectory $\{x_n\}$ always converges strongly to a point x_∞ (observe that $\|x_n - x_m\| \leq \|y_1\| \cdot |\sigma_n - \sigma_m|$), no matter whether the function has minimizers. Even when the minimizing set is nonempty, if the distance between the initial point x_0 and the minimizing set $\mathcal{S}$ is greater than $\sigma\|y_1\|$, then x_∞ cannot be a point in $\mathcal{S}$. However, one would like to know whether or not it is possible to attain the minimizing set if the initial condition is close enough to $\mathcal{S}$ (alternatively, if σ is large enough). We give a partial answer in terms of the smoothness of the objective function around $\mathcal{S}$.

In the sequel we assume $\mathcal{S}$ is nonempty. The following example includes the case where f is a differentiable function having a L-Lipschitz gradient in some η-neighborhood of $\mathcal{S}$ (consider the maximal monotone operator $A = \nabla f$):

Example 3 *Let A be a maximal monotone operator on H. Set $\mathcal{S} = A^{-1}0$ and let $d(u, \mathcal{S})$ denote the distance from u to $\mathcal{S}$. Assume there exist $\eta > 0$ and $L > 0$ such that if $d(u, \mathcal{S}) < \eta$, then $\sup_{v \in Au}\|v\| \leq L\ d(u, \mathcal{S})$. It is clear that the minimizing set cannot be attained in a finite number of steps. If $d(x_N, \mathcal{S}) < \eta$ for some N, the proximal iteration yields $\|x_n - x_{n-1}\| \leq \lambda_n L d(x_n, \mathcal{S})$ for all $n > N$. Since $\|x_n - x_{n-1}\| \geq d(x_{n-1}, \mathcal{S}) - d(x_n, \mathcal{S})$ we have $(1 + \lambda_n L) d(x_n, \mathcal{S}) \geq d(x_{n-1}, \mathcal{S})$. Therefore*

$$d(x_n, \mathcal{S}) \geq \left[\prod_{k=N+1}^{n} (1 + \lambda_k L)^{-1}\right] d(x_N, \mathcal{S}) \geq e^{-\sigma_n L} d(x_N, \mathcal{S}).$$

If $\sigma_n \to \sigma < \infty$, the sequence $\{x_n\}$ stays away from $\mathcal{S}$.

On the other hand, if the function f is somehow *pointed* on $\partial\mathcal{S}$ the sequence $\{x_n\}$ will converge if σ is large enough. To see this, let P denote the projection onto the nonempty closed convex set $\mathcal{S}$ and take $r > 0$. Notice that the following statementes are equivalent:

((1)) $\|v\| \geq r$ for all $v \in \partial f(x)$ such that $x \notin \mathcal{S}$;

((2)) $f(x) \geq f^* + r\|x - Px\|$ for all $x \in D$; and

((3)) $r(x - Px)\|x - Px\|^{-1} \in \partial f(Px)$ for all $x \notin \mathcal{S}$.

Proposition 3.10 *Let $f_n \equiv f$ and $\{\varepsilon_n\} \in \ell^1$. Assume there is $r > 0$ such that condition ((1)) holds and let $\sigma_n \to \sigma < \infty$.*

i) *If* $r^2\sigma > f(x_0) - f^* + \sum \varepsilon_k$*, then there are* $N \in \mathbf{N}$ *and* $x^* \in \mathcal{S}$ *such that* $x_n = x^*$ *for all* $n \geq N$*.*

ii) *If* $r^2\sigma = f(x_0) - f^* + \sum \varepsilon_k$*, then* $\lim_{n\to\infty} d(x_n, \mathcal{S}) = 0$*. Moreover, if the sequence* $\{\|y_n\|\}$ *turns out to be bounded,*[1] *then* x_n *converges to a point in* $\mathcal{S}$ *as* $n \to \infty$*.*

Proof. If $x_k \notin \mathcal{S}$ for $k = 1, \ldots, n$, then $f(x_{k-1}) - f(x_k) + \varepsilon_k \geq \lambda_k \|y_k\|^2 \geq r^2\lambda_k$. Summing up one gets $r^2\sigma_n \leq f(x_0) - f(x_n) + \sum \varepsilon_k$.

i) If $x_n \notin \mathcal{S}$ for all $n \in \mathbf{N}$ then $r^2\sigma \leq f(x_0) - f^* + \sum \varepsilon_k$, which is a contradiction.

ii) Clearly $\lim_{n\to\infty} f(x_n) = f^*$. By ((2)) we get $\lim_{n\to\infty} \|x_n - Px_n\| = 0$. Finally, if $\{\|y_n\|\}$ is bounded and $\sigma < \infty$ then $\{x_n\}$ is a Cauchy sequence. ■

Condition ((2)) above was used by Ferris in [51] to prove convergence in a finite number of steps. In the particular case where $\mathcal{S} = \{x^*\}$, the form ((3)) simplifies to read $\overline{B}(0, r) \subseteq \partial f(x^*)$, which is the assumption used by Rockafellar in [97] with the same purpose in that specific case. In the cited works, H is assumed to be $\mathbf{R}^n$ and the sequence $\{\lambda_n\}$ to be bounded from below by a positive constant, which is a more restrictive setting. Although these conditions seem rather restrictive we can easily find a function for which the proximal sequence stays away from $\mathcal{S}$.

Example 4 *Let* H *be separable and let* $\{e_k\}$ *be an orthonormal basis for* H*. Define* $H_n = span\{e_1, \ldots, e_n\}$ *and consider the function*

$$f(x) = \begin{cases} \sum \frac{1}{k} x_k & \text{if } \; x^k = \langle x, e_k\rangle \geq 0 \; \text{ for all } \; k \\ \infty & \text{otherwise.} \end{cases}$$

This function satisfies condition ((1)) on each H_n *with* $r = 1/n$ *and the union* $\cup H_n$ *is dense in* H *(so* f *does not satisfy condition ((1)) in all of* H *but almost). If we apply the proximal point algorithm with* $x_0 \in dom(f)$ *and* $\sigma_n \to \sigma$ *we find that the k-th component of the term* x_n *satisfies*

$$(x_n)^k \geq (x_{n-1})^k - \lambda_n/k \geq (x_0)^k - \sigma_n/k$$

and so

$$\liminf_{n\to\infty} (x_n)^k \geq (x_0)_k - \sigma/k$$

for each k*. If* x_0 *is selected so that* $\sup_k k(x_0)^k = \infty$ *we will have* $\liminf_{n\to\infty}(x_n)^k > 0$ *for some* k *no matter how large* σ *is. Therefore* x_n *does not converge to* 0*, which is the unique minimizer of* f*. Observe also that one can select* x_0 *arbitrarily close to* 0*, with the same result.*

3.2 The differential inclusion

In this section we present the main results concerning the asymptotic behavior of the solution u to the differential inclusion (2.7):

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

[1] For instance, if $\sum(\varepsilon_n + \varepsilon_{n-1})\lambda_n^{-1} < \infty$.

3.2.1 Weak Convergence

We present two conditions that guarantee weak convergence of the trajectories $u(t) = S(t)x$ for $x \in \overline{D(A)}$ and give a characterization for the weak limit. Unless otherwise stated the underlying space is Hilbert.

Demipositive Operators.

The classical weak convergence result that we present below was proved by Bruck in [36].

Theorem 3.11 *Let A be a demipositive operator and suppose $u : [0, \infty) \to H$ is the function given by Theorem 2.4. Then $u(t)$ converges weakly as $t \to \infty$ to an element of $\mathcal{S}$.*

Proof. We shall use Opial's lemma. First notice that for all $y \in \mathcal{S}$, the function $t \mapsto \frac{1}{2}\|u(t) - y\|^2 = \theta_y(t)$ is decreasing, thus $\|u(t) - y\|$ has a limit as $t \to \infty$ for all $y \in \mathcal{S}$. It remains to prove that every weak cluster point of $u(t)$ lies in $\mathcal{S}$. Let u_0 be as in the definition of demipositivity and let $u(t_n)$ converge weakly to u as $n \to \infty$. Set $v_n = u'(t_n)$. The sequence v_n is bounded. Define $h(t) = -\theta'_{u_0}(t)$ so that $h(t_n) = \langle v_n, u(t_n) - u_0\rangle$. Since θ_{u_0} is bounded, $h \in L^1$ and there is a subsequence t_{n_k} of t_n such that $h(t_{n_k}) \to 0$ as $k \to \infty$. But $u(t_{n_k}) \rightharpoonup u$, so $u \in \mathcal{S}$ by demipositivity. ■

Counterexample: The counterclockwise $\pi/2$-rotation $A(x, y) = (-y, x)$ in $\mathbf{R}^2$ is maximal monotone but not demipositive and the solutions to $u' \in -Au$ do not converge.

Extensions: The theorem was extended by G. Passty in [90] to the class of *φ-demipositive* operators.

Pazy's $\mathcal{L}$ condition.

Let A be a maximal monotone operator and S the semigroup generated by A. Denote by a^0 the element of minimal norm in $\overline{R(A)}$. A satisfies *Pazy's $\mathcal{L}$ condition* if

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

for every $h > 0$ and $x \in D(A)$. Alternatively, if for every $x \in D(A)$ one has

$$\lim_{t\to\infty} A^0 S(t)x = a^0.$$

As we shall se below in Proposition 3.26, a^0 is the only possible value for $\lim\limits_{t\to\infty} A^0 S(t)x$ whenever it exists. Pazy's $\mathcal{L}$ condition was presented in [91], where the reader will find the proof of Theorem 3.12 below and the following examples of operators satisfying the condition $\mathcal{L}$:

1. $A = \partial f$, where ϕ is a proper lower-semicontinuous convex function;
2. $A = \rho(I - T)$ with T nonexpansive and $\rho > 0$. In particular, the Yosida approximation of any maximal monotone operator.

An interesting task is to investigate the relationship between demipositivity and Pazy's $\mathcal{L}$ condition.

Theorem 3.12 *Let A satisfy Pazy's $\mathcal{L}$ condition and take $x \in \overline{D(A)}$.*

i) If $\mathcal{S} = \emptyset$ then $\|S(t)x\| \to \infty$ as $t \to \infty$; and

ii) If $\mathcal{S} \neq \emptyset$ then there is $p \in \mathcal{S}$ such that $S(t)x \rightharpoonup p$ as $t \to \infty$.

Proof. We only need to prove that every weak cluster point of the trajectory $S(t)x$ lies in $\mathcal{S}$. The rest follows as in Theorem 3.11. If $x \in D(A)$ we have $\lim_{t\to\infty} A^0 S(t)x = 0$ so the result follows from the weak-strong closedness of the graph of A. If $x \in \overline{D(A)}$ and $S(t_k)x \to y$, take a sequence $\{x_n\}$ in $D(A)$ such that $x_n \to x$. For fixed n there is a subsequence $\{t_{k_j}\}$ of $\{t_k\}$ such that $S(t_{k_j})x_n \rightharpoonup p_n \in \mathcal{S}$ as $j \to \infty$. The weak lower-semicontinuity of the norm implies

$$\|y - p_n\| \leq \liminf_{j\to\infty} \|S(t_{k_j})x - S(t_{k_j})x_n\| \leq \|x - x_n\|.$$

Since $\mathcal{S}$ is closed, $y \in \mathcal{S}$. ■

Remarks on the subdifferential case.

Proposition 3.13 *Let $A = \partial f$ with $f \in \Gamma_0(H)$. Then*

i) The function $t \mapsto f(u(t))$ is decreasing;

ii) $\lim_{t\to\infty} f(u(t)) = \inf_H f$;

Proof. Both assertions follow easily from the following estimation:

$$f(u(t)) + \frac{\|u(t) - u\|^2}{2t} \leq f(u) + \frac{\|x - u\|^2}{2t}, \tag{3.2}$$

which holds for every $u \in H$, $t > 0$, and is obtained by integrating the subdifferential inequality. ■

Remark 3.14 *Inequality (3.2) shows that $f(u(t))$ 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 [57]).*

3.2.2 Strong Convergence

Even in a Hilbert space H and 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 celebrated counterexample in [17]: the author defines a function $\phi \in \Gamma_0(\ell^2)$ having minimizers and proves that the trajectories converge weakly but not strongly. However, there is strong convergence in some cases:

Proposition 3.15 (Strong monotonicity I) *If A is α-strongly monotone for some $\alpha > 0$ then $u(t)$ converges strongly to the unique $x^* \in \mathcal{S}$ as $t \to \infty$.*

Proof. Strong monotonicity implies

$$\frac{1}{2}\frac{d}{dt}\|u(t)-v(t)\|^2 = \langle u'(t)-v'(t), u(t)-v(t)\rangle \leq -\alpha\|u(t)-v(t)\|^2$$

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

Extensions: 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 [87] the authors prove that if A satisfies Nevanlinna-Reich convergence condition then $u(t)$ converges strongly to some $x^* \in \mathcal{S}$ as $t \to \infty$. If X is not uniformly convex but X^* is, the result remains true provided Ax is proximinal and convex for every x (see [41]). If neither X nor X^* is uniformly convex, the result is still true if the semigroup is differentiable (see [87]).

Proposition 3.16 (Strong monotonicity II) *Let $\mathcal{S} \neq \emptyset$ and suppose there is $\alpha > 0$ such that for every $x \in D(A)$ we have*

$$\langle A^0 x, x - Px\rangle \geq \alpha\|x - Px\|^2.$$

Then $u(t)$ converges strongly to the unique $x^* \in \mathcal{S}$ as $t \to \infty$.

Proposition 3.17 (Nonempty interior) *Assume $\operatorname{int}\mathcal{S} \neq \emptyset$. Then $u(t)$ converges strongly to some $x^* \in \mathcal{S}$ as $t \to \infty$.*

Proof. Let $x_0 \in \operatorname{int}\mathcal{S}$ so that $B(x_0, r) \subset \mathcal{S}$ for some r. For all $h \in B(0,1)$ we have $\langle -u'(t), u(t) - x_0 - rh\rangle \geq 0$ and so $\langle -u'(t), u(t)-x_0\rangle \geq -r\langle u'(t), h\rangle$. Hence $r\|u'(t)\| \leq -\langle u'(t), u(t) - x_0\rangle$. Finally

$$\begin{aligned} r\|u(t)-u(s)\| &\leq r\int_s^t \|u'(\tau)\|\, d\tau \\ &\leq -\int_s^t \langle u'(\tau), u(\tau)-x_0\rangle\, d\tau \\ &\leq \frac{1}{2}\|u(s)-x_0\|^2 - \frac{1}{2}\|u(t)-x_0\|^2. \end{aligned}$$

Since $\|u(t)-x_0\|$ is convergent, $u(t)$ has the Cauchy property and converges as well. ■

Extensions: Theorem 4 in [87] 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 [41].

Proposition 3.18 (Pazy's $\mathcal{L}$ condition) *Let A satisfy Pazy's $\mathcal{L}$ condition. Take $x \in \overline{D(A)}$ and assume $0 \in \mathcal{S}$. If the function $t \mapsto \|S(t+h)x + S(t)x\|$ is nonincreasing for each $h \geq 0$ then $S(t)x$ converges strongly to some $p \in \mathcal{S}$ as $t \to \infty$.*

Proof. Since $\frac{1}{2}\frac{d}{dt}\|S(t)x\|^2 = -\langle A^0 S(t)x, S(t)x\rangle \leq 0$ we can define $d = \lim_{t\to\infty}\|S(t)x\|$. Hence

$$2d = 2\lim_{t\to\infty}\|S(t)x\| \leq \lim_{t\to\infty}\Big[\|S(t)x + S(t+s)x\| + \|S(t)x - S(t+s)x\|\Big]$$

for each s. By condition $\mathcal{L}$ we have $2d \leq \lim_{t\to\infty} \|S(t)x + S(t+s)x\|$ and by hypothesis we get $2d \leq \|S(t)x + S(t+s)x\|$ for each t and s. Finally,

$$\|S(t+s)x - S(t)\| \leq 4\|S(t)x\|^2 - 4d^2$$

so $\{S(t)x\}$ is a Cauchy net. ■

The hypotheses in the previous proposition from [91] hold if A satisfies Pazy's $\mathcal{L}$ condition and the minimal section A^0 is odd. In particular, if $A = \partial f$ where $f \in \Gamma_0(H)$ is even. In this case there is an easy independent proof, that we include for completeness:

Proof. Set $u(t) = S(t)x$. 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

$$\gamma'(t) = \langle u'(t), u(t) + u(s)\rangle \leq f(-u(s)) - f(u(t)) = f(u(s)) - f(u(t)) \leq 0.$$

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

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

But $\|u(t)\|$ converges as $t \to \infty$ because $0 \in \operatorname{Argmin} f$ and so $u(t)$ converges too. ■

The strong ω-limit set from a point x is defined by $\omega(x) = \bigcap_{t>0} \overline{\{S(s)x : s \geq t\}}$.

Proposition 3.19 (Nonempty strong ω-limit sets) *Let A satisfy Pazy's $\mathcal{L}$ condition and take $x \in \overline{D(A)}$. The following are equivalent:*

i) $S(t)x$ converges strongly;

ii) $\omega(x) = \{x^*\}$ and $S(t)x$ converges weakly to x^*;

iii) $\omega(x) \cap \mathcal{S} \neq \emptyset$.

Proof. Clearly *i*) implies *ii*). If *ii*) holds we have $\emptyset \neq \omega(x) \subset \mathcal{S}$ by Theorem 3.12. Finally suppose *iii*) holds. Let $S(t_n)x \to p \in \mathcal{S}$. If $t \geq t_n$ we have $\|S(t)x - p\| \leq \|S(t_n)x - p\|$, thus $S(t)x \to p$ as $t \to \infty$. ■

Corollary 3.20 *In Hilbert space, assume either that A is demipositive or that A satisfies Pazy's $\mathcal{L}$ condition and $\mathcal{S} \neq \emptyset$. If A has a compact resolvent, then for each $x \in \overline{D(A)}$, $S(t)x$ converges strongly to some $p \in \mathcal{S}$. This holds, in particular, if $A = \partial f$ and the set $\{\ u \in H \mid f(u) + \|u\|^2 \leq M\ \}$ is compact for every $M \geq 0$.*

Proof. $\omega(x) \neq \emptyset$ for all $x \in \overline{D(A)}$ if $\mathcal{S} \neq \emptyset$ and A has compact resolvent. Weak convergence is guaranteed by Theorems 3.11 and 3.12 respectively. The result follows from Proposition 3.19. ■

The case $A = \partial f$ was studied in [29].

Remark 3.21 *If $\mathcal{S}$ has nonempty interior we have $\omega(x) \neq \emptyset$ for all $x \in \overline{D(A)}$ and A is demipositive. Therefore, Proposition 3.17 can also be deduced from Proposition 3.19.*

3.2.3 Almost-convergence

The following result is from [84]:

Proposition 3.22 *Let X be a uniformly convex Banach space and assume $\mathcal{S} \neq \emptyset$.*

i) If the norm of X is Fréchet-differentiable or X satisfies Opial's condition then $u(t)$ almost-converges weakly as $t \to \infty$.

ii) If $\lim_{t\to\infty} \|u(t+h) - u(t)\| = \rho(h)$ exists uniformly in $h \geq 0$, then $u(t)$ almost-converges strongly as $t \to \infty$.

The second part had already been presented by the same authors in [68]. The remarkable improvement in [84] is that they prove the result not only for the trajectories, but for *almost-orbits*, which are approximate trajectories in a sense to be precised in Section 3.4.2.

3.2.4 Convergence in average

Let A be a maximal monotone operator on a Hilbert space H. Take $x \in \overline{D(A)}$ and 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 [19]. We first prove that weak cluster points of $\sigma(t)x$ are in $\mathcal{S}$; next, that the projection $P_{\mathcal{S}}S(t)x$ converges weakly to some u, which is the only weak cluster point of $\sigma(t)x$.

Lemma 3.23 *Every weak cluster point of $\sigma(t)x$ lies in $\mathcal{S}$.*

Proof. Indeed, assume $\sigma(t_k)x \rightharpoonup u$ as $k \to \infty$ and set $u(t) = S(t)x$. For any $v \in D(A)$ we have

$$2\int_0^{t_k} \langle u(t) - v, u'(t)\rangle dt = \|u(t_k) - v\|^2 - \|x - v\|^2.$$

Now take $w \in A(v)$, so that $\langle u(t)-v, -u'(t)-w\rangle \geq 0$ and $\langle u(t)-v, -w\rangle \geq \langle u(t)-v, u'(t)\rangle$. 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 $(v,w) \in G_r(A)$, so $0 \in Au$ by maximality. ■

This implies that if $\mathcal{S} = \emptyset$ then $\|\sigma(t)x\| \to \infty$ for every $x \in \overline{D(A)}$ as $t \to \infty$ (and also $\|u(t)\| \to \infty$). 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 $x \in \overline{D(A)}$.

Theorem 3.24 *If $\mathcal{S} \neq \emptyset$ then $\sigma(t)x$ converges weakly to $u = \lim_{t\to\infty} P_{\mathcal{S}}S(t)x$.*

[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 $\int_0^\infty |da_n| \to 0$.

Proof. Define $v(t) = P_{\mathcal{S}}S(t)x$. We first prove that $v(t)$ converges to a point $u \in \mathcal{S}$ as $t \to \infty$. To verify this, 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

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

Since $\mathcal{S}$ is convex, we have $\left\|\frac{v(t+h)+v(t)}{2} - S(t+h)x\right\| \geq \psi(t+h)$. 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 limit as $t \to \infty$.

It only remains to prove that $u = \lim_{t\to\infty} v(t)$ is the only weak cluster point of $\sigma(t)x$ as $t \to \infty$. Now $v(t)$ is a projection onto $\mathcal{S}$, so $\langle S(t)x - v(t), z - v(t)\rangle \leq 0$ for all $z \in \mathcal{S}$. Hence

$$\langle S(t)x - v(t), z - u\rangle \leq \langle S(t)x - v(t), v(t) - u\rangle \leq M\|v(t) - u\|$$

where $M = \sup_{t\geq 0}\psi(t)$. Integrating from 0 to T and dividing by T one gets

$$\left\langle \sigma_T(x) - \frac{1}{T}\int_0^T v(t)dt, z - u\right\rangle \leq \frac{M}{T}\int_0^T \|v(t) - u\|dt.$$

By passing to the limit we see that any weak cluster point, say w, of $\sigma(t)x$, must satisfy $\langle w - u, z - u\rangle \leq 0$ for all $z \in \mathcal{S}$. In particular for $z = w$, which gives $w = u$. ■

Extensions: Weak convergence in average is still true in uniformly convex Banach space with Fréchet-differentiable norm (see [93]) or satisfying Opial's condition (see [59]). The almost-convergence is strong if S is odd (see [15]).

If A is a subdifferential in Hilbert space, convergence in average guarantees the convergence of the trajectory (see [36]):

Proposition 3.25 *If $A = \partial f$ then $\lim_{t\to\infty}\left\|u(t) - \frac{1}{t}\int_0^t u(s)\ ds\right\| = 0$.*

Proof. Integration by parts gives $u(t) - \frac{1}{t}\int_0^t u(s)\ ds = \frac{1}{t}\int_0^t su'(s)\ ds$. Since $\|u'(t)\|$ is decreasing one has $\int_{t/2}^t s\|u'(s)\|^2\ ds \geq \|u'(t)\|^2\int_{t/2}^t s\ ds = \frac{3}{8}t^2\|u'(t)\|^2$. But since $A = \partial f$, the function $t \mapsto t\|u'(t)\|^2$ is in $L^1(0,\infty)$ (see [28]) so we have $\lim_{t\to\infty} t\|u'(t)\| = 0$ and the result follows. ■

Back to general operators, we also have the following result for the average velocity (see [91]):

Proposition 3.26 *Let a^0 be the least-norm element in $\overline{R(A)}$. For every $x \in \overline{D(A)}$ we have*

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t A^0S(\tau)x\ d\tau = \lim_{t\to\infty}\left(\frac{-S(t)x}{t}\right) = a^0.$$

Proof. Recall that $\int_0^t A^0 S(\tau)x\ d\tau = x - S(t)x$. The first equality follows immediately. For the second one, take a sequence $\{x_n\}$ such that $A^0 x_n \to a^0$ as $n \to \infty$. We have

$$\begin{aligned}
\left\| \frac{1}{t}\int_0^t A^0 S(\tau)x\ d\tau - a^0 \right\| &\leq \frac{1}{t}\left\| \int_0^t \left(A^0 S(\tau)x - A^0 S(\tau)x_n\right) + \left(A^0 S(\tau)x_n - a^0\right)\ d\tau \right\| \\
&\leq \frac{1}{t}\|x - S(t)x - x_n + S(t)x_n\| + \frac{1}{t}\int_0^t \left\|A^0 S(\tau)x_n - a^0\right\|\ d\tau \\
&\leq \frac{2}{t}\|x - x_n\| + \sup_{t\geq 0}\|A^0 S(t)x_n - a^0\|.
\end{aligned}$$

Notice that

$$\|A^0 S(t)x_n - a^0\|^2 \leq \|A^0 S(t)x_n\|^2 - \|a^0\|^2 \leq \|A^0 x_n\|^2 - \|a^0\|^2.$$

Given $\varepsilon > 0$ we can find $N \in \mathbf{N}$ such that $\sup_{t\geq 0}\|A^0 S(t)x_N - a^0\| < \varepsilon/2$. Finally, take $T > 0$ such that $2\|x - x_N\|/T < \varepsilon/2$ for all $t \geq T$. ∎

3.3 Euler's discretization

3.3.1 Weak Convergence

The following result from [41] works for demipositive operators in "a few" Banach spaces, namely $X = L^{2m}$, $m \in \mathbf{N}$ or $X = \ell^p$, $p \in (1,\infty)$. We state it here for Hilbert space, where it has a simpler form:

Proposition 3.27 *Let A be demipositive and $\{\lambda_n\} \in \ell^2 \setminus \ell^1$. If $z_n^* = (z_n - z_{n+1})/\lambda_n \in Az_n$ is bounded then z_n converges weakly to a zero of A.*

Proof. Let $y \in \mathcal{S}$. Since

$$\|z_{n+1} - y\|^2 \leq \|z_n - y\|^2 + \lambda_n^2\|z_n^*\|^2 \leq \|z_{n+1} - y\|^2 + \lambda_n^2 C$$

for some C, and $\{\lambda_n\}$ is in ℓ^2, the sequence $\|z_n - y\|$ is convergent. On the other hand, we have

$$\lambda_n\langle z_n^*, z_n - y\rangle \leq \lambda_n^2\|z_n^*\|^2 + \|z_n - y\|^2 - \|z_{n+1} - y\|^2.$$

Thus $\sum\limits_{n\geq 1}\lambda_n\langle z_n^*, z_n - y\rangle < \infty$. One concludes as in part ii) of Theorem 3.1 using an analogue of Lemma 3.2. ∎

In [38] the author had already proved a similar result by first establishing Proposition 3.30 below and then requiring asymptotic regularity in order to use Remark 1.5. His result is different from the one presented above: the hypothesis on the operator A is more restrictive ($A = I - T$); the hypothesis on the sequence z_n is less restrictive (only asymptotic regularity is needed); and the hypothesis on the stepsizes cannot be compared.

Extension: Let X be uniformly convex with Fréchet-differentiable norm, T nonexpansive having a fixed point, $A = I - T$ and $\{\lambda_n\}$ satisfying $0 \leq \lambda_n \leq 1$ and $\sum \lambda_n(1 - \lambda_n) = \infty$. Then $\{z_n\}$ converges weakly to a fixed point of T (see [95]).

3.3.2 Strong Convergence

The following holds if X and X^* are uniformly convex. We state it in Hilbert space where it has a simple form:

Theorem 3.28 *Let A be m-accretive and $\{\lambda_n\} \in \ell^2 \setminus \ell^1$. Assume $y_n = (z_n - z_{n+1})/\lambda_n$ is bounded. If A satisfies the Nevanlinna-Reich convergence condition (see Chapter 1, subsection 1.3.3), then $\{z_n\}$ converges strongly as $n \to \infty$.*

Proof. To simplify notation, write $P = P_{\mathcal{S}}$ and $j_n = x_n - Px_n$. We have

$$\|j_{n+1}\|^2 \leq \|x_{n+1} - Px_n\|^2 = \|j_n - \lambda_n y_n\|^2 = \|j_n\|^2 - 2\lambda_n \langle y_n, j_n \rangle + \lambda_n^2 \|y_n\|^2.$$

Since $\{y_n\}$ is bounded, so is $\{j_n\}$. Moreover,

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

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

Extensions: According to [41], 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 [87], the conclusion of Theorem 3.28 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.

Gradient's case

Let f be a convex function in $\mathcal{C}^1(H)$ and define the sequence $\{u_n\}$ by

$$u_{n+1} = u_n - \lambda_n \nabla f(u_n).$$

Theorem 3.29 *Set $\mu_n = \|u_{n+1} - u_n\|$. If $\{\mu_n\} \in \ell^2$ and $\{\lambda_n\} \notin \ell^1$ then u_n converges weakly to some $u^* \in \mathcal{S}$.*

Proof. Let $u \in \mathcal{S}$ and set $\theta_k = \frac{1}{2}\|u_k - u\|^2$. We have

$$\begin{aligned} \theta_{k+1} - \theta_k &= \tfrac{1}{2}\|u_{k+1} - u_k\|^2 + \langle u_{k+1} - u_k, u_k - u \rangle \\ &\leq \tfrac{1}{2}\|u_{k+1} - u_k\|^2 + \lambda_k (f(u) - f(u_k)) \\ &= \tfrac{1}{2}\mu_k^2 + \lambda_k (f^* - f(u_k)) \\ &\leq \tfrac{1}{2}\mu_k^2 \end{aligned}$$

But $\{\mu_k\} \in \ell^2$ and θ_k is bounded below, so θ_k must converge. By Opial's Lemma, convergence is guaranteed if every cluster point of the sequence $\{u_k\}$ lies in $\mathcal{S}$. Notice that

$$\sum_{k=1}^{\infty} \lambda_k (f(u_k) - f^*) \leq \frac{1}{2} \sum_{k=1}^{\infty} \mu_k^2 + \theta_0 < \infty.$$

Since $\{\lambda_n\} \notin \ell^1$, we must have $\liminf_{n\to\infty} f(u_n) = f^*$. Let $u_{k_n} \rightharpoonup \bar{u}$. Lower-semicontinuity implies $f(\bar{u}) \leq \liminf_{n\to\infty} f(u_n) = f^*$ and so $\bar{u} \in \mathcal{S}$. ■

If $A = I - T$, where T is a strict contraction ($\|Tx - Ty\| \leq k\|x - y\|$ with $k < 1$) on any Banach space, it is easy to prove that for all $x \in X$ the sequence $T^n x$ converges strongly to the unique fixed point of T (Banach's Fixed-Point Theorem).

3.3.3 Almost-convergence.

The basic result for nonexpansive mappings in Hilbert space is the following (see [38]):

Proposition 3.30 *Let C be a closed convex subset of a Hilbert space H and let T be a nonexpansive mapping on C having a fixed point. Then for each $x \in C$ the sequence $T^n x$ almost-converges weakly to some $c \in \mathcal{S}$. Moreover, c can be characterized as the asymptotic center in the sense of Edelstein.*

Extensions: Weak almost-convergence to a fixed-point of T remains true in uniformly convex Banach spaces in the following cases: X satisfies Opial's condition or C is compact (see [59]); X has a weakly sequentially continuous duality map and $C = X$ ([20]); the norm of X is Fréchet-differentiable ([39, 95]); or X is superreflexive ([18]).

For the strong almost-convergence we have a result from [68]:

Proposition 3.31 *Let X be a uniformly convex Banach space and $C \subseteq X$ be closed and convex. Suppose T is a nonexpansive mapping having a fixed point. If $\lim_{n\to\infty} \|T^n x - T^{n+k}x\| = \rho(k)$ exists uniformly in $k \in \mathbf{N}$ then $T^n x$ almost-converges strongly as $n \to \infty$ to a point in $\mathcal{S}$.*

A similar result had been already proved in [38] for asymptotically isometric operators in Hilbert space. A nonexpansive mapping T is *asymptotically isometric* on a set D if the limit $\lim_{n\to\infty} \|T^{n+k}x - T^n y\|$ exists for each $x, y \in D$, uniformly in $k \in \mathbf{N}$.

3.3.4 Convergence in average.

For nonexpansive mappings weak convergence in average was estalished in [14]:

Proposition 3.32 *Let T be a nonexpansive mapping on a bounded closed convex subset C of a Hilbert space H and set $A = I - T$ and $\lambda_n \equiv 1$. For every initial point z_0 the sequence $z_n = T^n z_0$ converges weakly in average to a fixed point of T, which is the strong limit of the sequence $P_{\mathcal{S}} T^n z_0$.*

Extensions: The previous result holds also if X is uniformly convex with Fréchet-differentiable norm and $\lambda_n \to 1$ or if X is superreflexive ([93]). The convergence in average is strong if C is symmetric and T is odd (see [16]).

A fairly general result is the following from [37]. Set $\bar{z}_n = \frac{1}{\sigma_n} \sum_{k=1}^{n} z_k$, where z_n is given in (2.10).

Theorem 3.33 (Convergence in average) *Let A be a maximal monotone operator on a closed convex subset C of a Hilbert space H. Assume $\sum \lambda_n = \infty$ and $\sum \|\bar{z}_n - \bar{z}_{n-1}\|^2 < \infty$. If $\mathcal{S} \neq \emptyset$ then $\bar{z}_n$ converges weakly in average to a point $w = \lim_{n\to\infty} P_{\mathcal{S}} \bar{z}_n$. Otherwise $\lim_{n\to\infty} \|\bar{z}_n\| = \infty$.*

Proof. As in the proof of Theorem 3.24 we first prove that every weak cluster point of the sequence $\{\bar{z}_n\}$ lies in $\mathcal{S}$. Then we show that the projections $w_n = P_{\mathcal{S}} \bar{z}_n$ converge to some $w \in \mathcal{S}$ and finally that w can be the only weak cluster point of the bounded sequence $\{\bar{z}_n\}$.

First, let $[u, v] \in A$ and set $\bar{z}_n^* = (\bar{z}_n - \bar{z}_{n+1})/\lambda_n \in A\bar{z}_n$. We have

$$\begin{aligned} \|\bar{z}_{n+1} - u\|^2 &= \|\bar{z}_n - \lambda_n \bar{z}_n^* - u\|^2 \\ &= \|\bar{z}_n - u\|^2 + \|\lambda_n \bar{z}_n^*\|^2 + 2\lambda_n \langle \bar{z}_n^*, u - \bar{z}_n \rangle \\ &\leq \|\bar{z}_n - u\|^2 + \|\lambda_n \bar{z}_n^*\|^2 + 2\lambda_n \langle v, u - \bar{z}_n \rangle. \end{aligned} \tag{3.3}$$

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

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

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 $\mathcal{S}$.

Next, take $u \in \mathcal{S}$. From equation (3.3) we get

$$\|\bar{z}_{n+1} - u\|^2 \leq \|\bar{z}_n - u\|^2 + \|\lambda_n \bar{z}_n^*\|^2,$$

thus for each $p \in \mathbf{N}$ we have

$$\|\bar{z}_{n+p} - u\|^2 \leq \|\bar{z}_n - u\|^2 + \rho_n, \tag{3.4}$$

where $\rho_n = \sum_{k\geq n} \|\lambda_k \bar{z}_k^*\|^2$, which tends to 0 as $n \to \infty$. On the other hand, using the parallelogram identity and the convexity of $\mathcal{S}$ we obtain

$$\begin{aligned} \|w_{n+p} - w_n\|^2 &= 2\|\bar{z}_{n+p} - w_n\|^2 + 2\|\bar{z}_{n+p} - w_{n+p}\|^2 - 4\|\bar{z}_{n+p} - \tfrac{1}{2}(w_n + w_{n+p})\|^2 \\ &\leq 2\|\bar{z}_{n+p} - w_n\|^2 - 2\|\bar{z}_{n+p} - w_{n+p}\|^2. \end{aligned}$$

Using inequality (3.4) with $u = w_n$ we get

$$0 \leq \|w_{n+p} - w_n\|^2 \leq 2\rho_n + 2\|\bar{z}_n - w_n\|^2 - 2\|\bar{z}_{n+p} - w_{n+p}\|^2.$$

This implies that the sequence $\{\|\bar{z}_n - w_n\|\}$ converges and we deduce that the sequence $\{w_n\}$ converges as well. Clearly the limit, w, must be in the closed set $\mathcal{S}$.

Finally, let $z_{n_k} \rightharpoonup \zeta \in \mathcal{S}$. We shall prove that $\zeta = w$. Since $\langle \zeta - w_n, z_n - w_n\rangle \leq 0$ we have

$$\langle \zeta - w, z_n - w_n\rangle \leq \langle w_n - w, z_n - w_n\rangle \leq M\|w_n - w\|,$$

where $M = \sup \|w_n - z_n\|$. This implies

$$\left\langle \zeta - w, \bar{z}_n - \frac{1}{\sigma_n}\sum_{k=1}^{n} \lambda_k w_k \right\rangle \leq \frac{M}{\sigma_n}\sum_{k=1}^{n} \lambda_k \|w_k - w\|.$$

Taking $n = n_k$ and letting $k \to \infty$ we get $\langle \zeta - w, \zeta - w\rangle \leq 0$, which implies $\zeta = w$ and completes the proof. ■

3.4 Asymptotic equivalence

In this section we explain how to deduce qualitative information on the asymptotic behavior of the systems defined by (2.1), (2.7) and (2.10). We provide a comparison tool that guarantees that two evolution systems share certain symptotic properties. For the complete abstract theory see Chapter 7.

3.4.1 Contracting evolution systems and almost-orbits

Let C be a convex subset of a Banach space X and let I denote the identity operator in X. An *evolution system* 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) $V(t,s)V(s,r) = V(t,r)$.

An evolution system is *contracting* if it satisfies

iii) $\|V(t,s)x - V(t,s)y\| \le \|x - y\|$.

Example 5 *The contraction semigroups and the products of resolvents define contracting evolution systems.*

Let V be an evolution system on C. A locally bounded function $u : \mathbf{R}_+ \to C$ is an *almost-orbit* of V if

$$\lim_{t\to\infty} \|u(t+h) - V(t+h,t)u(t)\| = 0 \qquad \text{uniformly in} \quad h \ge 0. \tag{3.5}$$

A locally bounded trajectory of the form $t \mapsto V(t,s)x$ for s and x fixed is called an *orbit* of V.

3.4.2 Asymptotic properties

The proof of the following result, under a more general form, can be found in Chapter 7.

Theorem 3.34 *Let V be an evolution system. We have the following:*

1. *If $V(t,s)x$ converges weakly (resp. strongly) as $t\to\infty$ for all x and s, then so does every almost-orbit.*

2. *If $V(t,s)x$ almost-converges weakly (resp. strongly) as $t\to\infty$ for all x and s, then so does every almost-orbit.*

3. *If $V(t,s)x$ converges weakly (resp. strongly) in average as $t\to\infty$ for all x and s, then so does every almost-orbit.*

Remark 3.35 *The first part of the previous theorem was proved in [90] if V is defined by a semigroup of contractions or if the almost-orbits are orbits of a semigroup of contractions. The second part was proved in [84] under the following supplementary assumptions:*

1. *V is defined by a strongly continuous semigroup of contractions;*

2. *The set of common fixed points of the semigroup is nonempty; and*

3. *For the weak topology, the space X is* weakly sequentially complete*, which means that every weak Cauchy sequence converges (weakly) to an element in X. 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 [77]).*

3.4.3 A comparison tool

Proposition 3.36 *Let A be a m-accretive operator on a Banach space X and let S be the semigroup generated by A. Assume one of the following conditions holds:*

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

ii) X is Hilbert space, $A = \partial f$ and $\{\lambda_n\} \notin \ell^1$.

The following are equivalent:

(1) For every initial condition $x_0 \in X$ the proximal sequence x_n converges weakly (strongly) as $n\to\infty$ to a zero of A; and

(2) For every initial condition $x \in X$ the continuous trajectory $S(t)x$ converges weakly (strongly) as $t \to \infty$ to a zero of A.

If $\{\lambda_n\} \in \ell^2 \setminus \ell^1$ and $A = I - T$, where T is nonexpansive, then the preceding statements are equivalent to

(3) For every initial condition $z_0 \in X$ the sequence z_n given by Euler's discretization converges weakly (strongly) as $n \to \infty$ to a zero of A.

This result was proved in [70] under i) and in [56] under ii). Both proofs rely on Kobayashi's inequality.

Chapter 4

Some nonautonomous systems

4.1 General existence results

Assume that $\{A(t)\}$ is a family of m-accretive operators on a common domain D and consider the problem of finding a function $x(t)$ satisfying

$$\begin{cases} x'(t) & \in & -A(t)x(t) & \text{a.e. on } (0,\infty) \\ x(s) & = & x_s & \in & D(A). \end{cases} \tag{4.1}$$

Existence of such a function was proved in [49] under any of the following conditions on the evolution of $A(t)$. Set $J_\lambda(t) = (I + \lambda A(t))^{-1}$.

- $\|J_\lambda(t)x - J_\lambda(\tau)x\| \le \lambda \|f(t) - f(\tau)\| L(\|x\|)$, where f is continuous and L is increasing.
- $\|J_\lambda(t)x - J_\lambda(\tau)x\| \le \lambda \|f(t) - f(\tau)\| L(\|x\|)[1 + \|A(\tau)x\|]$, where f is of bounded variation and L is increasing.

They find the corresponding solution following the ideas in [47] (discretizing and then taking limit by refinement of the mesh). They show that for each $T > 0$ the sequence of functions $u_n^s : [0,T] \to X$ defined by

$$u_n^s(t) = \prod_{k=1}^{n} \Big[\; I + \left(\tfrac{t-s}{n}\right) A\left(s + \tfrac{k}{n}(t-s)\right) \; \Big]^{-1} x_s$$

converges uniformly on $[0,T]$ and its limit defines a contracting evolution system $U(t,s)$ that solves the problem.

4.2 Steepest descent with perturbations

Let $f \in \Gamma_0(H)$. In order to minimize f one may use the following scheme: Take a family $\{f_\varepsilon\}_{\varepsilon>0}$ of functions in $\Gamma_0(H)$ approximating f, in some sense, as $\varepsilon \to 0$. Select a parameterization (as regular as necessary) $t \mapsto \varepsilon(t)$ with $\lim\limits_{t\to\infty} \varepsilon(t) = 0$ and consider the differential inclusion

$$-x'(t) \in \partial f_{\varepsilon(t)}(x(t)). \tag{4.2}$$

We assume that it has a solution $\hat{x}$. For conditions ensuring this see, for example, [8]. In some cases the functions $t \mapsto f(\hat{x}(t))$ and $t \mapsto \hat{x}(t)$ can give information about the optimal value $f^* = \inf f$ and the minimizing set $\mathcal{S} = \text{Argmin} f$, respectively. In what follows, we assume that for all sufficiently small ε there is a function x_ε satisfying

$$-x_\varepsilon'(t) \in \partial f_\varepsilon(x_\varepsilon(t)).$$

We also assume that for each ε the set $\mathcal{S}_\varepsilon = \text{Argmin} f_\varepsilon$ reduces to a point x^*_ε. Assume also that the function ξ defined by $\xi(\varepsilon) = x^*_\varepsilon$ is absolutely continuous on compact intervals $[\varepsilon_1, \varepsilon_2]$ and $\lim_{\varepsilon\to 0} \xi(\varepsilon) = x^* \in \mathcal{S}$.

4.2.1 Slow parameterization

In what follows we assume that each f_ε is strictly convex and for each $R > 0$ there is a positive function β_R such that

$$\langle y - v, x - u\rangle \geq \beta_R(\varepsilon)|x - u|^2 \tag{4.3}$$

for any $[x, y],\ [u, v] \in \partial f_\varepsilon$ with $x, u \in \bar{B}(0, R)$. Let $\varepsilon(t)$ be a decreasing function such that

$$\int_0^\infty \beta_R(\varepsilon(s))ds = +\infty.$$

This imposes that the convergence of the function $t \mapsto \varepsilon(t)$ to 0 is *sufficiently slow.*

Example 6 (Tikhonov regularization) *Let* $f_\varepsilon(x) = f(x) + \frac{\varepsilon}{2}\|x\|^2$. *Here* $\beta(\varepsilon) = \varepsilon$ *and* $\xi(\varepsilon)$ *is determined by*

$$\partial f(\xi(\varepsilon)) + \varepsilon\xi(\varepsilon) \ni 0 \qquad \textit{which is} \qquad \xi(\varepsilon) = (I + (1/\varepsilon)\partial f)^{-1}(0).$$

The following results first appeared in [6]:

Theorem 4.1 (Short optimal trajectory) *Assume* $\xi(\varepsilon)$ *has finite length; in other words, that there is* $\varepsilon_0 > 0$ *such that* $\int_0^{\varepsilon_0} \|\xi'(\varepsilon)\|d\varepsilon < \infty$. *Then* $\hat{x}(t)$ *converges strongly to* x^*.

Theorem 4.2 (Long optimal trajectory) *Let* β *satisfy inequality* (4.3) *for all* $x, u \in H$. *Assume* $\xi(\varepsilon)$ *satisfies* $\|\xi'(\varepsilon)\| \leq \frac{1}{\gamma(\varepsilon)}$ *for some positive function* γ *and choose* $t \mapsto \varepsilon(t)$ *with*

$$\lim_{t\to+\infty} \frac{\varepsilon'(t)}{\beta(\varepsilon(s))\gamma(\varepsilon(s))} = 0.$$

Then $\hat{x}(t)$ *converges strongly to* x^*.

In order to prove Theorems 4.1 and 4.2 above we need the following Gronwall-type inequality from [6], which is easy to prove:

Lemma 4.3 *Let* $m, \theta : [t_0, t_1] \to [0, \infty)$ *with* m *integrable and* θ *absolutely continuous such that*

$$\theta'(t) \leq m(t)\sqrt{\theta(t)} \qquad \textit{a.e. on} \qquad [t_0, t_1].$$

Then for all $t \in [t_0, t_1]$ *we have*

$$\sqrt{\theta(t)} \leq \sqrt{\theta(t_0)} + \frac{1}{2}\int_{t_0}^t m(s)\ ds.$$

If $\hat{x}(\cdot)$ satisfies (4.2) define $\phi(t) = \frac{1}{2}\|\hat{x}(t) - \xi(\varepsilon(t))\|^2$. Since $\xi(\varepsilon(t)) \to x^*$ as $t \to \infty$ it suffices to prove that $\phi(t) \to 0$ as $t \to \infty$.

Proof of Theorem 4.1 First observe that

$$\begin{aligned}
\phi'(t) &= \langle \hat{x}'(t) - \xi'(\varepsilon(t))\varepsilon'(t), \hat{x}(t) - \xi(\varepsilon(t))\rangle \\
&\leq \langle \hat{x}'(t), \hat{x}(t) - \xi(\varepsilon(t))\rangle - \varepsilon'(t)\|\xi'(\varepsilon(t))\|\|\hat{x}(t) - \xi(\varepsilon(t))\| \\
&\leq -\beta_R(\varepsilon(t))\|\hat{x}(t) - x(\varepsilon(t))\|^2 - \varepsilon'(t)\|\xi'(\varepsilon(t))\|\|\hat{x}(t) - x(\varepsilon(t))\| \qquad (4.4)
\end{aligned}$$

and so

$$\phi'(t) \le -\varepsilon'(t)\|\xi'(\varepsilon(t))\|\|\hat{x}(t) - x(\varepsilon(t))\|.$$

Using Lemma 4.3 and a change of variable we obtain

$$\sqrt{\phi(t)} \le \sqrt{\phi(0)} + \frac{1}{\sqrt{2}}\int_0^{\varepsilon(0)} \|\xi'(\varepsilon)\|\, d\varepsilon,$$

so $u(t)$ is bounded. Take $R > 0$ such that $u(t)$ and $x(\varepsilon(t))$ belong to $\bar{B}(0,R)$ for all t. From (4.4) we get

$$\phi'(t) + 2\beta_R(\varepsilon(t))\phi(t) \le -\sqrt{2}\|\xi'(\varepsilon)\|\varepsilon'(t)\sqrt{\phi(t)}. \tag{4.5}$$

Define $E(t) = \int_{t_0}^t \beta_R(\varepsilon(s))\, ds$ and multiply (4.5) by $2e^{E(t)}$. Apply Lemma 4.3 and a change of variable to obtain

$$\sqrt{\phi(t)} \le e^{E(t_0)-E(t)}\sqrt{\phi(t_0)} + \frac{1}{\sqrt{2}}\int_{\varepsilon(t)}^{\varepsilon(t_0)} \|\xi'(\varepsilon)\|\, d\varepsilon,$$

which implies

$$\limsup_{t\to\infty} \sqrt{\phi(t)} \le \frac{1}{\sqrt{2}}\int_0^{\varepsilon(t_0)} \|\xi'(\varepsilon)\|\, d\varepsilon.$$

Finally let $t_0 \to \infty$ to conclude that $\lim_{t\to\infty} \phi(t) = 0$. ■

Proof of Theorem 4.2 Proceeding as in the proof of Theorem 4.1 with β instead of β_R we obtain

$$\sqrt{\phi(t)} \le e^{E(t_0)-E(t)}\sqrt{\phi(t_0)} - \frac{1}{\sqrt{2}}\int_{t_0}^{t} e^{E(s)-E(t)}\|\xi'(\varepsilon(s))\|\varepsilon'(s)\, ds.$$

Set

$$h(t) = \sup_{s\ge t} \frac{|\varepsilon'(s)|}{\beta(\varepsilon(s))\gamma(\varepsilon(s))},$$

so that $h(t) \to 0$ as $t \to \infty$. We have

$$\sqrt{\phi(t)} \le e^{E(t_0)-E(t)}\sqrt{\phi(t_0)} + \frac{e^{-E(t)}h(t_0)}{\sqrt{2}}\int_{t_0}^{t} e^{E(s)}E'(s)\, ds$$

and so

$$\limsup_{t\to\infty} \sqrt{\phi(t)} \le \frac{1}{\sqrt{2}}h(t_0).$$

Finally we let $t_0 \to \infty$ and the result follows. ■

4.2.2 Fast parameterization

The results presented in this part are due to Cominetti and Alemany [43].

We make the following hypotheses:

1. $\|\xi'(\varepsilon)\| \le \frac{1}{\gamma(\varepsilon)}$ for some positive bounded function γ;
2. $\xi(\varepsilon)$ has finite length: there is $\varepsilon_0 > 0$ such that $\int_0^{\varepsilon_0} \|\xi'(\varepsilon)\|\, d\varepsilon < \infty$;
3. $f(x) \le \liminf_{k\to\infty} f_{\varepsilon_k}(x_k)$ for all $x \in H$, $\varepsilon_k \to 0$, $x_k \rightharpoonup x$;

4. The parameterization ε is differentiable, decreasing and for every $y \in \mathcal{S}$ there is $\eta > 0$ such that
$$\sup_{T>0} \int_0^T \left[f_{\varepsilon(t)} \left(\xi(\varepsilon(t)) - \eta(y - x^*) \right) - f^*_{\varepsilon(t)} \right] dt < \infty.$$

Proposition 4.4 *Assume the conditions above hold.*

i) If H is finite dimensional then $\hat{x}(t)$ converges to some $x_\infty \in \mathcal{S}$; and

ii) If the function $\varepsilon \mapsto f_\varepsilon(x)$ is nondecreasing for every $x \in H$, then $\hat{x}(t)$ converges weakly to some $x_\infty \in \mathcal{S}$.

The transition from fast to slow parameterization is continuous. To see this, let $\{\varepsilon_k\}$ be a sequence of parameterizations such that $x^k_\infty = \lim_{t\to\infty} x_k(t)$ exists for each $k \in \mathbf{N}$. Assume that $\xi(\varepsilon)$ has finite length.

Proposition 4.5 *If $\int_0^\infty \beta(\varepsilon_k(t))\, dt \to \infty$ as $k \to \infty$ then $x^k_\infty \to x^*$ as $k \to \infty$.*

4.2.3 Tikhonov dynamics

Let $f_\varepsilon(x) = f(x) + \frac{\varepsilon}{2}|x|^2$. It is well known that the unique minimizer $\xi(\varepsilon)$ of f_ε converges strongly to the point $x^* \in S$ having minimal norm.

According to the results in the preceding section, whenever $\int_0^\infty \varepsilon(t)dt = \infty$ the trajectory $t \mapsto \hat{x}(t)$ converges strongly to x^* as $t \to \infty$, provided one of the following additional assumptions hold:

(a) The optimal path $\varepsilon \mapsto \xi(\varepsilon)$ has finite length, *i.e.* $\int_0^{\varepsilon_0} \|\xi'(\varepsilon)\| d\varepsilon < \infty$; or

(b) $\varepsilon'(t)/\varepsilon(t)^2 \to 0$ when $t \to \infty$.

We shall prove next that, in this particular case, neither (a) nor (b) are necessary. In addition we establish that when $\int_0^\infty \varepsilon(t)dt < \infty$ we still have weak convergence of $\hat{x}(t)$ towards a point $x_\infty \in \mathcal{S}$.

Theorem 4.6 *With the preceding notation,*

i) If $\int_0^\infty \varepsilon(t)dt = \infty$ then $\hat{x}(t)$ converges strongly to x^*; and

ii) If $\int_0^\infty \varepsilon(t)dt < \infty$ then $\hat{x}(t)$ converges weakly to some $x_\infty \in S$.

Proof. For part *i)* define $\theta(t) = \frac{1}{2}\|\hat{x}(t) - x^*\|^2$ so that $\dot{\theta}(t) = \langle \hat{x}'(t), \hat{x}(t) - x^* \rangle$. Using the strong convexity of f_ε and the differential inclusion we get
$$f_{\varepsilon(t)}(\hat{x}(t)) - \langle \hat{x}'(t), x^* - \hat{x}(t) \rangle + \tfrac{\varepsilon(t)}{2} \|\hat{x}(t) - x^*\|^2 \leq f_{\varepsilon(t)}(x^*)$$
that is to say
$$\theta'(t) + \varepsilon(t)\theta(t) \leq f_{\varepsilon(t)}(x^*) - f_{\varepsilon(t)}(\hat{x}(t)).$$
Since $f_{\varepsilon(t)}(\xi(\varepsilon(t))) \leq f_{\varepsilon(t)}(\hat{x}(t))$ and $f(x^*) \leq f(\xi(\varepsilon(t)))$ we deduce
$$\theta'(t) + \varepsilon(t)\theta(t) \leq \tfrac{\varepsilon(t)}{2} \left[\|x^*\|^2 - \|\xi(\varepsilon(t))\|^2 \right].$$

Now, $\|\xi(\varepsilon)\|^2 \to \|x^*\|^2$ as $\varepsilon \to 0$, so that for each $\eta > 0$ we may find $T \geq 0$ such that $\theta'(t) + \varepsilon(t)\theta(t) \leq \eta\varepsilon(t)$ for $t \geq T$. Multiplying this inequality by $\varphi(t) = \exp(\int_0^t \varepsilon(\tau)\, d\tau)$ we get $\frac{d}{dt}[\varphi(t)(\theta(t) - \eta)] \leq 0$ so that $\varphi(t)(\theta(t) - \eta)$ is decreasing and hence bounded. Since $\varphi(t) \to \infty$ we deduce $\limsup_{t\to\infty} \theta(t) \leq \eta$. Since η is arbitrary, $\theta(t) \to 0$ thus $\hat{x}(t) \to x^*$ as

$t \to \infty$.

For part *ii)* we take $\bar{x} \in \mathcal{S}$ and set $\theta(t) = \frac{1}{2}\|\hat{x}(t) - \bar{x}\|^2$. As before we get

$$\theta'(t) + \varepsilon(t)\theta(t) \leq f(\bar{x}) - f(\hat{x}(t)) + \tfrac{\varepsilon(t)}{2}\left[\|\bar{x}\|^2 - \|\hat{x}(t)\|^2\right] \tag{4.6}$$

from which it follows easily that $\theta'(t) \leq \frac{\|\bar{x}\|^2}{2}\varepsilon(t)$. Thus $\theta(t) - \frac{\|\bar{x}\|^2}{2}\int_0^t \varepsilon(\tau)\,d\tau$ is decreasing and hence converges as $t \to \infty$. Since the integral $\int_0^t \varepsilon(\tau)\,d\tau$ is also convergent the same holds for $\theta(t)$. By virtue of Opial's Lemma it only remains to show that every weak cluster point of $\hat{x}(t)$ belongs to $\mathcal{S}$. To that end, it suffices to establish that $f(\hat{x}(t)) \to f^*$. Let D^- be the upper left-handed Dini derivative[1]. We have

$$D^-\left[f(\hat{x}(t))\right] \leq -\langle \hat{x}'(t) + \epsilon(t)\hat{x}(t), \hat{x}'(t)\rangle = \tfrac{\varepsilon(t)^2}{4}\|\hat{x}(t)\|^2 - \left\|\hat{x}'(t) + \tfrac{\varepsilon(t)}{2}\hat{x}(t)\right\|^2 \leq K\varepsilon(t)$$

for some constant $K > 0$ and all sufficiently large t. It follows that

$$D^-\left[f(\hat{x}(t)) - K\int_0^t \varepsilon(\tau)\,d\tau\right] \leq 0$$

so that $f(\hat{x}(t)) - K\int_0^t \varepsilon(\tau)\,d\tau$ is decreasing and hence convergent. It follows that $f(\hat{x}(t))$ converges as well. Now, using (4.6) we obtain $f(\hat{x}(t)) - f^* \leq -\theta'(t) + \frac{\varepsilon(t)}{2}\|\bar{x}\|^2$ so that

$$0 \leq \int_0^t [f(\hat{x}(\tau)) - f^*]dt \leq \theta(0) - \theta(t) + \tfrac{\|\bar{x}\|^2}{2}\int_0^t \varepsilon(\tau)\,d\tau \leq \theta(0) + \tfrac{\|\bar{x}\|^2}{2}\int_0^\infty \varepsilon(\tau)\,d\tau < \infty$$

and $f(\hat{x}(t)) \to f^*$ as $t \to \infty$ as required. ■

Theorem 4.6 still holds if the subdifferential ∂f is replaced by any maximal monotone operator, with some minor additional assumptions. Part i) was proved in [34] and [94] for ε decreasing and in [45] for ε of bounded variation. Example 7 (taken from [45]) below shows that it does not hold for general ε. On the other hand, part ii) holds if A is demipositive (from Theorem 3.11) or if it satisfies Pazy's $\mathcal{L}$ condition (from part ii) of Theorem 3.12), by virtue of Proposition 8.6 and Theorem 3.34.

Example 7 *Consider the $\pi/2$ clockwise rotation on the plane around the point $(1,1)$. In order to simplify the computations we identify $\mathbf{R}^2$ with $\mathcal{C}$ and use the properties of complex numbers. Setting $p = 1 + i$ the inclusion $-\dot{u}(t) \in Au(t) + \varepsilon(t)u(t)$ can be rewritten as*

$$u'(t) = -i(u(t) - p) - \varepsilon(t)u(t). \tag{4.7}$$

Now let ε_n be a sequence of positive real numbers such that $\lim\limits_{n\to\infty} \varepsilon_n = 0$ and $\sum \varepsilon_n = \infty$. Take $a_0 = 0$ and let $b_n = a_n + \tau_n$, $a_{n+1} = b_n + \sigma_n$ for $n \in \mathbf{N}$. The positive sequences τ_n and σ_n will be given afterwards but they will be bounded below away from zero. Define the step function ε by

$$\varepsilon(t) = \begin{cases} \varepsilon_n & \text{if} \quad a_n \leq t < b_n \\ 0 & \text{if} \quad b_n \leq t < a_{n+1}. \end{cases}$$

Clearly, $\lim\limits_{t\to\infty} \varepsilon(t) = 0$ and $\varepsilon \notin L^1(0,\infty)$ since τ_n is bounded below away from zero.

[1]Let g be a continuous real-valued function. The upper left-handed Dini derivative of g at a point x is defined by $D^-g(x) = \limsup\limits_{h\to 0^+} \frac{g(x)-g(x-h)}{h}$. For the other Dini derivatives see [100].

Let $u(a_n) = 1$. On the interval $[a_n, b_n)$, the solution to equation (4.7) is

$$u(t) = \frac{1}{\varepsilon_n + i}\left[i - 1 + (1+\varepsilon_n)e^{-(\varepsilon_n+i)(t-a_n)}\right]. \tag{4.8}$$

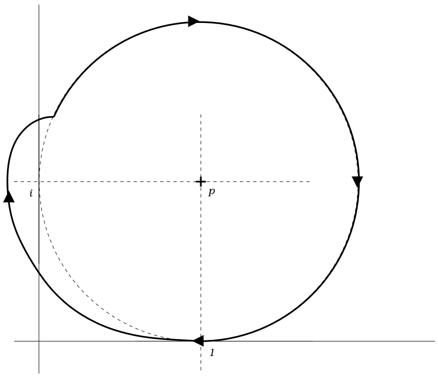
Define τ_n as the first positive zero of the function

$$\psi_n(s) = (1+\varepsilon_n)e^{-2\varepsilon_n s} - 2\varepsilon_n e^{-\varepsilon_n s}[\sin(s) - \cos(s)] + \varepsilon_n - 1.$$

This coincides with the first time after a_n when $|u(t) - p| = 1$ (the trajectory intersects the circle of radius 1 centered at p). One can prove that $\tau_n \geq 1/3$ for all n if $\varepsilon_n \leq 2$ (because ψ_n' is strictly positive on $(0, 1/3]$) and so $\varepsilon \notin L^1(0,\infty)$.

On the interval $[b_n, a_{n+1})$ the solution is

$$u(t) = p + (u(b_n) - p)e^{-i(t-b_n)}.$$



Now we pick σ_n such that $u(a_{n+1}) = 1$ in order for the solution to cycle indefinitely. More precisely, let σ_n be the first positive solution of $e^{is} = i(u(b_n) - p)$. Such positive solution exists because $|u(b_n) - p| = 1$.

On the interval $[b_n, a_{n+1})$, the trajectory $u(t)$ travels from $u(b_n)$ to 1 along the circle $|z-p| = 1$. Now, equation (4.8) implies that the real part of $u(b_n)$ is strictly less than 1. Therefore, the trajectory covers at least the arc joining (clockwise) the points $1+2i$ and 1 on the circle $|z-p| = 1$ as t goes from b_n to a_{n+1}, so it cannot converge as $t \to \infty$. The figure represents the trajectory $u(t)$ on the interval $[a_n, a_{n+1}]$, starting from 1 and back.

4.2.4 Convergence to the set $\mathcal{S}$.

In $\mathbf{R}^n$, under fairly general conditions the solutions of the nonautonomous system can be proved to approach the optimal set $\mathcal{S}$. This is done in [21]:

Theorem 4.7 *Let $\mathcal{S}$ be nonempty and bounded and assume that*

i) $f(y_\infty) \leq \lim f_{\varepsilon(t_k)}(y_k)$ whenever $y_k \to y_\infty$, $t_k \to \infty$; and

ii) $v(t) = \sup\left\{ f_{\varepsilon(t)}(y) \mid y \in S_\infty \right\} \to f^*$ as $t \to \infty$.

If $\hat{x}(t)$ is a solution of $x'(t) + \partial f_{\varepsilon(t)}(x(t)) \ni 0$ then $d(\hat{x}(t), S) \to 0$.[2]

Proof. Setting $\Theta(z) = \frac{1}{2}d(z,\mathcal{S})$, let $\theta(t) = \Theta(\hat{x}(t))$ and $h(t) = \sup\{\Theta(z) \mid f_{\varepsilon(t)}(z) \leq v(t)\}$. Observe that $h(t) \to 0$ as $t \to \infty$. Otherwise we could find $\varepsilon > 0$, $t_k \to \infty$ and $z_k \to z_\infty$ such that $f_{\varepsilon(t_k)}(z_k) \leq v(t_k)$ and $\Theta(z_k) = \varepsilon$. By i) and ii) $z_\infty \in \mathcal{S}$, which contradicts $\Theta(z_k) = \varepsilon$.

Now $\theta'(t) \leq v(t) - f_{\varepsilon(t)}(\hat{x}(t))$ almost everywhere and so $\theta(t) \geq h(t)$ implies $\theta'(t) \leq 0$. Since $h(t) \to 0$ as $t \to \infty$, the limit $\lim_{t\to\infty} \theta(t)$ exists (this is easy to prove, see [21]). We shall find a subsequence that converges to 0. Since $\int_0^\infty [f_{\varepsilon(\tau)}(\hat{x}(\tau)) - v(\tau)]\, d\tau \leq \theta(0) < \infty$ and $\hat{x}(t)$ is bounded, there is $t_k \to \infty$ such that $\hat{x}(t_k) \to x_\infty$ as $k \to \infty$ and the sequence $f_{\varepsilon(t_k)}(\hat{x}(t_k))$ is minimizing. Hence $x_\infty \in \mathcal{S}$ and $\lim_{k\to\infty} \theta(t_k) = 0$. ■

[2] In fact the authors prove this result for the inclusion $x'(t) + \alpha(t)\partial f_t(x(t)) \ni 0$, where $\alpha : [0,\infty) \to [0,\infty)$ is such that $\int_t^\infty \alpha(s)\, ds = \infty$ for each $t \geq 0$.

Part II

Asymptotic analysis of nonautonomous evolution systems

Chapter 5

Asymptotic convergence to the optimal value of diagonal proximal iterations in convex minimization

Introduction

Let $\{f_n\}$ be a family of proper, lower-semicontinuous convex functions on a real Hilbert space H with a common (convex) domain D. The functions f_n are meant to approximate an objective function f, with domain D_f, in some sense. We assume that $\inf_H f = \inf_D f \geq -\infty$ and we denote this quantity by f^*.

Take a starting point $x_0 \in H$ and two sequences of real parameters $\{\lambda_k\} \subset (0, \Lambda]$ and $\{\varepsilon_k\} \subset [0, \infty)$. A sequence $\{x_k\} \subset D$ is said to be an *inexact diagonal proximal sequence* generated by $(x_0, \{\lambda_k\}, \{f_k\}, \{\varepsilon_k\})$ if

$$y_k := \frac{x_{k-1} - x_k}{\lambda_k} \in \partial_{\varepsilon_k} f_k(x_k) \tag{5.1}$$

for all $k \geq 1$, where the approximate $\varepsilon-$subdifferential ∂_ε is defined by

$$\partial_\varepsilon g(u) = \{\ u^* \in H \mid g(v) \geq g(u) + \langle u^*, v - u\rangle - \varepsilon \ \ \forall v \in H\ \} \qquad \varepsilon \geq 0.$$

Observe that in order to construct such a sequence it is necessary to solve an optimization problem at each iteration. This is usually done by means of another algorithm (e.g. bundle methods) so the use of the approximate $\varepsilon-$subdifferential is very important for practical purposes.

If $f_n \equiv f$ and $\varepsilon_n \equiv 0$, this is the standard *proximal point algorithm*, which was first introduced in [81] for solving variational inequalities. This algorithm has been used extensively in convex minimization (see [56] and [96]) in the search for the value f^* and, if there are any, the points at which it is attained. It has also been useful to study difference approximation of differential inclusions governed by monotone operators (see [47] and [66]). For weak and strong convergence of the sequence $\{x_n\}$ see [97] when $\lambda_n \equiv \lambda$ and [33] for more general cases. Several variations have been considered: see [57] for weaker conditions on the stepsizes; [55] for *relaxation*; [101] for hybrid projection-proximal point algorithm.

The term *diagonal* refers to the fact that the objective function is updated at each iteration (thus one could think of a sequence of sequences, each one corresponding to a different function, from which we take only the terms on the diagonal). This is not a new approach; many authors have defined approximation methods for a given objective function f and combined them with the proximal point algorithm. The common feature is that a family $\{f_n\}$ is chosen so that the functions are better behaved (in terms of smoothness, computability, coerciveness, etc) and converge to f in some sense. A pioneer work is [61], where some interior penalty methods are considered (see also [62]). For exterior penalties and variational convergence see [1]. For Tikhonov, see [86]. From the point of view of computation and bundle methods, see [10] and [11]. In the latter, a sequence of exterior penalties is considered and their algorithm is proved to converge in a finite number of steps in the linear case. On the other hand, in [42] and [44] the authors study exponential penalty and log-barrier in linear programming using properties of the limit (unperturbed) dynamics.

We shall present some estimations concerning the trajectory $\{x_n\}$, the values $\{f_n(x_n)\}$ and the velocities $\{y_n\}$, which will be used to derive qualitative and quantitative convergence results provided $f_n \to f$ in a certain way. Some of our results generalize the corresponding ones in [56], where $f_n \equiv f$. The present work has three distinctive marks: first, some results can be applied to find the optimal value even if the objective function has no minimizers and is not coercive (cf [1]); second, the results do not depend explicitly on the type of approximation; and third, when there are minimizers it is easy to see how the proximal point algorithm and the family $\{f_n\}$ influence the rate of convergence to the value f^*.

The paper is organized as follows: in the first section we give the basic inequalities and derive convergence of the sequence $\{f_n(x_n)\}$; section 2 deals with rates of convergence; section 3 shows some results concerning the velocities and the proximal sequence itself; in the fourth section we give a short overview of finite convergence; and finally in section 5 there are some ideas for possible extensions.

5.1 Convergence of the values $f_n(x_n)$.

Let us start by giving some notation. Set $\sigma_0 = 0$ and for $n \geq 1$, $\sigma_n = \sum_{k=1}^n \lambda_k$. Unless otherwise stated we assume $\lim_{n\to\infty} \sigma_n = \infty$. For a sequence $\{z_n\}$ denote by $\overline{z_n} = \frac{1}{\sigma_n}\sum_{k=1}^n \lambda_k z_k$ the sequence of its averages weighted by the stepsizes $\{\lambda_k\}$. Notice that $\lim_{n\to\infty} z_n = L$ implies $\lim_{n\to\infty} \overline{z_n} = L$. Finally, let $\{p_n\}$ and $\{\varrho_n\}$ be two sequences of positive numbers. We shall write $p_n = \mathcal{O}(\varrho_n)$ if the ratio p_n/ϱ_n is bounded. Observe that if $p_n = \mathcal{O}(\varrho_n)$ then $\overline{p_n} = \mathcal{O}(\overline{\varrho_n})$. If $\lim_{n\to\infty} p_n/\varrho_n = 0$ we write $p_n = o(\varrho_n)$.

Let us also recall that the Kronecker Theorem[1], implies that if a sequence $\{p_n\}$ of positive numbers is in ℓ^1, then $\lim_{n\to\infty} \frac{1}{\sigma_n}\sum_{k=1}^n \sigma_k p_k = 0$.

Lemma 5.1 *For any $u \in D$ we have*

$$\overline{f_n(x_n)} - \overline{f_n(u)} \leq \frac{\|u - x_0\|^2}{2\sigma_n} - \frac{1}{2\sigma_n}\sum_{k=1}^n \lambda_k^2\|y_k\|^2 - \frac{\|u - x_n\|^2}{2\sigma_n} + \overline{\varepsilon_n}.$$

Proof. Let $u \in D$. The subdifferential inequality gives

$$f_k(u) - f_k(x_k) \geq \frac{1}{\lambda_k}\langle x_{k-1} - x_k, u - x_k\rangle - \varepsilon_k \tag{5.2}$$

[1]See, for example, [65], p. 129.

and so

$$\begin{aligned}2\lambda_k(f_k(u)-f_k(x_k)) &\geq 2\langle x_{k-1}-x_k, u-x_k\rangle - 2\lambda_k\varepsilon_k \\ &= \|x_{k-1}-x_k\|^2+\|u-x_k\|^2-\|u-x_{k-1}\|^2-2\lambda_k\varepsilon_k \\ &= \lambda_k^2\|y_k\|^2+\|u-x_k\|^2-\|u-x_{k-1}\|^2-2\lambda_k\varepsilon_k. \end{aligned} \tag{5.3}$$

Summing for $k=1, \ldots n$ we get

$$2\sum_{k=1}^{n}\lambda_k f_k(u) - 2\sum_{k=1}^{n}\lambda_k f_k(x_k) \geq \sum_{k=1}^{n}\lambda_k^2\|y_k\|^2+\|u-x_n\|^2-\|u-x_0\|^2-2\sum_{k=1}^{n}\lambda_k\varepsilon_k. \tag{5.4}$$

Dividing by $2\sigma_n$ we obtain the desired inequality. ■

Let $\mathcal{S} = \text{Argmin} f$. For $r > 0$ define $\mathcal{S}_r = \{ u \in D \cap D_f \mid f(u) \leq f^* + r \}$ whenever $f^* > -\infty$ and $\mathcal{S}_r = \{ u \in D \cap D_f \mid f(u) \leq -r \}$ otherwise. Consider the following hypothesis:

Hypothesis $\mathbf{H_1}$: $\limsup_{n\to\infty} \overline{f_n(u)} \leq f(u)$ for all $u \in \mathcal{S}_r$ and some $r > 0$.

Finally set $f_n^* = \inf f_n$. The following result is an immediate consequence of the previous lemma:

Corollary 5.2 *Let $\lim_{n\to\infty} \overline{\varepsilon_n} = 0$ and assume hypothesis $\mathbf{H_1}$ holds. We have*

i) $\limsup_{n\to\infty} \overline{f_n(x_n)} \leq f^*$*;*

ii) If $f^* \leq \liminf_{n\to\infty} \overline{f_n^*}$, then $\lim_{n\to\infty} \overline{f_n(x_n)} = f^*$;

iii) If $f^* \leq \liminf_{n\to\infty} f_n^*$, then $\liminf_{n\to\infty} f_n(x_n) = f^*$.

This result is similar to Theorem 2.1 in [13], though the hypotheses here are weaker and more concise. Also, we show three different "degrees" in the conclusion, which help understand more clearly the effect of averaging. We notice that the sequence $\overline{f_n(x_n)}$ approximates the optimal value f^* under fairly weak convergence hypotheses on the sequence f_n and the errors ε_n.

Remark 5.3 *In the special case where there is $x^* \in \mathcal{S} \cap D$ we have*

$$\overline{f_n(x_n)} - f^* \leq \frac{\|x^*-x_0\|^2}{2\sigma_n} + \left[\overline{f_n(x^*)} - f^*\right] + \overline{\varepsilon_n}.$$

Suppose for simplicity that $f^* \leq \overline{f_n^*}$ so that the left-hand side is nonnegative. The rate of convergence of $\overline{f_n(x_n)}$ to f^* depends on three parameters, namely:

- *σ_n, which seems to be intrinsic to the proximal iterations;*
- *ε_n, which has to do with computational precision;*
- *the behavior of f_n on the minimizing set $\mathcal{S}$.*

If $\sum \lambda_k\varepsilon_k < \infty$ and $\overline{f_n(x^*)} - f^* = \mathcal{O}(\frac{1}{\sigma_n})$ for some $x^* \in \mathcal{S}$, then $\overline{f_n(x_n)} - f^* = \mathcal{O}(\frac{1}{\sigma_n})$. The same speed of convergence is provided by Theorem 2.1 in [56] where $f_n \equiv f$ and $\varepsilon_n \equiv 0$.

In order for the values $\{f_n(x_n)\}$ to converge we need further hypotheses on the way the sequence $\{f_n\}$ evolves:

Hypothesis $\mathbf{H_2}$: There exist a set $K \subseteq H$ containing the trajectory $\{x_n\}$ and a nonnegative sequence $\{a_n\}$ such that $f_n(x) \leq f_{n-1}(x) + a_{n-1}$ for all $n \geq 2$ and all $x \in K$.

Remark 5.4 *The inequality in Hypothesis* $\mathbf{H_2}$ *is somewhat uniform and therefore difficult to establish in the whole domain without any further assumptions. However, the fact that the set* K *is chosen according to* $\{x_n\}$ *plays a key role in practice since it is possible to get* a priori *estimations for the sequence in many interesting applications. For example, the verification of hypothesis* $\mathbf{H_2}$ *is considerably simplified if one knows beforehand that the sequence belongs to a bounded or compact set (which occurs under coerciveness or inf-compactness assumptions on the family* $\{f_n\}$*).*

Example 8 *Hypothesis* $\mathbf{H_2}$ *holds trivially if the sequence* $\{f_n\}$ *is decreasing (here* $K = D$ *and* $a_n \equiv 0$*); which happens, for instance, for Tikhonov's approximation or if we consider penalization schemes as described in [44] using the log- or the inverse barrier.*

Example 9 *Hypothesis* $\mathbf{H_2}$ *is true if the sequence converges uniformly. This might seem to be restrictive at a first glance but it actually occurs in some interesting applications. For example, let* $g_1, \ldots, g_m \in \Gamma_0(H)$ *and define* $f(u) = \max\{ g_i(u) \mid 1 \leq i \leq m \}$*. A standard way to approximate* f *is to set* $F(u, \eta) = \eta \log \sum_{i=1}^{m} \exp\left(\frac{g_i(u)}{\eta}\right)$ *for* $\eta > 0$*. It is easy to see that* $f(u) \leq F(u, \eta) \leq f(u) + \eta \log(m)$*. For a positive sequence* $\eta_n \to 0$ *one defines* $f_n(u) = F(u, \eta_n)$ *so that* $f_n \to f$ *uniformly as* $n \to \infty$*. This approximation happens to be decreasing too.*

Lemma 5.5 *Assume hypothesis* $\mathbf{H_2}$ *holds. For every* $u \in D$ *we have*

$$f_n(x_n) - \overline{f_n(u)} \leq \frac{\|u - x_0\|^2}{2\sigma_n} + \frac{1}{\sigma_n} \sum_{k=1}^{n} \sigma_k (a_k + \varepsilon_k) - \frac{1}{2\sigma_n} \sum_{k=1}^{n} \lambda_k (\sigma_{k-1} + \sigma_k) \|y_k\|^2 - \frac{\|u - x_n\|^2}{2\sigma_n}.$$

Proof. We set $u = x_{k-1}$ in (5.2), multiply the resulting inequality by σ_{k-1} and use hypothesis $\mathbf{H_2}$ to obtain

$$\sigma_{k-1} f_{k-1}(x_{k-1}) - \sigma_k f_k(x_k) + \lambda_k f_k(x_k) + \sigma_{k-1} a_{k-1} \geq \sigma_{k-1} \lambda_k \|y_k\|^2 - \sigma_{k-1} \varepsilon_k.$$

Summing for $k = 1, \ldots n$ we get

$$-\sigma_n f_n(x_n) + \sum_{k=1}^{n} \lambda_k f_k(x_k) + \sum_{k=2}^{n} \sigma_{k-1} a_{k-1} \geq \sum_{k=2}^{n} \sigma_{k-1} \lambda_k \|y_k\|^2 - \sum_{k=2}^{n} \sigma_{k-1} \varepsilon_k.$$

Adding twice this inequality to (5.4) we get

$$\begin{aligned} 2\sum_{k=1}^{n} \lambda_k f_k(u) - 2\sigma_n f_n(x_n) + 2\sum_{k=2}^{n} \sigma_{k-1} a_{k-1} & \\ \geq \quad 2\sum_{k=2}^{n} \sigma_{k-1} \lambda_k \|y_k\|^2 + \sum_{k=1}^{n} \lambda_k^2 \|y_k\|^2 + \|u - x_n\|^2 - \|u - x_0\|^2 - 2\sum_{k=1}^{n} \sigma_k \varepsilon_k & \\ = \quad \sum_{k=1}^{n} \lambda_k (\sigma_k + \sigma_{k-1}) \|y_k\|^2 + \|u - x_n\|^2 - \|u - x_0\|^2 - 2\sum_{k=1}^{n} \sigma_k \varepsilon_k. & \end{aligned}$$

Dividing by $2\sigma_n$ and rearranging the terms we obtain the desired inequality. ■

In particular if $\mathbf{H_2}$ holds we have

$$f_n(x_n) - \overline{f_n(u)} \leq \frac{\|u - x_0\|^2}{2\sigma_n} + \frac{1}{\sigma_n} \sum_{k=1}^{n} \sigma_k (a_k + \varepsilon_k). \tag{5.5}$$

Rearranging the terms and using $\mathbf{H_2}$ once more we get

$$-\sum_{k=n}^{\infty} a_k \leq f_n(x_n) - f^* \leq \frac{1}{\sigma_n}\sum_{k=1}^{n}\sigma_k(a_k+\varepsilon_k) + \frac{\|u-x_0\|^2}{2\sigma_n} + \left[\overline{f_n(u)} - f^*\right]. \tag{5.6}$$

The first inequality in (5.6) gives

$$\liminf f_n(x_n) \geq f^*. \tag{5.7}$$

We shall prove that the sequence $f_n(x_n)$ approaches the optimal value f^* without any further assumptions. Convergence of the values $f_n(x_n)$ has been proved in a monotone setting (see Theorem 3.1 in [1] and Corollary 3.1 in [73]) or assuming that $\mathcal{S} \neq \emptyset$, that $f_n \to f$ in the sense of Mosco[2] and that $\{\lambda_n\}$ is bounded away from zero (see Theorems 3.2 and 3.3 in [1] and Theorem 2.2 in [13]).

Theorem 5.6 *Assume hypotheses* $\mathbf{H_1}$ *and* $\mathbf{H_2}$ *hold. If the sequences* $\{a_n\}$ *and* $\{\varepsilon_n\}$ *are in* ℓ^1 *then* $\lim_{n\to\infty} f_n(x_n) = f^*$. *If moreover* $f^* > -\infty$ *then*

i) $\lim_{n\to\infty} \overline{\sigma_n\|y_n\|^2} = 0$;

ii) $\lim_{n\to\infty} \frac{\|x_n\|^2}{\sigma_n} = 0$; *and*

iii) $\sum_{k=1}^{\infty} \lambda_k\|y_k\|^2 < \infty$.

Proof. First recall from (5.7) that $\liminf_{n\to\infty} f_n(x_n) \geq f^*$. Now take $r > 0$ and let $n \to \infty$ in inequality (5.5), so that hypothesis $\mathbf{H_1}$ and the Kronecker Theorem give $\limsup_{n\to\infty} f_n(x_n) \leq f(u)$ for all $u \in \mathcal{S}_r$. We conclude that $\limsup_{n\to\infty} f_n(x_n) \leq f^*$ and so $\lim_{n\to\infty} f_n(x_n) = f^*$.[3] Now assume $f^* > -\infty$.

i) For $\eta > 0$ take x_η such that $f(x_\eta) < f^* + \eta$. Using $\mathbf{H_1}$, the estimation given by Lemma 5.5 with $u = x_\eta$ and the fact that $\lim_{n\to\infty} f_n(x_n) = f^*$ we get

$$\limsup_{n\to\infty} \frac{1}{\sigma_n}\sum_{k=1}^{n} \lambda_k\sigma_k\|y_k\|^2 \leq f(x_\eta) - f^* < \eta.$$

ii) The proof is similar.

iii) Setting $u = x_{k-1}$ in (5.2) and using hypothesis $\mathbf{H_2}$ we find

$$\lambda_k\|y_k\|^2 \leq f_{k-1}(x_{k-1}) - f_k(x_k) + a_{k-1} + \varepsilon_k$$

for $k \geq 2$. Summing for $k = 2 \ldots n$ we get

$$\sum_{k=2}^{n} \lambda_k\|y_k\|^2 \leq f_1(x_1) - f_n(x_n) + \sum_{k=2}^{n} a_{n-1} + \sum_{k=1}^{n} \varepsilon_k.$$

From the first part, the right-hand side converges as $n \to \infty$. ∎

[2] Epi-convergence both for the weak and the strong topologies (see below).

[3] Notice that this argument is still valid if $f^* = -\infty$.

Part i) implies $\lim_{n\to\infty} \overline{\|y_n\|^2} = 0$ and so $\lim_{n\to\infty} \overline{\|y_n\|} = 0$.[4] On the other hand, part iii) reveals that $\lim_{n\to\infty} \lambda_n \|y_n\|^2 = 0$. Therefore, if $\lambda_n \geq \lambda > 0$ then $\lim_{n\to\infty} \|y_n\| = 0$.

Observe that, under epi-convergence, every cluster point of the sequence $\{x_n\}$ will be a minimizer of the limit function f. More precisely, let τ be the strong or the weak topology in H. We have the following:

Corollary 5.7 *In addition to the hypotheses of Theorem 5.6, suppose that for all $x \in D_f$ and for all sequences $\{z_n\}$ in D converging to x for the topology τ we have $f(x) \leq \liminf_{n\to\infty} f_n(z_n)$. Then every $\tau-$cluster point of the proximal sequence $\{x_n\}$ is a minimizer of f.*

A similar result has been proved in [1] under different hypotheses including boundedness of the sequence $\{x_n\}$.

5.2 Some remarks on the rate of convergence.

It is natural to expect the rate of convergence of $f_n(x_n)$ to depend on the way the sequence f_n converges to f. A positive feature of our approach is that when the function f has minimizers we can give precise estimates on the rate of convergence of the values in terms of the rate of pointwise convergence of the sequence $\{f_n\}$ on the optimal set. One drawback is that the results that we present in this section about the rate of convergence require the set $\mathcal{S} \cap D$ to be nonempty. This is the case for exterior penalties but not for barriers. Apart from this section, this hypothesis is also necessary in the proof of Proposition 5.10, and there only.

Proposition 5.8 *Suppose $\mathbf{H_2}$ holds and there exists $x^* \in \mathcal{S} \cap D$ such that $\overline{|f_n(x^*) - f(x^*)|} = \mathcal{O}(\frac{1}{\sigma_n})$. If $\sum_{k=1}^{n} \sigma_k(a_k + \varepsilon_k) < \infty$ then $|f_n(x_n) - f^*| = \mathcal{O}(\frac{1}{\sigma_n})$.*

Proof. It suffices to apply inequality (5.6) noticing that $\sum_{k\geq n} a_k = \frac{1}{\sigma_n} \sum_{k\geq n} \sigma_n a_k \leq \frac{1}{\sigma_n} \sum_{k\geq n} \sigma_k a_k$. ■

In [56] the author studies the case $f_n \equiv 0$, $\varepsilon_n \equiv 0$ and proves that if the proximal sequence x_n happens to converge strongly to some $x^* \in \mathcal{S}$ then $f(x_n) - f^* = o(\frac{1}{\sigma_n})$. For a sequence f_n this need not be the case. However, it *is* the case if the convergence is sufficiently fast on at least one minimizer.

Corollary 5.9 *Assume the sequence f_n is bounded from below by f^*. If x_n converges strongly to some $x^* \in \mathcal{S} \cap D$ such that $\zeta_n(x^*) = o(\frac{1}{\sigma_n})$. If $\|y_n\| = \mathcal{O}(\frac{1}{\sigma_n})$ and $\varepsilon_n = o(\frac{1}{\sigma_n})$, then $|f_n(x_n) - f^*| = o(\frac{1}{\sigma_n})$.*

Proof. The subdifferential inequality gives $f_n(x^*) \geq f_n(x_n) + \frac{1}{\lambda_n}\langle x_{n-1} - x_n, x^* - x_n\rangle - \varepsilon_n$. Thus

$$f_n(x_n) - f^* \leq \|y_n\| \cdot \|x^* - x_n\| + [f_n(x^*) - f^*] + \varepsilon_n$$

[4] One can prove the following: Let $\psi : \mathbf{R}_+ \to \mathbf{R}_+$ be a continuous function with $\psi(0) = 0$ and let $\{\alpha_n\}$ be a sequence of positive numbers. Assume whether that $\{\alpha_n\}$ is bounded or that ψ satisfies $\psi(x) \leq M(1+x)$ for some $M > 0$. Then $\overline{\alpha_n} \to 0$ implies $\overline{\psi(\alpha_n)} \to 0$. This can be applied to $\psi(x) = \sqrt{x}$.

and the result follows immediately. ■

Observe that by virtue of Theorem 9 in [33] (or Theorem 2.2 in [56]), when $f_n \equiv 0$ and $\varepsilon_n \equiv 0$ we always have $\|y_n\| = \mathcal{O}(\frac{1}{\sigma_n})$, so that Theorem 3.1 in [56] is a consequence of Corollary 5.9 above. It is important to point out that the proof in [56] uses a clever (but unnecessarily sophisticated) argument instead of the trivial observation $\|y_n\| = \mathcal{O}(\frac{1}{\sigma_n})$.

In general, it is difficult to prove that $\|y_n\| = \mathcal{O}(\frac{1}{\sigma_n})$ because the speed depends strongly on the evolution of the sequence $\{f_n\}$ so one has to study each particular case.

5.3 Further results

The following is an immediate consequence of Lemma 5.5 and Proposition 5.8:

Proposition 5.10 *Assume hypotheses* $\mathbf{H_1}$ *and* $\mathbf{H_2}$ *hold and suppose also that* $\sum_{k=1}^{n} \sigma_k(a_k + \varepsilon_k) < \infty$. *If* $\overline{|f_n(x^*) - f^*|} = \mathcal{O}(\frac{1}{\sigma_n})$ *for some* $x^* \in \mathcal{S} \cap D$*, then*

i) $\sum_{k=1}^{\infty} \lambda_k \sigma_k \|y_k\|^2 < \infty$*; and*

ii) The sequence $\{x_n\}$ *is bounded.*

Remark 5.11 *Some comments are in order:*

1. *If* $\lambda_n \sigma_n \geq \lambda > 0$ *then* $\lim_{n\to\infty} \|y_n\| = 0$*; and if* $\lambda_n \geq \lambda > 0$*, then* $\lim_{n\to\infty} \sigma_n \|y_n\|^2 = 0$*, thus* $\|y_n\| = o(\frac{1}{\sqrt{\sigma_n}})$.
2. *From part i) we deduce that* $\overline{\sigma_n \|y_n\|^2} = \mathcal{O}(\frac{1}{\sigma_n})$*, which implies that* $\overline{\sigma_n^2 \|y_n\|^2}$ *is bounded and so is* $\overline{\sigma_n \|y_n\|}$. *This does not fulfill the hypothesis in Theorem 5.9, but is close.*
3. *The rate of approximation is crucial for the boundedness of the sequence* $\{x_n\}$. *To see this, take* $f \equiv 0$ *on* $D = [0, \infty)$ *and* $f_n(x) = \delta_n e^{-x}$*, where the sequence* $\{\delta_n\}$ *decreases to* 0. *It is easy to verify that this approximation satisfies all our assumptions and if* $\delta_n \to 0$ *sufficiently slow, we will have* $x_n \to \infty$.

Finally, we shall present another type of condition assuring that the sequence of velocities converges to zero. This will be done by considering how the shapes of the functions f_n evolve with n. Let us assume there exists a sequence θ_n of nonnegative numbers such that for all $u_n^* \in \partial f_n(u_n)$ and $u_{n+1}^* \in \partial f_{n+1}(u_{n+1})$ we have

$$\langle\, u_{n+1}^* - u_n^*, u_{n+1} - u_n \,\rangle \geq -\theta_n \|u_{n+1} - u_n\|. \tag{5.8}$$

Example 10 *Let* g *be a convex* $\mathcal{C}^1$ *function and suppose its gradient is bounded by a constant* B *on* D. *Take a sequence* $\{r_n\}$ *of positive numbers and define* $f_n = f + r_n g$. *Then it is easy to verify that* (5.8) *is satisfied for* $\theta_n = B|r_n - r_{n-1}|$. *In a similar fashion, if* $\{g_n\}$ *is a family of convex* $\mathcal{C}^1$ *functions such that* $\{\nabla g_n\}$ *converges uniformly, one can take* $\theta_n = \|\nabla g_n - \nabla g_{n-1}\|_\infty$.

Proposition 5.12 *Assume hypotheses* $\mathbf{H_1}$ *and* $\mathbf{H_2}$ *hold. If* $f^* > -\infty$ *and* $\{\theta_n\} \in \ell^1$*, then* $\lim_{n\to\infty} \|y_n\| = 0$.

Proof. Inequality (5.8) implies $\|y_{n+1}\| \leq \|y_n\| + \theta_n$. Since $\{\theta_n\} \in \ell^1$ and $\|y_n\| \geq 0$ the limit $\lim_{n\to\infty} \|y_n\|$ exists. It must be 0 because $\lim_{n\to\infty} \overline{\|y_n\|} = 0$. ■

5.4 A word on finite convergence

For the moment, assume $f_n \equiv f$ and $\varepsilon_n \equiv 0$. If the sequence $\{\lambda_n\}$ of stepsizes is in ℓ^1 ($\sigma_n \to \sigma < \infty$), the trajectory $\{x_n\}$ always converges strongly to a point x_∞ (observe that $\|x_n - x_m\| \le \|y_1\| \cdot |\sigma_n - \sigma_m|$), no matter whether the function has minimizers. Even when the minimizing set is nonempty, if the distance between the initial point x_0 and the minimizing set $\mathcal{S}$ is greater than $\sigma\|y_1\|$, then x_∞ cannot be a point in $\mathcal{S}$. However, one would like to know whether or not it is possible to attain the minimizing set if the initial condition is close enough to $\mathcal{S}$ (alternatively, if σ is large enough). We give a partial answer in terms of the smoothness of the objective function around $\mathcal{S}$.

In the sequel we assume $\mathcal{S}$ is nonempty. The following example includes the case where f is a differentiable function having a L-Lipschitz gradient in some η-neighborhood of $\mathcal{S}$ (consider the maximal monotone operator $A = \nabla f$):

Example 11 *Let A be a maximal monotone operator on H. Set $\mathcal{S} = A^{-1}0$ and let $d(u, \mathcal{S})$ denote the distance from u to $\mathcal{S}$. Assume there exist $\eta > 0$ and $L > 0$ such that if $d(u, \mathcal{S}) < \eta$, then $\sup_{v \in Au} \|v\| \le L\ d(u, \mathcal{S})$. It is clear that the minimizing set cannot be attained in a finite number of steps. If $d(x_N, \mathcal{S}) < \eta$ for some N, the proximal iteration yields $\|x_n - x_{n-1}\| \le \lambda_n L d(x_n, \mathcal{S})$ for all $n > N$. Since $\|x_n - x_{n-1}\| \ge d(x_{n-1}, \mathcal{S}) - d(x_n, \mathcal{S})$ we have $(1 + \lambda_n L) d(x_n, \mathcal{S}) \ge d(x_{n-1}, \mathcal{S})$. Therefore*

$$d(x_n, \mathcal{S}) \ge \left[\prod_{k=N+1}^{n} (1 + \lambda_k L)^{-1} \right] d(x_N, \mathcal{S}) \ge e^{-\sigma_n L} d(x_N, \mathcal{S}).$$

If $\sigma_n \to \sigma < \infty$, the sequence $\{x_n\}$ stays away from $\mathcal{S}$.

On the other hand, if the function f is somehow *pointed* on $\partial\mathcal{S}$ the sequence $\{x_n\}$ will converge if σ is large enough. To see this, let P denote the projection onto the nonempty closed convex set $\mathcal{S}$ and take $r > 0$. Notice that the following statementes are equivalent:

((1)) $\|v\| \ge r$ for all $v \in \partial f(x)$ such that $x \notin \mathcal{S}$;

((2)) $f(x) \ge f^* + r\|x - Px\|$ for all $x \in D$; and

((3)) $r(x - Px)\|x - Px\|^{-1} \in \partial f(Px)$ for all $x \notin \mathcal{S}$.

Proposition 5.13 *Let $f_n \equiv f$ and $\{\varepsilon_n\} \in \ell^1$. Assume there is $r > 0$ such that condition ((1)) holds and let $\sigma_n \to \sigma < \infty$.*

i) If $r^2\sigma > f(x_0) - f^* + \sum \varepsilon_k$, then there are $N \in \mathbf{N}$ and $x^* \in \mathcal{S}$ such that $x_n = x^*$ for all $n \ge N$.

ii) If $r^2\sigma = f(x_0) - f^* + \sum \varepsilon_k$, then $\lim_{n\to\infty} d(x_n, \mathcal{S}) = 0$. Moreover, if the sequence $\{\|y_n\|\}$ turns out to be bounded,[5] then x_n converges to a point in $\mathcal{S}$ as $n \to \infty$.

Proof. If $x_k \notin \mathcal{S}$ for $k = 1, \dots, n$, then $f(x_{k-1}) - f(x_k) + \varepsilon_k \ge \lambda_k \|y_k\|^2 \ge r^2\lambda_k$. Summing up one gets $r^2\sigma_n \le f(x_0) - f(x_n) + \sum \varepsilon_k$.

i) If $x_n \notin \mathcal{S}$ for all $n \in \mathbf{N}$ then $r^2\sigma \le f(x_0) - f^* + \sum \varepsilon_k$, which is a contradiction.

ii) Clearly $\lim_{n\to\infty} f(x_n) = f^*$. By ((2)) we get $\lim_{n\to\infty} \|x_n - Px_n\| = 0$. Finally, if $\{\|y_n\|\}$ is bounded and $\sigma < \infty$ then $\{x_n\}$ is a Cauchy sequence. ■

[5] For instance, if $\sum(\varepsilon_n + \varepsilon_{n-1})\lambda_n^{-1} < \infty$.

Condition ((2)) above was used by Ferris in [51] to prove convergence in a finite number of steps. In the particular case where $\mathcal{S} = \{x^*\}$, the form ((3)) simplifies to read $\overline{B}(0,r) \subseteq \partial f(x^*)$, which is the assumption used by Rockafellar in [97] with the same purpose in that specific case. In the cited works, H is assumed to be $\mathbf{R}^n$ and the sequence $\{\lambda_n\}$ to be bounded from below by a positive constant, which is a more restrictive setting. Although these conditions seem rather restrictive we can easily find a function for which the proximal sequence stays away from $\mathcal{S}$.

Example 12 *Let H be separable and let $\{e_k\}$ be an orthonormal basis for H. Define H_n =span$\{e_1, \ldots, e_n\}$ and consider the function*

$$f(x) = \begin{cases} \sum \frac{1}{k} x_k & \text{if} \quad x^k = \langle x, e_k \rangle \geq 0 \quad \text{for all} \quad k \\ \infty & \text{otherwise.} \end{cases}$$

This function satisfies condition ((1)) on each H_n with $r = 1/n$ and the union $\cup H_n$ is dense in H (so f does not satisfy condition ((1)) in all of H but almost). If we apply the proximal point algorithm with $x_0 \in dom(f)$ and $\sigma_n \to \sigma$ we find that the k-th component of the term x_n satisfies

$$(x_n)^k \geq (x_{n-1})^k - \lambda_n/k \geq (x_0)^k - \sigma_n/k$$

and so

$$\liminf_{n\to\infty} (x_n)^k \geq (x_0)_k - \sigma/k$$

for each k. If x_0 is selected so that $\sup_k k(x_0)^k = \infty$ we will have $\liminf_{n\to\infty}(x_n)^k > 0$ for some k no matter how large σ is. Therefore x_n does not converge to 0, which is the unique minimizer of f. Observe also that one can select x_0 arbitrarily close to 0, with the same result.

Remark 5.14 *If the sequence $\{f_k\}$ is not constant, one can require each f_k to satisfy the hypothesis of Proposition 5.13 for a certain r_k. The same argument leads us to*

$$\sum_{k=1}^{n} \lambda_k \|y_k\|^2 \leq f_0(x_0) - f_n(x_n) + \sum_{k=1}^{n} (\varepsilon_k + a_{k-1}).$$

The difficulty lies in the fact that no lower bound can be obtained for $\|y_k\|$ because the point x_k could be precisely the minimizer of f_k.

5.5 Numerical tests

Let us go back to Example 9. Since the sequence f_n is decreasing we can take $a_n \equiv 0$. If $x^* \in \mathcal{S}$ and $\sum \lambda_k \eta_k < \infty$, then $|\overline{f_n(x^*)} - f^*| = \mathcal{O}(\frac{1}{\sigma_n})$. For the exact subdifferentials Proposition 5.8 yields $|f_n(x_n) - f^*| = \mathcal{O}(\frac{1}{\sigma_n})$.

Let $g_1, g_2, g_3 \in \Gamma_0(\mathbf{R}^2)$ be given by $g_1(x,y) = x + 2y - 1$, $g_2(x,y) = 2x - y$ and $g_3(x,y) = 1 - x$. The function $f = \max\{g_1, g_2, g_3\}$ is bounded below, continuous and coercive. Hence $\mathcal{S} \neq \emptyset$. In fact, $\mathcal{S} = \{(1/2, 1/2)\}$ and $f^* = 1/2$. Since the hypotheses of Corollary 5.7 also hold we must have $x_n \to (1/2, 1/2)$. Strong convergence then implies $\sigma_n f_n(x_n) \to 0$ as $n \to \infty$ by virtue of Theorem 5.9 provided $\eta_n \to 0$ fast enough.

Figure 1 shows the first 30 iterations of the algorithm, Figure 2 shows $f_n(x_n)$ vs n and Figure 3 shows $\sigma_n(f_n(x_n) - f^*)$ vs n. Here $\lambda_n \equiv 1$, $\eta_n = 1/n^2$ and $x_0 = (0,0)$. Figure 4, where we have $\sigma_n^2(f_n(x_n) - f^*)$ vs n, suggests that this quantity converges as $n \to \infty$, so that would be the fastest rate one could expect.

Figure 1

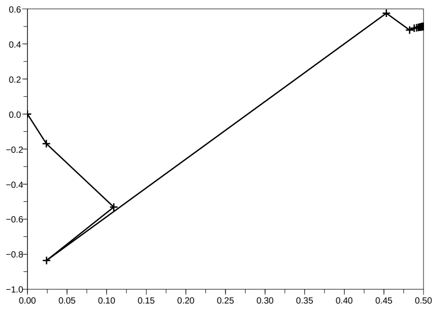


Figure 2

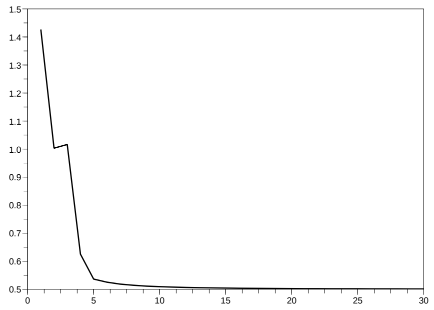


Figure 3

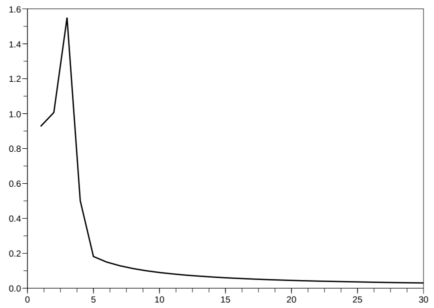
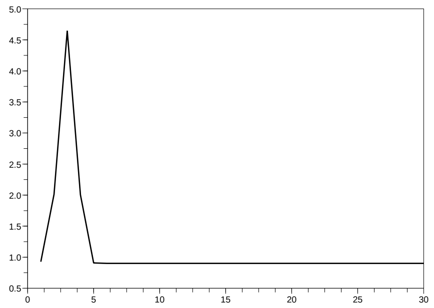


Figure 4



Chapter 6

Asymptotic equivalence and Kobayashi-type estimates for nonautonomous monotone operators in Banach spaces

Introduction

Motivated by either the existence or the algorithmic approximation of solutions to a differential inclusion problem of the type

$$\dot{x} + A(t)x \ni 0, \tag{6.1}$$

where $A(t)$ is a possibly time-dependent m-accretive operator with domain in a Banach space, several authors have considered some special implicit discretization schemes.

In the autonomous case where $A(t) \equiv A$, Crandall and Liggett introduced in [47] the following limit:

$$S(t)x_0 = \lim_{n\to\infty} \left(I + \frac{t}{n}A\right)^{-n} x_0 = \lim_{n\to\infty} (J^A_{t/n})^n x_0, \tag{6.2}$$

where $J^A_\lambda = (I + \lambda A)^{-1}$ is the resolvent of A. Under some closedness assumptions on the operator A, they proved that this limit exists and defines a strongly continuous semigroup $\{S(t)\}_{t\geq 0}$ on X such that $x(t) := S(t - t_0)x_0$ is the strong solution to $\dot{x} + Ax \ni 0$ that satisfies $x(t_0) = x_0$. They also provided some estimates on $\|(J^A_\lambda)^n x_0 - (J^A_\mu)^m x_0\|$, and established the Lipschitz continuity of the solution. Later Kobayashi recovered in [66] similar existence results for the autonomous case, with fundamental improvements concerning certain estimates and some continuity properties. In fact, he constructed sequences of approximate solutions which converge in an appropriate sense to a solution to the differential inclusion. The key argument is an inequality that provides an estimate for the distance between arbitrary points of two independent sequences generated by the so called *proximal iterations*. More precisely, in the case where $x_k = J^A_{\lambda_k} x_{k-1}$ and $\widehat{x}_l = J^A_{\widehat{\lambda}_l} \widehat{x}_{l-1}$ with (possibly nonconstant) stepsizes $\{\lambda_k\} \subset (0, \Lambda]$ and $\{\widehat{\lambda}_l\} \subset (0, \widehat{\Lambda}]$, Kobayashi's inequality establishes that:

$$\|x_k - \widehat{x}_l\| \leq \|x_0 - u\| + \|\widehat{x}_0 - u\| + |\!|\!| Au |\!|\!| \sqrt{(\sigma_k - \widehat{\sigma}_l)^2 + \Lambda\sigma_k + \widehat{\Lambda}\widehat{\sigma}_l}, \tag{6.3}$$

where $\sigma_k = \sum_{i=1}^k \lambda_i$ (similar for $\widehat{\sigma}_l$), u is any element in the domain of A and $|\!|\!| Au |\!|\!| = \inf_{[u,v]\in A} \|v\|$. Kobayashi showed that such a remarkable inequality also holds for some *inexact* proximal iterations by adding certain terms in the right-hand side of the estimate (see

Remark 6.3 below).

Kobayashi's inequality is a powerful tool. Some Lipschitz continuity properties of a limit as (6.2) follow easily from such an inequality, giving explicit Lipschitz constants in terms of the data. Moreover, passing to the limit in only one of the sequences, it is possible to compare the continuous and discrete trajectories, namely, we have an estimate of the type

$$\|x_k - x(t)\| \leq \|x_0 - u\| + |\!|\!|Au|\!|\!| \sqrt{(\sigma_k - t)^2 + \Lambda\sigma_k}.$$

In [84], and still for the autonomous case, Miyadera and Kobayasi introduced the notion of *almost-orbit*, which is a kind of approximate solution to $\dot{x} + Ax \ni 0$. They used Kobayashi's inequality to prove that the continuous path constructed by linear interpolation of some proximal iterations is indeed an almost-orbit for the semigroup generated by the operator A. A converse result is given in [70]. It is known that several asymptotic properties of the orbits are inherited by the almost-orbits (see [90], [67] and [5]). More recently, Güler proves in [56] that if the operator A is the *subdifferential* of a closed, proper and convex function in a Hilbert space, both the continuous trajectory defined by the semigroup and the discrete proximal iterations either converge or diverge simultaneously; this being valid both for the strong and the weak topologies. Besides some technical difficulties, Güler's proof relies on Kobayashi's inequality together with some clever ideas borrowed from Passty, who had already obtained in [90] a similar conclusion under different but complementary assumptions. The powerful results in [90] and [56] reveal a sort of equivalence in the asymptotic behavior of the continuous and discrete trajectories.

Concerning the nonautonomous case, Crandall and Pazy provided in [49] a suitable generalization of the limiting formula (6.2). On the other hand, under some additional conditions on the operator-valued function $t \mapsto A(t)$, Kobayasi *et al.* gave in [67, Lemma 3.4] a first nonautonomous version of the original Kobayashi inequality. Then, they obtained important properties of the corresponding continuous dynamics by passing to the limit in an appropriate manner. However, their nonautonomous Kabayashi-type inequality and the resulting estimates in [67] are rather involved, and based on some extrapolations and not on optimal bounds.

In this work we pretend to provide two different and independent contributions to the existing theory and also show, through an example, how these two instruments can be combined.

Keeping this in mind, one of the goals of this paper is to prove some alternative nonautonomous Kobayashi-type estimates which are in general sharper than those provided in [67], and valid in a more general setting with respect to the properties required on the family of operators. With our approach, no sophisticated mathematical tools are needed, and the consequent estimates for discrete proximal iterates as well as continuous trajectories show explicitly and separately the different effects of the time dependence.

Our second goal is to extend a theory of asymptotic equivalence started by Passty in [90] to a larger class or families of operators to include the nonautonomous case. This theory allows to ensure that some "approximate trayjectories" enjoy the same asymptotic properties as the orbits of some evolution systems. To illustrate how these two theories can be combined, we exploit the sharper estimations mentioned above to obtain equivalence results for the asymptotic behavior of continuous and discrete trajectories, similar but complementary to those one can find in [90, 70, 56].

This paper is organized as follows. In section 2 we recall some definitions and results concerning monotone or accretive operators and evolution equations. We also set some notation and state the main hypotheses along with some examples. Section 3 contains an abstract asymptotic equivalence result for evolution systems that generalizes [90, Lemma 1], which we apply then to a pair of differential inclusions in a setting that includes the quasi-autonomous case (see [27]). The second application uses some new Kobayashi-type estimates, which we state in section 4 and prove in section 5.

6.1 Preliminaries

Let $(X, \|\cdot\|)$ be a Banach space and denote by X^* its dual, which is endowed with the dual norm defined by $\|f\|_* = \sup_{\|u\|\leq 1} f(u)$. The duality product $\langle\cdot,\cdot\rangle : X \times X^* \to \mathbf{R}$ is defined by $\langle u, f\rangle = f(u)$ for all $u \in X$ and $f \in X^*$. The duality mapping $\mathcal{J} : X \rightrightarrows X^*$ is given by

$$\mathcal{J}(u) = \{\ f \in X^* \mid \|f\|_* = \|u\| \text{ and } \langle u, f\rangle = \|u\|^2\ \}.$$

Given a set-valued mapping $A : X \rightrightarrows X$, its domain is given by $D(A) = \{u \in X \mid Au \neq \emptyset\}$. For convenience of notation, sometimes we identify A with its graph by writing $[u, v] \in A$ for $v \in Au$. If $u \in D(A)$ then we set

$$|\!|\!|Au|\!|\!| = \inf_{[u,v]\in A} \|v\|.$$

A mapping $A : X \rightrightarrows X$ is said to be a *monotone operator* if for all $[u_1, v_1], [u_2, v_2] \in A$ there exists $f \in \mathcal{J}(u_1 - u_2)$ such that

$$\langle v_1 - v_2, f\rangle \geq 0. \tag{6.4}$$

A monotone operator is said to be *maximal* if its graph is not properly contained in the graph of any other monotone operator. Let I be the identity mapping in X. For $\lambda > 0$, the *resolvent* of A is defined as the mapping $J^A_\lambda = (I + \lambda A)^{-1}$. An operator A is said to be *accretive* if for all $\lambda > 0$, and $[u_1, v_1], [u_2, v_2] \in A$ one has

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

This implies that J^A_λ is a single-valued nonexpansive mapping. If, in addition, the range of $I + \lambda A$ equals X for all $\lambda > 0$, the operator A is said to be *m-accretive*. A well-known consequence of general results in [82, 63] is the following:

i) A is monotone if, and only if, A is accretive.

ii) If A is m-accretive, then it is maximal monotone. The converse is true in Hilbert space but not in general Banach spaces (see [60, Chap. V] for a counterexample).

Let D be a nonempty subset of X and define

$$\mathcal{M}(D) = \{A : X \rightrightarrows X \mid A \text{ is } m\text{-accretive and } D(A) = D\}.$$

A collection of four sequences $(\{x_k\}, \{\lambda_k\}, \{A_k\}, \{\varepsilon_k\})$ with $\{A_k\} \subset \mathcal{M}(D)$ is said to be a *discrete proximal scheme* if $\lambda_k > 0$, and

$$(x_k - x_{k-1})/\lambda_k + A_k x_k \ni \varepsilon_k, \tag{6.5}$$

for all $k \geq 1$. The corresponding sequence $\{x_k\}$ is said to be a *discrete proximal trajectory* starting from the point $x_0 \in X$ and generated by $(\{\lambda_k\}, \{A_k\}, \{\varepsilon_k\})$.

Next, given points $x \neq \widehat{x}$ and $v, \widehat{v}$ in X, define

$$\Delta([x, v], [\widehat{x}, \widehat{v}]) = \inf_{f\in\mathcal{J}(x-\widehat{x})} \frac{\langle\widehat{v} - v, f\rangle}{\|x - \widehat{x}\|}. \tag{6.6}$$

If $x = \widehat{x}$ then we set $\Delta([x,v],[\widehat{x},\widehat{v}]) = 0$. Notice that $\Delta([x,v],[\widehat{x},\widehat{v}]) \leq \|v - \widehat{v}\|$. If $[x,v],[\widehat{x},\widehat{v}] \in A$ for some $A \in \mathcal{M}(D)$ then, by monotonicity, we have that $\Delta([x,v],[\widehat{x},\widehat{v}]) \leq 0$.

Let us consider two sequences $\{A_k\}$ and $\{\widehat{A}_l\}$ in $\mathcal{M}(D)$. As in [67], we shall assume that the following condition holds:

$$\forall k,l \geq 1,\ \exists \Theta_{k,l} \geq 0,\ \forall\,[x,v] \in A_k,\ \forall\,[\widehat{x},\widehat{v}] \in \widehat{A}_l,\ \Delta([x,v],[\widehat{x},\widehat{v}]) \leq \Theta_{k,l}. \tag{6.7}$$

Remark 6.1 *Condition* (6.7) *above was introduced in [67] to determine the existence of weak solutions for a nonautonomous differential inclusion. The same hypothesis was used in [88] for a parabolic system describing the behavior of the HIV virus in the human body and the reaction to treatment by chemotherapy (see also [104], where the author first introduces a model without treatment). See* (6.11) *below for a continuous version of this condition.*

Example 13 *Let $A \in \mathcal{M}(D)$ and $B : X \rightrightarrows X$ a strongly monotone mapping such that $\|B\|_{\infty,D} := \sup_{x\in D}\sup_{v\in B(x)} \|v\| < \infty$. Set $A(r) = A + rB \in \mathcal{M}(D)$. Given two sequences $\{r_k\}$ and $\{\widehat{r}_l\}$ of positive numbers define accordingly $A_k = A(r_k)$ and $\widehat{A}_l = A(\widehat{r}_l)$ for $k,l \geq 1$. Then it is easy to verify that* (6.7) *is satisfied for $\Theta_{k,l} = |r_k - \widehat{r}_l|\|B\|_{\infty,D}$.*

Let $A, \widehat{A} : [0,\infty) \longrightarrow \mathcal{M}(D)$. For $m \in \mathbf{N}$ and $t > t_0 \geq 0$, consider the *finite* discrete proximal trajectories $\{x_k\}_{k=0}^m$ and $\{\widehat{x}_l\}_{l=0}^m$ defined by

$$x_0 = u \text{ and } x_k = \left(I - \tfrac{t-t_0}{m} A(t_0 + \tfrac{k(t-t_0)}{m})\right)^{-1} x_{k-1}, \quad \text{for } k = 1,\dots,m. \tag{6.8a}$$

$$\widehat{x}_0 = u \text{ and } \widehat{x}_l = \left(I - \tfrac{t-t_0}{m} \widehat{A}(t_0 + \tfrac{l(t-t_0)}{m})\right)^{-1} \widehat{x}_{l-1}, \quad \text{for } l = 1,\dots,m. \tag{6.8b}$$

From now on, we assume that x_m and $\widehat{x}_m$ converge[1] in X as $m \to \infty$, and we respectively denote by $U(t,t_0)u$ and $\widehat{U}(t,t_0)u$ their limits, that is, for all $u \in D$ and $t > t_0 \geq 0$ we set

$$U(t,t_0)u = \lim_{m\to\infty} x_m = \lim_{m\to\infty} \prod_{k=1}^{m} \left(I - \tfrac{t-t_0}{m} A(t_0 + \tfrac{k(t-t_0)}{m})\right)^{-1} u \tag{6.9a}$$

$$\widehat{U}(t,t_0)u = \lim_{m\to\infty} \widehat{x}_m = \lim_{m\to\infty} \prod_{l=1}^{m} \left(I - \tfrac{t-t_0}{m} \widehat{A}\left(t_0 + \tfrac{l(t-t_0)}{m}\right)\right)^{-1} u. \tag{6.9b}$$

We set $U(t_0,t_0)u = \widehat{U}(t_0,t_0)u = u$. In addition, we will assume that for each $u \in D$, the functions

$$t \mapsto |\!|\!|A(t)u|\!|\!| \text{ and } t \mapsto |\!|\!|\widehat{A}(t)u|\!|\!| \text{ are bounded and locally Riemann-integrable.} \tag{6.10}$$

Sufficient conditions on $\{A(t)\}_{t\in[0,\infty)}$ ensuring that $U(t,t_0)u$ is well defined are given in [49]. The function $U(\cdot,t_0)u$ is said to be a *weak* or *DS-limit solution*[2] of inclusion (6.1). In a time-independent domain these generalized solutions happen to coincide with *integral solutions* in the sense of Bénilan ([27]) under hypothesis (6.11) below (see Theorem 2.4 in [67]). The trajectory $t \mapsto U(t,t_0)u$ can also be proved to satisfy (6.1) under supplementary assumptions. We shall not go further on this matter here but only mention that such conditions imply in particular the continuity of $[0,\infty) \ni t \mapsto |\!|\!|A(t)u|\!|\!|$, hence (6.10). Defining evolution systems, such as (6.9a) and (6.9b), by a limiting process involving (piecewise constant interpolations of) certain discretizations of the evolution equation and its variants is a common practice. Classical references are [47], [49] and [66]. More recent examples can be found in [88] (discretization of the differential inclusion), [12] (approximations of the

[1] As we shall see, by virtue of the weak lower-semicontinuity of the norm, it suffices to assume weak convergence.

[2] DS for *discrete scheme*. The term was used in [67] and introduced in [66] in the autonomous setting $A(t) \equiv A$.

nonautonomous operator) or [54] (discretization of the weak solution equation).

Finally, we assume the continuous version of (6.7): there exists a bounded Riemann-integrable function $\Theta : [0,\infty) \times [0,\infty) \to [0,\infty)$ satisfying

$$\forall t,s \in [0,\infty),\ \forall [x,v] \in A(t),\ \forall [\widehat{x},\widehat{v}] \in \widehat{A}(s),\ \Delta([x,v],[\widehat{x},\widehat{v}]) \leq \Theta(t,s), \tag{6.11}$$

for $\Delta(\cdot,\cdot)$ given by (6.6) (see, for instance, Example 13). This is precisely Hypothesis $(\mathbf{H};C)$ in [67] but no continuity hypotheses are made here.

6.2 Kobayashi-type estimates

In this section we first derive bounds for the distance between arbitrary points in two discrete proximal trajectories. Next we use this estimation to compare a discrete trajectory with a continuous one of the form (6.9). Finally we are able to provide an estimation for two continuous trajectories.

6.2.1 Discrete-discrete estimate

The main result of this section is the following:

Theorem 6.2 *Let $(\{x_k\},\{\lambda_k\},\{A_k\},\{\varepsilon_k\})$ and $(\{\widehat{x}_l\},\{\widehat{\lambda}_l\},\{\widehat{A}_l\},\{\widehat{\varepsilon}_l\})$ be two discrete proximal schemes* (6.5)*. If* (6.7) *holds then for every $u \in D$ and for all k,l we have that*

$$\begin{aligned} \|x_k - \widehat{x}_l\| \quad &\leq \quad \|x_0 - u\| + \|\widehat{x}_0 - u\| + \alpha_{k,l} + \beta_{k,l} \\ &\qquad + \sqrt{(\gamma_k(u) - \widehat{\gamma}_l(u))^2 + \delta_k(u) + \widehat{\delta}_l(u) + \eta_{k,l}(u)}, \end{aligned} \tag{6.12}$$

where

$$\gamma_k(u) = \sum_{i=1}^{k} \lambda_i |\!|\!| A_i u |\!|\!|, \qquad \delta_k(u) = \sum_{i=1}^{k} \lambda_i^2 |\!|\!| A_i u |\!|\!|^2 \qquad \textit{(similar for } \widehat{\gamma}_l \textit{ and } \widehat{\delta}_l\textit{)}, \tag{6.13}$$

and $\alpha_{k,l}$, $\beta_{k,l}$, and $\eta_{k,l}$ are defined recursively as follows:

$$\left\{ \begin{array}{l} \alpha_{k,0} = e_k = \sum\limits_{i=1}^{k} \lambda_i \|\varepsilon_i\|,\ \alpha_{0,l} = \widehat{e}_l = \sum\limits_{i=j}^{l} \widehat{\lambda}_j \|\widehat{\varepsilon}_j\|, \ \textit{and} \\ \alpha_{k,l} = \frac{\widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l} \alpha_{k-1,l} + \frac{\lambda_k}{\lambda_k + \widehat{\lambda}_l} \alpha_{k,l-1} + \frac{\lambda_k \widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l} \|\varepsilon_k - \widehat{\varepsilon}_l\|. \end{array} \right. \tag{6.14}$$

$$\left\{ \begin{array}{l} \beta_{k,0} = \beta_{0,l} = 0, \ \textit{and} \\ \beta_{k,l} = \frac{\widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l} \beta_{k-1,l} + \frac{\lambda_k}{\lambda_k + \widehat{\lambda}_l} \beta_{k,l-1} + \frac{\lambda_k \widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l} \Theta_{k,l}. \end{array} \right. \tag{6.15}$$

$$\left\{ \begin{array}{l} \eta_{k,0}(u) = \eta_{0,l}(u) = 0, \ \textit{and} \\ \eta_{k,l}(u) = \frac{\widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l} \eta_{k-1,l}(u) + \frac{\lambda_k}{\lambda_k + \widehat{\lambda}_l} \eta_{k,l-1}(u) \\ \qquad\qquad + \frac{\lambda_k \widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l} 2(\gamma_k - \widehat{\gamma}_l) \left[|\!|\!| \widehat{A}_l u |\!|\!| - |\!|\!| A_k u |\!|\!| \right]. \end{array} \right. \tag{6.16}$$

Remark 6.3 *Notice that $\alpha_{k,l} \leq e_k + \widehat{e}_l$ for all $k,l \geq 0$.*

Remark 6.4 *In the specific case where $A_k \equiv \widehat{A}_l \equiv A$, we can take $\Theta_{k,l} \equiv 0$ and get $\beta_{k,l} \equiv 0$. We also have $\eta_{k,l} \equiv 0$. Then* (6.12) *amounts to*

$$\|x_k - \widehat{x}_l\| \leq \|x_0 - u\| + \|\widehat{x}_0 - u\| + \alpha_{k,l} + |\!|\!| Au |\!|\!| \sqrt{(\sigma_k - \widehat{\sigma}_l)^2 + \tau_k + \widehat{\tau}_l},$$

where $\sigma_k = \sum_{i=1}^{k} \lambda_i$ and $\tau_k = \sum_{i=1}^{k} \lambda_i^2$ (similar for $\widehat{\sigma}_l$ and $\widehat{\tau}_l$). We thus recover the inequality obtained by Kobayashi in [66], where $\alpha_{k,l}$ is replaced with $e_k + \widehat{e}_l$ (see Remark 6.3).

Remark 6.5 *Kobayashi* et al. *considered a similar problem in [67] in the interesting specific case where both* $\{A_k\}$ *and* $\{\widehat{A}_l\}$ *are generated by a continuously parameterized family* $\{A(t)\}_{t\in I}$. *Their estimate resembles (6.12) but there is a fundamental difference: Our estimation keeps track of all the information contained in the sequences* $\{|\!|\!| A_k u |\!|\!|\}$ *and* $\{|\!|\!| \widehat{A}_l u |\!|\!|\}$, *while their bound involves the value* $|\!|\!| A(t_0)u |\!|\!|$ *for one specific* t_0 *and makes some extrapolations using a modulus of continuity of the function* $t \mapsto |\!|\!| A(t)u |\!|\!|$. *As a consequence, their estimation looks simpler. However, it is clear that using our approach we get a sharper bound in a more general setting. In this sense, the two results are complementary. Another advantage of our estimation is that, as we shall see in the next sections, it is easy to derive information about a continuous-in-time trajectory by taking limits.*

6.2.2 Discrete-continuous estimate

Theorem 6.6 *Let* $A : [0,\infty) \to \mathcal{M}(D)$ *and* $\widehat{A} : [0,\infty) \to \mathcal{M}(D)$ *satisfying* (6.9), (6.10) *and* (6.11). *Let* $t > t_0 \geq 0$ *and take a discrete proximal scheme* $(\{x_k\},\{\lambda_k\},\{A_k\},\{\varepsilon_k\})$ *with* $A_k := A(s_0+\sigma_k)$ *for* $\sigma_k = \sum_{i=1}^k \lambda_i$ *and some* $s_0 \in [0,\infty)$. *For every* $u \in D$ *and* $k \in \mathbf{N}$, *we have that*

$$\|x_k - \widehat{U}(t,t_0)u\| \leq \|x_0 - u\| + \alpha_k + \beta_k + \sqrt{[\gamma_k(u) - \widehat{\mathcal{A}}_u(t,t_0)]^2 + \delta_k(u) + \eta_k(u)}, \tag{6.17}$$

where

$$\widehat{\mathcal{A}}_u(t,t_0) := \int_{t_0}^t |\!|\!| \widehat{A}(\xi)u |\!|\!| \; d\xi \tag{6.18}$$

and $\alpha_k = \sum_{i=1}^k \lambda_i\|\varepsilon_i\|$, $\beta_k = \limsup_{m\to\infty} \beta_{k,m} < \infty$ *and* $\eta_k(u) = \limsup_{m\to\infty} \eta_{k,m}(u) < \infty$ *for the sequences given by* (6.13)-(6.16) *with*

$$\widehat{A}_l = \widehat{A}\left(t_0 + \tfrac{l(t-t_0)}{m}\right) \text{ and } \Theta_{k,l} = \Theta\left(s_0 + \sigma_k, t_0 + \tfrac{l(t-t_0)}{m}\right).$$

Remark 6.7 *Taking* $k = 0$ *and* $x_0 = u$ *we get* $\|u - \widehat{U}(t,t_0)u\| \leq \widehat{\mathcal{A}}_u(t,t_0)$.

6.2.3 Continuous-continuous estimate

Theorem 6.8 *Let* $u \in D$. *Suppose* (6.9), (6.10) *and* (6.11) *hold and take* $t - t_0 \leq s - s_0$. *We have*

$$\|U(s,s_0)u - \widehat{U}(t,t_0)u\| \leq \sqrt{2\left[\mathcal{A}_u(s,s_0) - \widehat{\mathcal{A}}_u(t,t_0)\right]^2 - [\mathcal{A}_u(t_0+s-t,s_0)]^2 + \int_0^\tau \Theta(s-\xi,t-\xi)\; d\xi}.$$

The inequality above is an equality if one takes $\widehat{A}(t)x = A(t)x \equiv c \in X$.

Remark 6.9 *Set* $\widehat{A} = A$. *If the function* $t \mapsto |\!|\!| A(t)u |\!|\!|$ *is nonincreasing then*

$$\sqrt{2\left[\mathcal{A}_u(s,s_0) - \mathcal{A}_u(t,t_0)\right]^2 - [\mathcal{A}_u(t_0+s-t,s_0)]^2} \leq |\mathcal{A}_u(s,s_0) - \mathcal{A}_u(t,t_0)|.$$

Theorem 6.8 shows that the function U automatically inherits continuity properties from the function Θ. For instance, if $\Theta(t,s)$ is locally bounded by the difference $|t-s|$, the function U is locally Lipschitz-continuous in the pair (t_0,t).

Corollary 6.10 *Let* C *be a compact subset of the triangle* $t \geq s \geq 0$ *and assume* $\Theta(t,s) \leq L|t-s|$ *on* C. *For each* $u \in D$, *the function* $(t,s) \mapsto U(t,s)u$ *is Lipschitz-continuous with constant*[3]

$$\sqrt{2} \sup_{\delta_1 \leq \xi \leq \delta_2} |\!|\!| A(\xi)u |\!|\!| + L(\delta_2 - \delta_1),$$

where $\delta_1 = \min\{\ s \mid (t,s) \in C\ \}$, $\delta_2 = \max\{\ t \mid (t,s) \in C\ \}$.

[3]For the ℓ^1 norm.

Lipschitz-continuity of the function $t \mapsto U(t, t_0)u$ had already been proved in [49] and [67]. In the first article, even for monotone operators, their constant depends exponentially on the length of the interval (with, sometimes, a linear-affine coefficient) unless the function $t \mapsto A(t)$ is constant. In the second cited article, the authors prove Lipschitz continuity for *weak* solutions and find a constant depending on $\|\|A(0)u\|\|$ and a global bound for Θ. Our Proposition 6.8 shows that the constant, in fact, depends on the local (rather than global) behavior of Θ and $\|\|A(\cdot)u\|\|$. Moreover, if the function $t \mapsto A(t)$ is not constant, their constant grows linearly with the length of the interval. Therefore one cannot get a global Lipschitz constant even if the function has a very small variation.

Example 14 *Let $X = \mathbf{R}$. With the notation introduced in Example 13 take $A \equiv 0$, $B \equiv 1$ and parameterize ε by a nonincreasing positive function $\varepsilon(t)$. We have $\Theta(t, s) = |\varepsilon(t) - \varepsilon(s)|$ and $\|\|A(t)u\|\| = \varepsilon(t)$ for all u. For simplicity set $t_0 = s_0 = 0$ and $t \leq s$. If the function $\varepsilon(\cdot)$ is Lipschitz-continuous with constant L, we can apply the results in [67] on the interval $[0, T]$ to get $LT + 2\varepsilon(0) - \varepsilon(T)$ as a Lipschitz constant for U, which tends to ∞ with the length of the interval. On the other hand, according to our Proposition 6.8 and Remark 6.9 we have $\|U(s, 0)u - U(t, 0)u\| \leq \int_t^s \varepsilon(\xi)\, d\xi + \int_0^t [\varepsilon(t-\xi) - \varepsilon(s-\xi)]\, d\xi = \int_0^{s-t} \varepsilon(\xi)\, d\xi \leq \varepsilon(0)|s-t|$.*

6.3 Proofs of the estimates

This section contains the proofs of the estimations contained in Subsections 6.2.1, 6.2.2 and 6.2.3, respectively.

6.3.1 Discrete-discrete

In order to prove Theorem 6.2 we first establish two auxiliary lemmas.

Lemma 6.11 *Let $(\{x_k\}, \{\lambda_k\}, \{A_k\}, \{\varepsilon_k\})$ be a discrete proximal scheme. For every $u \in D$ and all $k \geq 1$ we have*

$$\|x_k - u\| \leq \|x_0 - u\| + \gamma_k(u) + e_k. \tag{6.19}$$

Proof. Let $u \in D$. We shall prove the result by induction. The estimate (6.19) is trivially satisfied for $k = 0$. Suppose (6.19) holds for some $k \geq 0$ and set

$$v_{k+1} = (x_{k+1} - x_k)/\lambda_{k+1} - \varepsilon_{k+1} \in -A_{k+1}x_{k+1}.$$

Take $y \in X$ such that $[u, -y] \in A_{k+1}$. Since $\|x_{k+1} - u\|^2 = \langle x_{k+1} - u, f\rangle$, for any $f \in \mathcal{J}(x_{k+1} - u)$, we have

$$\begin{aligned}
\|x_{k+1} - u\|^2 &= \langle x_{k+1} - u + \lambda_{k+1}(y - v_{k+1}) - \lambda_{k+1}\varepsilon_{k+1}, f\rangle - \lambda_{k+1}\langle y - v_{k+1}, f\rangle \\
&\quad + \lambda_{k+1}\langle \varepsilon_{k+1}, f\rangle \\
&\leq \|x_k - u + \lambda_{k+1}y\|\|x_{k+1} - u\| - \lambda_{k+1}\langle y - v_{k+1}, f\rangle \\
&\quad + \lambda_{k+1}\|\varepsilon_{k+1}\|\|x_{k+1} - u\|.
\end{aligned}$$

The monotonicity of A_{k+1} and the induction hypothesis imply

$$\begin{aligned}
\|x_{k+1} - u\| &\leq \|x_k - u + \lambda_{k+1}y\| + \lambda_{k+1}\|\varepsilon_{k+1}\| \\
&\leq \|x_k - u\| + \lambda_{k+1}\|y\| + \lambda_{k+1}\|\varepsilon_{k+1}\| \\
&\leq \|x_0 - u\| + \gamma_k + e_k + \lambda_{k+1}\|y\| + \lambda_{k+1}\|\varepsilon_{k+1}\| \\
&\leq \|x_0 - u\| + \gamma_k + \lambda_{k+1}\|y\| + e_{k+1}.
\end{aligned}$$

As y was arbitrarily chosen so that $[u, -y] \in A_{k+1}$, we conclude that

$$\|x_{k+1} - u\| \leq \|x_0 - u\| + \gamma_k + \lambda_{k+1}\|\|A_{k+1}u\|\| + e_{k+1} = \|x_0 - u\| + \gamma_{k+1} + e_{k+1},$$

which completes the proof of Lemma 6.11. ■

Lemma 6.12 *Let $x \neq \widehat{x}$, $v, \widehat{v}, \varepsilon, \widehat{\varepsilon} \in X$ and λ, $\widehat{\lambda} \in (0, \infty)$. Then*

$$\begin{aligned}(\lambda + \widehat{\lambda})\|x - \widehat{x}\| \quad &\leq \quad \lambda\|\widehat{x} + \widehat{\lambda}(\widehat{v} - \widehat{\varepsilon}) - x\| + \widehat{\lambda}\|x + \lambda(v - \varepsilon) - \widehat{x}\| \\ &\qquad + \lambda\widehat{\lambda}\Delta([x, v], [\widehat{x}, \widehat{v}]) + \lambda\widehat{\lambda}\|\varepsilon - \widehat{\varepsilon}\|. \end{aligned} \tag{6.20}$$

Proof. Suppose $x \neq \widehat{x}$; otherwise there is nothing to prove. If $f \in \mathcal{J}(x - \widehat{x})$ then

$$\begin{aligned}(\lambda + \widehat{\lambda})\|x - \widehat{x}\|^2 \quad &= \quad \lambda\langle \widehat{x} - x,\ -f\rangle + \widehat{\lambda}\langle x - \widehat{x},\ f\rangle \\ &= \quad \lambda\langle \widehat{x} + \widehat{\lambda}(\widehat{v} - \widehat{\varepsilon}) - x,\ -f\rangle + \widehat{\lambda}\langle x + \lambda(v - \varepsilon) - \widehat{x},\ f\rangle \\ &\qquad + \lambda\widehat{\lambda}\langle \widehat{v} - v,\ f\rangle + \lambda\widehat{\lambda}\langle \varepsilon - \widehat{\varepsilon},\ f\rangle \\ &\leq \quad \Big[\lambda\|\widehat{x} + \widehat{\lambda}(\widehat{v} - \widehat{\varepsilon}) - x\| + \widehat{\lambda}\|x + \lambda(v - \varepsilon) - \widehat{x}\|\Big]\ \|x - \widehat{x}\| \\ &\qquad + \lambda\widehat{\lambda}\langle \widehat{v} - v,\ f\rangle + \lambda\widehat{\lambda}\|\varepsilon - \widehat{\varepsilon}\|\|x - \widehat{x}\|. \end{aligned}$$

Dividing by $\|x - \widehat{x}\|$ and taking infimum with respect to f, we conclude the proof of Lemma 6.12. ■

Proof of Theorem 6.2. For simplicity of notation, we drop the dependence on u. Setting

$$\omega_{k,l} = (\gamma_k - \widehat{\gamma}_l)^2 + \delta_k + \widehat{\delta}_l + \eta_{k,l},$$

we must prove that for all $k, l \geq 0$:

$$\omega_{k,l} \geq 0 \quad \text{and} \quad \|x_k - \widehat{x}_l\| \leq \|x_0 - u\| + \|\widehat{x}_0 - u\| + \alpha_{k,l} + \beta_{k,l} + \sqrt{\omega_{k,l}}. \tag{6.21}$$

We will argue by induction on the pair (k, l).

First, by virtue of Lemma 6.11, we have that

$$\|x_k - \widehat{x}_0\| \leq \|x_k - u\| + \|u - \widehat{x}_0\| \leq \|x_0 - u\| + \gamma_k + e_k + \|\widehat{x}_0 - u\|,$$

which proves (6.21) for any pair $(k, 0)$ with $k \geq 0$ because $\widehat{\gamma}_0 = \widehat{\delta}_0 = \eta_{k,0} = \beta_{k,0} = 0$ so that $\gamma_k^2 \leq \omega_{k,0}$. Similarly, (6.21) holds with $(0, l)$ for all $l \geq 0$.

Now suppose (6.21) is true for $(k-1, l)$ and $(k, l-1)$. Take v_k and $\widehat{v}_l$ such that $x_{k-1} = x_k + \lambda_k(v_k - \varepsilon_k)$ and $\widehat{x}_{l-1} = \widehat{x}_l + \widehat{\lambda}_l(\widehat{v}_l - \widehat{\varepsilon}_l)$. We have $[x_k, v_k] \in A_k$ and $[\widehat{x}_l, \widehat{v}_l] \in \widehat{A}_l$. By Lemma 6.12 together with (6.7), we have that

$$(\lambda_k + \widehat{\lambda}_l)\|x_k - \widehat{x}_l\| \leq \lambda_k\|\widehat{x}_{l-1} - x_k\| + \widehat{\lambda}_l\|x_{k-1} - \widehat{x}_l\| + \lambda_k\widehat{\lambda}_l\Theta_{k,l} + \lambda_k\widehat{\lambda}_l\|\varepsilon_k - \widehat{\varepsilon}_l\|.$$

Therefore,

$$\begin{aligned}\|x_k - \widehat{x}_l\| \quad &\leq \quad \tfrac{\widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l}\|x_{k-1} - \widehat{x}_l\| + \tfrac{\lambda_k}{\lambda_k + \widehat{\lambda}_l}\|\widehat{x}_{l-1} - x_k\| \\ &\qquad + \tfrac{\lambda_k\widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l}\Theta_{k,l} + \tfrac{\lambda_k\widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l}\|\varepsilon_k - \widehat{\varepsilon}_l\| \\ &\leq \quad \tfrac{\widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l}\left[\|x_0 - u\| + \|\widehat{x}_0 - u\| + \alpha_{k-1,l} + \beta_{k-1,l} + \sqrt{\omega_{k-1,l}}\right] \\ &\qquad + \tfrac{\lambda_k}{\lambda_k + \widehat{\lambda}_l}\left[\|x_0 - u\| + \|\widehat{x}_0 - u\| + \alpha_{k,l-1} + \beta_{k,l-1} + \sqrt{\omega_{k,l-1}}\right] \\ &\qquad + \tfrac{\lambda_k\widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l}\Theta_{k,l} + \tfrac{\lambda_k\widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l}\|\varepsilon_k - \widehat{\varepsilon}_l\|. \end{aligned}$$

Using (6.14) and (6.15), we conclude that

$$\|x_k - \widehat{x}_l\| \leq \|x_0 - u\| + \|\widehat{x}_0 - u\| + \alpha_{k,l} + \beta_{k,l} + \left(\tfrac{\widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l} \sqrt{\omega_{k-1,l}} + \tfrac{\lambda_k}{\lambda_k + \widehat{\lambda}_l} \sqrt{\omega_{k,l-1}} \right). \quad (6.22)$$

We claim that $\omega_{k,l} \geq 0$ and

$$\tfrac{\widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l} \sqrt{\omega_{k-1,l}} + \tfrac{\lambda_k}{\lambda_k + \widehat{\lambda}_l} \sqrt{\omega_{k,l-1}} \leq \sqrt{\omega_{k,l}}. \quad (6.23)$$

Indeed, we have that

$$\left(\tfrac{\widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l} \sqrt{\omega_{k-1,l}} + \tfrac{\lambda_k}{\lambda_k + \widehat{\lambda}_l} \sqrt{\omega_{k,l-1}} \right)^2 \leq \tfrac{\widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l} \omega_{k-1,l} + \tfrac{\lambda_k}{\lambda_k + \widehat{\lambda}_l} \omega_{k,l-1}.$$

But direct computations show that

$$(\gamma_{k-1} - \widehat{\gamma}_l)^2 + \delta_{k-1} = (\gamma_k - \widehat{\gamma}_l)^2 + \delta_k - 2(\gamma_k - \widehat{\gamma}_l)\lambda_k |\!|\!| A_k u |\!|\!|,$$

and

$$(\gamma_k - \widehat{\gamma}_{l-1})^2 + \widehat{\delta}_{l-1} = (\gamma_k - \widehat{\gamma}_l)^2 + \widehat{\delta}_l + 2(\gamma_k - \widehat{\gamma}_l)\widehat{\lambda}_l |\!|\!| \widehat{A}_l u |\!|\!|.$$

It follows from this and (6.16) that

$$\begin{aligned} \tfrac{\widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l} \omega_{k-1,l} + \tfrac{\lambda_k}{\lambda_k + \widehat{\lambda}_l} \omega_{k,l-1} &= (\gamma_k - \widehat{\gamma}_l)^2 + \delta_k + \widehat{\delta}_l + \tfrac{\widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l} \eta_{k-1,l} + \tfrac{\lambda_k}{\lambda_k + \widehat{\lambda}_l} \eta_{k,l-1} \\ &\quad + \tfrac{\lambda_k \widehat{\lambda}_l}{\lambda_k + \widehat{\lambda}_l} 2(\gamma_k - \widehat{\gamma}_l) \left[|\!|\!| \widehat{A}_l u |\!|\!| - |\!|\!| A_k u |\!|\!| \right] \\ &= (\gamma_k - \widehat{\gamma}_l)^2 + \delta_k + \widehat{\delta}_l + \eta_{k,l} \\ &= \omega_{k,l}, \end{aligned}$$

which proves our claim and completes the proof of Theorem 6.2. ■

6.3.2 Discrete-continuous

In order to prove Theorem 6.6 we consider the points $\{\widehat{x}_l\}_{l=0}^m$ defined by (6.8b) and pass to the limit in inequality (6.12). Inequality (6.18) follows almost immediately, but the fact that β_k and $\eta_k(u)$ are finite is a more delicate issue.

First observe that using the recurrence formulae (6.15) and (6.16) repeatedly, along with the initial conditions, we can obtain closed expressions for $\beta_{k,l}$ and $\eta_{k,l}(u)$, respectively. Several summation and multiplication operations are involved in such expressions, making them unpractical for useful computations. Nevertheless, some estimates in terms of the local maxima and minima of the sequences $\{\lambda_k\}$ and $\{\widehat{\lambda}_l\}$ can be given. More precisely, define $\lambda_{n_+} = \max_{1 \leq k \leq n}\{\lambda_k\}$ and $\lambda_{n_-} = \min_{1 \leq k \leq n}\{\lambda_k\}$, and analogously for the sequence $\{\widehat{\lambda}_l\}$. Next, set $\mu_{k,l_+} = \frac{\widehat{\lambda}_{l_+}}{\widehat{\lambda}_{l_+} + \lambda_{k_-}}$, $\mu_{k,l_-} = \frac{\widehat{\lambda}_{l_-}}{\widehat{\lambda}_{l_-} + \lambda_{k_+}}$, $\nu_{k,l_+} = \frac{\lambda_{k_+}}{\widehat{\lambda}_{l_-} + \lambda_{k_+}}$ and $\nu_{k,l_-} = \frac{\lambda_{k_-}}{\widehat{\lambda}_{l_-} + \lambda_{k_+}}$. With this notation, we have

$$\beta_{n,m} \leq B_{n,m}, \qquad \text{and} \qquad \eta_{n,m}(u) \leq H_{n,m}(u), \quad (6.24)$$

where

$$B_{n,m} = \lambda_{n_+} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} \binom{i+j}{j} \mu_{n,m_+}^{i+1} \nu_{n,m_+}^{j} \Theta_{n-i,m-j} \quad (6.25)$$

and

$$\begin{aligned}
H_{n,m}(u) &= 2\lambda_{n_+}\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}\binom{i+j}{j}\mu_{n,m_+}^{i+1}\nu_{n,m_+}^{j}\gamma_{n-i}(u)\|A_{n-i}u\| \\
&- 2\lambda_{n_-}\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}\binom{i+j}{j}\mu_{n,m_-}^{i+1}\nu_{n,m_-}^{j}\gamma_{n-i}(u)\|\widehat{A}_{m-j}u\| \\
&- 2\lambda_{n_-}\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}\binom{i+j}{j}\mu_{n,m_-}^{i+1}\nu_{n,m_-}^{j}\widehat{\gamma}_{m-j}(u)\|A_{n-i}u\| \\
&+ 2\lambda_{n_+}\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}\binom{i+j}{j}\mu_{n,m_+}^{i+1}\nu_{n,m_+}^{j}\widehat{\gamma}_{m-j}(u)\|\widehat{A}_{m-j}u\|. \qquad (6.26)
\end{aligned}$$

Remark 6.13 *If $\widehat{\lambda}_l \equiv \frac{t}{m}$ for some $t > 0$ and $m \geq 1$, then the expression $\mu_{n,m_\pm}^{i+1}$ in (6.25) and (6.26) can be replaced by $\prod_{k=n-i}^{n}\mu_k$, where $\mu_k = \frac{\frac{t}{m}}{\lambda_k+\frac{t}{m}}$. Also, $\nu_{n,m_+} = \frac{\lambda_{n_+}}{\lambda_{n_+}+\frac{t}{m}}$ and similarly for ν_{n,m_-}.*

We shall prove that $B_{n,m}$ and $H_{n,m}(u)$ converge as $m \to \infty$, from which we deduce that β_n and $\eta_n(u)$ are finite. In order to simplify notation let $\Lambda_{n,i} = \prod_{p=n-i}^{n}\lambda_p$.

Lemma 6.14 *Let $\tau = t - t_0$. With the notation introduced above, the following holds for each $n \geq 1$:*

$$\lim_{m\to\infty} B_{n,m} = \lambda_{n_+}\int_0^{\tau}\sum_{i=0}^{n-1}\frac{1}{i!}\Theta(s_0+\sigma_n-\sigma_i, t-\xi)\left(\frac{\xi^i e^{-\frac{\xi}{\lambda_{n_+}}}}{\Lambda_{n,i}}\right)d\xi$$

and

$$\begin{aligned}
\lim_{m\to\infty} H_{n,m}(u) &= \lambda_{n_+}\int_0^{\tau}\sum_{i=0}^{n-1}\frac{2}{i!}\gamma_{n-i}\|A_{n-i}u\|\left(\frac{\xi^i e^{-\frac{\xi}{\lambda_{n_+}}}}{\Lambda_{n,i}}\right)d\xi \\
&- \lambda_{n_-}\int_0^{\tau}\sum_{i=0}^{n-1}\frac{2}{i!}\gamma_{n-i}\|A(t-\xi)u\|\left(\frac{\xi^i e^{-\frac{\xi}{\lambda_{n_-}}}}{\Lambda_{n,i}}\right)d\xi \\
&- \lambda_{n_-}\int_0^{\tau}\sum_{i=0}^{n-1}\frac{2}{i!}\mathcal{A}_u(t-\xi)\ \|A_{n-i}u\|\left(\frac{\xi^i e^{-\frac{\xi}{\lambda_{n_-}}}}{\Lambda_{n,i}}\right)d\xi \\
&+ \lambda_{n_+}\int_0^{\tau}\sum_{i=0}^{n-1}\frac{2}{i!}\mathcal{A}_u(t-\xi)\ \|A(t-\xi)u\|\left(\frac{\xi^i e^{-\frac{\xi}{\lambda_{n_+}}}}{\Lambda_{n,i}}\right)d\xi.
\end{aligned}$$

Proof. The idea is to express $B_{n,m}$ and $H_{n,m}$ as Riemann sums of certain step functions and apply the Bounded Convergence Theorem. First observe that, according to Remark 6.13, we have

$$\sum_{j=0}^{m-1}\binom{i+j}{j}\mu_{n,m_\pm}^{i+1}\nu_{n,m_\pm}^{j}$$

$$
\begin{aligned}
&= \sum_{j=0}^{m-1} \frac{(j+1)(j+2)\cdots(j+i)}{i!} \prod_{p=n-i}^{n} \left(\frac{\frac{\tau}{m}}{\lambda_p + \frac{\tau}{m}} \right) \left(\frac{\lambda_{n\pm}}{\lambda_{n\pm} + \frac{\tau}{m}} \right)^j \\
&= \frac{1}{i!} \sum_{j=0}^{m-1} \left[\prod_{p=n-i}^{n} \left(\frac{1}{\lambda_p + \frac{\tau}{m}} \right) \right] \left[\prod_{q=1}^{i} \left(\frac{j\tau}{m} + \frac{q\tau}{m} \right) \right] \left[\left(1 + \frac{\tau}{m\lambda_{n\pm}} \right)^{\frac{m}{\tau}} \right]^{-\frac{j\tau}{m}} \frac{\tau}{m} \\
&= \int_0^{\tau} \psi_m(\xi)\, d\xi,
\end{aligned}
$$

where ψ_m is a step function defined as follows: if $\frac{j\tau}{m} \leq \xi < \frac{(j+1)\tau}{m}$ then

$$
\begin{aligned}
\psi_m(\xi) &= \frac{1}{i!} \left[\prod_{p=n-i}^{n} \left(\frac{1}{\lambda_p + \frac{\tau}{m}} \right) \right] \left[\prod_{q=1}^{i} \left(\frac{j\tau}{m} + \frac{q\tau}{m} \right) \right] \left[\left(1 + \frac{\tau}{m\lambda_{n\pm}} \right)^{\frac{m}{\tau}} \right]^{-\frac{j\tau}{m}} \qquad (6.27) \\
&= \frac{m}{\tau} \binom{i+j}{j} \mu_{n,m\pm}^{i+1} \nu_{n,m\pm}^{j}.
\end{aligned}
$$

Notice that for $\frac{j\tau}{m} \leq \xi < \frac{(j+1)\tau}{m}$ we have

$$
\begin{aligned}
\psi_m(\xi) &\leq \frac{m}{\tau} \binom{i+j}{j} \mu_{n,m\pm}^{i+1} \\
&= \frac{m}{\tau} \frac{(j+i)!}{i!\, j!} \prod_{p=n-i}^{n} \left(\frac{\frac{\tau}{m}}{\lambda_p + \frac{\tau}{m}} \right) \\
&\leq \frac{m}{\tau} \frac{\tau^{i+1}}{m^{i+1}} \frac{(j+1)(j+2)\cdots(j+i)}{i!} \prod_{p=n-i}^{n} \left(\frac{1}{\lambda_p} \right) \\
&\leq \frac{\tau^i\, (2m)^i}{i!\, m^i\, \Lambda_{n,i}} \\
&= \frac{(2\tau)^i}{i!\, \Lambda_{n,i}}
\end{aligned}
$$

and so, the sequence $\{\psi_m\}$ is uniformly bounded. Moreover,

$$
\lim_{m\to\infty} \psi_m(\xi) = \frac{\xi^i e^{-\frac{\xi}{\lambda_{n\pm}}}}{i!\, \Lambda_{n,i}}
$$

on $[0,\tau]$. To see this we use representation (6.27). The only difficult part is the middle bracket. But $\prod_{q=1}^{i} \left(\frac{j\tau}{m} + \frac{q\tau}{m} \right)$ is a polynomial in $\frac{j\tau}{m}$ of degree i. The leading coefficient is 1, while the rest are bounded by a constant (depending only on i and τ) times $\frac{1}{m}$ and so they vanish as $m \to \infty$.

We have proved that the sequence $\{\psi_m\}$ is uniformly bounded and pointwise convergent on $[0,\tau]$. But the same is true for the sequences

$$
\Theta_{n-i,m-j}, \qquad \|\widehat{A}_{m-j} u\| \qquad \text{and} \qquad \widehat{\gamma}_{m-j}(u),
$$

which converge, almost everywhere, to

$$
\Theta(s_0 + \sigma_n - \sigma_i, t-\xi), \qquad \|\widehat{A}(t-\xi)u\| \qquad \text{and} \qquad \widehat{\mathcal{A}}_u(t-\xi),
$$

respectively[4] due to the hypothesized Riemann-integrability. The result then follows from the Bounded Convergence Theorem. ∎

[4]Again, j is related to ξ via $\frac{j\tau}{m} \leq \xi < \frac{(j+1)\tau}{m}$.

Proof of Theorem 6.6. Let $\{\widehat{x}_l\}_{l=0}^m$ be the points defined by (6.8b). By virtue of Theorem 6.2, we have

$$\left\| x_k - \prod_{l=1}^m \left(I - \frac{\tau}{m}\widehat{A}_l\right)^{-1} u \right\| \leq \|x_0 - u\| + \alpha_{k,m} + \beta_{k,m}$$
$$+\sqrt{(\gamma_k(u) - \widehat{\gamma}_m(u))^2 + \delta_k(u) + \widehat{\delta}_m(u) + \eta_{k,m}(u)}.$$

Since the subjacent proximal scheme is exact, we have $\alpha_{k,m} = e_k$ for all m. It is easy to see that $\lim_{m\to\infty} \widehat{\delta}_m(u) = 0$, while $\lim_{m\to\infty} \widehat{\gamma}_m(u) = \mathcal{A}_u(t, t_0)$. Letting $m \to \infty$ in the previous inequality, we obtain (6.17). Finally, β_n and $\eta_n(u)$ are finite by virtue of (6.24) and Lemma 6.14. ■

6.3.3 Continuous-continuous

The idea is to use Theorem 6.6 and pass to the limit once more. To do this, we shall compute $\lim_{n\to\infty} \beta_n$ and $\lim_{n\to\infty} \eta_n(u)$.

It is not difficult to verify that $\beta_n = \lim_{m\to\infty} B_{n,m}$ while $\eta_n(u) = \lim_{m\to\infty} H_{n.m}(u)$ because $\lambda_k \equiv \frac{\sigma}{n}$. On the other hand, for $\xi \in [0,\tau]$, $\zeta \in [0,\sigma]$ and $0 \leq i \leq n-1$ define

$$f_n(\xi, \zeta) = \frac{2}{i!\,\xi}\left(\frac{n\xi}{\sigma}\right)^{i+1} e^{-\frac{n\xi}{\sigma}} \qquad \text{with} \quad \frac{i\sigma}{n} \leq \zeta < \frac{(i+1)\sigma}{n}.$$

The expressions in Lemma 6.14 become

$$\beta_n = \int_0^\tau \left\{ \sum_{i=0}^{n-1} \frac{1}{2}\Theta(s_0 + \sigma_n - \sigma_i, t-\xi) f_n\left(\xi, \tfrac{i\sigma}{n}\right)\frac{\sigma}{n} \right\} d\xi \tag{6.28}$$

and

$$\begin{aligned}
\eta_n \;=\; & \int_0^\tau \left\{ \sum_{i=0}^{n-1} \gamma_{n-i} \|A_{n-i}u\|\; f_n\left(\xi, \tfrac{i\sigma}{n}\right)\frac{\sigma}{n} \right\} d\xi \\
& - \int_0^\tau \left\{ \sum_{i=0}^{n-1} \gamma_{n-i}\; \|A(t-\xi)u\| f_n\left(\xi, \tfrac{i\sigma}{n}\right)\frac{\sigma}{n} \right\} d\xi \\
& - \int_0^\tau \left\{ \sum_{i=0}^{n-1} \mathcal{A}_u(t-\xi, t_0) \|A_{n-i}u\|\; f_n\left(\xi, \tfrac{i\sigma}{n}\right)\frac{\sigma}{n} \right\} d\xi \\
& + \int_0^\tau \left\{ \sum_{i=0}^{n-1} \mathcal{A}_u(t-\xi, t_0)\; \|A(t-\xi)u\| f_n\left(\xi, \tfrac{i\sigma}{n}\right)\frac{\sigma}{n} \right\} d\xi,
\end{aligned} \tag{6.29}$$

respectively.

Lemma 6.15 *Fix* $\xi \in [0,\tau]$. *The sequence* $\{f_n(\xi, \cdot\,)\}$ *converges uniformly to zero on every closed subset of* $[0,\sigma]$ *not containing* ξ.

Proof. Take $\zeta \in [0,\sigma]$ and define $i_n = \left\lfloor \frac{n\zeta}{\sigma} \right\rfloor$ so that

$$\frac{n\zeta}{\sigma} - 1 < i_n \leq \frac{n\zeta}{\sigma} \tag{6.30}$$

and

$$f_n(\xi,\zeta) = \frac{1}{i_n!\;\xi}\left(\frac{n\xi}{\sigma}\right)^{i_n+1} e^{-\frac{n\xi}{\sigma}}$$

for each n. Stirling's Formula states that

$$\lim_{m\to\infty} \frac{\sqrt{2\pi}\; m^{m+1/2}}{e^m\; m!} = 1.$$

The sequence being convergent, there exists a constant $M > 0$ such that

$$f_n(\xi,\zeta) \le \frac{M}{\xi\;(i_n)^{i_n+1/2}}\left(\frac{n\xi}{\sigma}\right)^{i_n+1} e^{i_n-\frac{n\xi}{\sigma}}$$

for all n.

Denoting $\frac{\xi}{\sigma}$ by a and $\frac{\zeta}{\sigma}$ by b, by virtue of the double inequality (6.30), we have[5]

$$\begin{aligned}
f_n(\xi,\zeta) &\le \frac{M(an)^{bn+1}e^{bn-an}}{\xi\;(bn-1)^{bn-1/2}} \\
&= \frac{M}{\xi}\;an\sqrt{bn}\left(\frac{an}{bn-1}\right)^{bn} e^{n(b-a)} \\
&= \frac{M}{\sigma}\left(\frac{bn}{bn-1}\right)^{bn}\sqrt{b}n^{3/2}\left(\frac{a}{b}\right)^{bn} e^{n(b-a)} \\
&= \frac{M\sqrt{b}}{\sigma}\left(\frac{bn}{bn-1}\right)^{bn} n^{3/2}\left[\left(\frac{a}{b}\right)^{b} e^{(b-a)}\right]^n \\
&\le Cn^{3/2}\left[e^{\left(b-a+b\ln\left(\frac{a}{b}\right)\right)}\right]^n,
\end{aligned}$$

where $C = \frac{4M\sqrt{b}}{\sigma}$. Now, if ζ is in a closed set not containing ξ then $|b-a| \ge c$ for some $c > 0$. By continuity, $b - a + b\ln\left(\frac{a}{b}\right) \le -d$ for some $d > 0$ and the result follows. ■

Remark 6.16 *Each of the sums in braces on the right-hand sides of equations (6.28) and (6.29) can be interpreted as integrals of the form*

$$\int_0^\sigma \phi_n(\xi,\zeta) f_n(\xi,\zeta)\; d\zeta,$$

where the sequences $\{\phi_n\}$ and $\{f_n\}$ have the following properties:

i) For each $\xi \in [0,\tau]$, and each n, $\phi_n(\xi,\cdot\,)$ is a step function and the sequence $\{\phi_n(\xi,\cdot\,)\}$ converges uniformly to a continuous function $\phi(\xi,\cdot)$ as $n \to \infty$.

ii) For each $\xi \in [0,\tau]$, and each n, $f_n(\xi,\cdot\,)$ is a step function and the sequence $\{f_n(\xi,\cdot\,)\}$ converges uniformly (and thus in L^1) to zero on every closed subset of $[0,\sigma]$ not containing ξ.

iii) For $\xi \in [0,\tau]$ set

$$I_n(\xi) := \int_0^\sigma f_n(\xi,\zeta)\; d\zeta = \sum_{i=0}^{n-1} \frac{2}{i!}\left(\frac{n\xi}{\sigma}\right)^i e^{-\frac{n\xi}{\sigma}}.$$

[5] Since the convergence around zero is straightforward, we may assume $b \ge 2/n$ and so $bn - 1 \ge 1$

Then[6]

$$\lim_{n\to\infty} I_n(\xi) = \begin{cases} 2 & \textit{if} \quad \xi < \sigma \\ 1 & \textit{if} \quad \xi = \sigma \\ 0 & \textit{if} \quad \xi > \sigma. \end{cases}$$

iv) The integrals are bounded functions of ξ*.*

The following is an immediate consequence of the Bounded Convergence Theorem.

Lemma 6.17 *Let* $\{\phi_n\}$ *and* $\{f_n\}$ *satisfy i), ii), iii) and iv). Then*

$$\lim_{n\to\infty} \int_0^\tau \int_0^\sigma \phi_n(\xi,\zeta) f_n(\xi,\zeta)\ d\zeta\ d\xi = 2\int_0^\delta \phi(\xi,\xi)\ d\xi,$$

where $\delta = \min\{\tau,\sigma\}$*.*

Proof of Theorem 6.8. According to Remark 6.16 and Lemma 6.17, we have

$$\begin{aligned}
\lim_{n\to\infty} \eta_n &= 2\int_0^\tau \mathcal{A}_u(s-\xi,s_0)\|\!|A(s-\xi)u\|\!|\ d\xi - 2\int_0^\tau \mathcal{A}_u(s-\xi,s_0)\|\!|\widehat{A}(t-\xi)u\|\!|\ d\xi \\
&\quad - 2\int_0^\tau \widehat{\mathcal{A}}_u(t-\xi,t_0)\|\!|A(s-\xi)u\|\!|\ d\xi + 2\int_0^\tau \widehat{\mathcal{A}}_u(t-\xi,t_0)\|\!|\widehat{A}(t-\xi)u\|\!|\ d\xi \\
&= 2\int_0^\tau \Big(\mathcal{A}_u(s-\xi,s_0) - \widehat{\mathcal{A}}_u(t-\xi,t_0)\Big)\Big(\|\!|A(s-\xi)u\|\!| - \|\!|\widehat{A}(t-\xi)u\|\!|\Big)\ d\xi \\
&= \Big[\mathcal{A}_u(s,s_0) - \widehat{\mathcal{A}}_u(t,t_0)\Big]^2 - [\mathcal{A}_u(t_0+s-t,s_0)]^2
\end{aligned}$$

while

$$\lim_{n\to\infty} \beta_n = \int_0^\tau \Theta(s-\xi,t-\xi)\ d\xi.$$

The result follows from Theorem 6.6. ■

[6] To see this, consider the normalized partial averages of a sequence of independent Poisson-distributed random variables with parameter $a = \xi/\sigma$. According to the Central Limit Theorem the distribution functions F_n converge uniformly to the normal distribution function. Finally, the sum above corresponds to $F_n(\xi_n)$ where $\xi_n = \sqrt{\frac{n}{a}}(1-a)$

Chapter 7

Asymptotic almost-equivalence of abstract evolution systems

7.1 Introduction and preliminaries

Let C be a nonempty Borel subset of a Banach space $(X, \|\cdot\|)$. An *evolution system* (ES for short) on C is a two-parameter family $U = \{U(t,s) \mid t \geq s \geq 0\}$ of possibly non-linear maps from C into itself satisfying:

i) $\forall t \geq 0$, $\forall x \in C$, $U(t,t)x = x$; and

ii) $\forall t \geq s \geq r \geq 0$, $\forall x \in C$, $U(t,s)U(s,r)x = U(t,r)x$.

We call U an *M-Lipschitz evolution system* (*M-LES*) if, for some $M > 0$, we have in addition:

iii) $\forall t \geq s \geq 0$, $\forall x, y \in C$, $\|U(t,s)x - U(t,s)y\| \leq M\|x - y\|$.

A *contracting evolution system* (*CES*) is a 1-LES.

We say that an evolution system U is *autonomous* if for all $t, s \geq 0$ we have $U(t,0) = U(t+s,s)$. For such an ES, the family $T = \{T(t) := U(t,0) \mid t \geq 0\}$ defines a *semigroup*, that is

a) $\forall x \in C$, $T(0)x = x$; and

b) $\forall t, s \geq 0$, $\forall x \in C$, $T(t)T(s)x = T(t+s)x$.

In this case, we have the relation $U(t,s) = T(t-s)$.

Example 15 *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*

$$\begin{cases} u'(t) \in F(t, u(t)) & \text{for almost every } t \in (0,\infty) \\ u(s) = x \end{cases}$$

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 16 *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*

$$V(t,s) = \prod_{n=\nu(s)+1}^{\nu(t)} F_n,$$

where the product represents the composition of functions. Then V is an ES. If each F_n is M_n-Lipschitz and the infinite product $\prod_{n=1}^{\infty} M_n$ is bounded from above by M, then V is an M-LES. *This holds, for instance, if each F_n is of the form $F_n = (I + A_n)^{-1}$, where $\{A_n\}$ is a family of m-accretive operators on C. In this case, $M_n = 1$ for all n so the piecewise constant interpolation of infinite product of resolvents is a* CES.

If U is an ES on C, an *orbit* of U is a function $u : [0, \infty) \to C$ such that for some $t_0 \geq 0$ and $x_0 \in C$, $u(t) = U(t, t_0)x_0$ for all $t \geq t_0$. Throughout this paper, all orbits are assumed to be measurable and locally bounded, hence locally integrable on $[0, \infty)$. More generally, we say that a function $u \in L^\infty_{\text{loc}}(0, \infty; C)$ is an *almost-orbit* of U if

$$\lim_{t\to\infty} \sup_{h\geq 0} \|u(t + h) - U(t + h, t)u(t)\| = 0. \tag{7.1}$$

Remark 7.1 *The term "almost-orbit" was introduced in [84] for a* continuous *function satisfying* (7.1)*. Later, in [83], the author gives a different definition, just requiring*

$$\lim_{t,h\to\infty} \|u(t + h) - U(t + h, t)u(t)\| = 0,$$

but still for continuous functions. The latter condition is strictly but slightly weaker than (7.1) *for practical purposes. In fact, the example provided in [83] to motivate the interest of studying almost-orbits also satisfies* (7.1)*. In both cited works the authors give some criteria that can be applied to an almost-orbit in order to guarantee certain asymptotic behavior. The same approach is used in [106]. In [99, 98], the authors carry out a similar analysis for* uniformly asymptotically almost nonexpansive curves *(a concept that includes almost-orbits* of almost nonexpansive semigroups*) in Hilbert space. Other results on the asymptotic behavior of almost-orbits of nonexpansive semigroups can be found in [72] (see also the references therein). It is important to mention that [84, 83] contain versions of Proposition 7.13 below in the corresponding setting. Our intention is to show how to derive many asymptotic properties of almost-orbits by studying only the orbits. In that sense our work is different but complementary to [84, 83].*

Clearly the orbits of U are almost-orbits as well. We say that two ES, namely U and V, are *asymptotically almost-equivalent* if every orbit of U is an almost-orbit of V and viceversa.

Remark 7.2 *Let U be an autonomous* CES *and suppose V is asymptotically almost-equivalent to U. Then V is an* asymptotic semigroup *in the sense introduced in [90], where the author proves that in this setting every orbit of U converges strongly (or weakly) if, and only if, every orbit of V does (see Proposition 7.9 below).*

Remark 7.3 *Suppose U is a* M-LES *on C. If u is an almost-orbit of U and some function $v \in L^\infty_{loc}(0, \infty; C)$ satisfies $\lim_{t\to\infty} \|v(t) - u(t)\| = 0$ then v is also an almost-orbit of U.*

Remark 7.4 *Suppose that for each $r > 0$ there is $G_r : \mathbf{R}^2_+ \to \mathbf{R}_+$ such that for all $x \in B(0, r)$ and $t \geq s \geq 0$ we have $\|U(t, s)x - V(t, s)x\| \leq G_r(t, s)$ and $\lim_{t\to\infty} G_r(t + h, t) = 0$ uniformly in $h \geq 0$. Then every bounded orbit of U is an almost-orbit of V and viceversa.*[1]

Example 17 *Let A be a m-accretive operator on X and let U be the autonomous* CES *defined by the differential inclusion $-\dot{u} \in Au$ as in Example 15.*

1. *Take $f \in L^1(0, \infty; X)$ and let V be the evolution system defined by the integral solutions of $-\dot{u} \in Au + f$. The orbits of V are almost-orbits of U (see [84]). This result is generalized in CITE{APK}.*

[1] If the same function G works for all $r > 0$, then the boundedness assumption is unnecessary.

2. *For a sequence $\{\lambda_n\}$ of positive numbers define $W(t,s) = \prod_{n=\nu(s)+1}^{\nu(t)} (I + \lambda_n A)^{-1}$ as in Example 16. If $\{\lambda_n\} \in \ell^2 \setminus \ell^1$ then U and W are asymptotically almost-equivalent (see [70], although the fact that the orbits of W are almost-orbits of U had already been proved in [84]). The same equivalence appeared earlier in [90] assuming A is single-valued and Lipschitz.*

3. *In the special case where A is the subdifferential of a proper, lower-semicontinuous convex function in Hilbert space, U and W are asymptotically almost-equivalent if only $\{\lambda_n\} \notin \ell^1$ (see [56]).*

It is important to underscore the fact that U, V and W are CES.

The aim of this paper is to show that under suitable conditions, orbits and almost-orbits of a given ES have the same asymptotic behavior in terms of boundedness, convergence, convergence in average and other related properties. As a consequence, the same holds true for all orbits of asymptotically almost-equivalent evolution systems. A few results of this kind can be found in [90, 84, 70, 56] in the context of autonomous differential inclusions of the type $-u'(t) \in Au(t)$, where A is a m-accretive operator. A first attempt to deal with certain nonautonomous and non-Lipschitz systems can be found in [83].

7.2 Lipschitz evolution systems

7.2.1 On the boundedness of almost-orbits

First we prove that for a *M-LES* either all almost-orbits are bounded or none is so. We need the following extension of [84, Lemma 3.1]:

Lemma 7.5 *Let U be a* M-LES *and u_1, u_2 two almost-orbits of U. Then*

$$\limsup_{t\to\infty} \|u_1(t) - u_2(t)\| \leq M \liminf_{t\to\infty} \|u_1(t) - u_2(t)\| < \infty.$$

Proof. For $i = 1,2$ let $\psi_i(t) = \sup_{h\geq 0} \|u_i(t+h) - U(t+h,t)u_i(t)\|$. For every $h \geq 0$ we have

$$\begin{aligned} \|u_1(t+h) - u_2(t+h)\| &\leq \psi_1(t) + \psi_2(t) + \|U(t+h,t)u_1(t) - U(t+h,t)u_2(t)\| \\ &\leq \psi_1(t) + \psi_2(t) + M\|u_1(t) - u_2(t)\|. \end{aligned}$$

Letting $h \to \infty$ yields

$$\limsup_{h\to\infty} \|u_1(h) - u_2(h)\| \leq \psi_1(t) + \psi_2(t) + M\|u_1(t) - u_2(t)\| < \infty$$

and as $t \to \infty$ we get

$$\limsup_{h\to\infty} \|u_1(h) - u_2(h)\| \leq M \liminf_{t\to\infty} \|u_1(t) - u_2(t)\|.$$

■

Corollary 7.6 *Let U be an* M-LES *and u_1, u_2 two almost-orbits of U. We have*

i) If one almost-orbit of U is bounded, then every almost-orbit of U is bounded as well.

ii) If 0 is a cluster point of $\|u_1(t) - u_2(t)\|$ then $\lim_{t\to\infty} \|u_1(t) - u_2(t)\| = 0$.

iii) If U is a CES*, the limit* $\lim_{t\to\infty} \|u_1(t) - u_2(t)\|$ *always exists.*

Remark 7.7 *The previous corollary implies that if U and V are asymptotically almost-equivalent* M-LES*, and if* **one** *almost-orbit of U* **or** *V is bounded, then* **every** *almost-orbit of U* **and** *V is bounded.*

Remark 7.8 *Let u be a bounded almost-orbit of an* M-LES *U so that $\|u\|_\infty = \sup_t \|u(t)\| < \infty$. Since u is an almost-orbit, there exists $p_0 \geq 0$ such that for all $p \geq p_0$, we have*

$$\|u(p+h) - U(p+h,p)u(p)\| \leq 1 \textit{ for all } h \geq 0.$$

Hence, for all $p \geq p_0$ and $h \geq 0$ we get $\|U(p+h,p)u(p)\| \leq 1 + \|u\|_\infty$.

7.2.2 Strong and weak convergence

The following result and its proof are inspired by [90, Lemma 1], where the author studies two special cases: when U is an autonomous *CES*, and when the almost-orbits are in fact the orbits of a semigroup of contractions. Here we give a shorter proof in a more general context.

Proposition 7.9 *Let U be an* M-LES*. If every orbit of U converges strongly (resp. weakly) as time goes to infinity, then the same holds true for every almost-orbit of U.*

Proof. Let τ denote the hypothesized topology. And suppose that the $\tau-$limit of $U(t,s)x$ as $t \to \infty$ exists for all x and s. Let u be an almost-orbit of U. Take $p \geq 0$ and set $\zeta(p) = \tau - \lim_{t\to\infty} U(t,p)u(p)$. We have

$$\zeta(p+h) - \zeta(p) = \tau - \lim_{t\to\infty} \{U(t,p+h)u(p+h) - U(t,p)u(p)\}.$$

But for all $t \geq p+h$ we have

$$\begin{aligned} \|U(t,p+h)u(p+h) - U(t,p)u(p)\| &= \|U(t,p+h)u(p+h) - U(t,p+h)U(p+h,p)u(p)\| \\ &\leq M\|u(p+h) - U(p+h,p)u(p)\| \end{aligned}$$

and by τ-lower semicontinuity of the norm we get

$$\|\zeta(p+h) - \zeta(p)\| \leq M\|u(p+h) - U(p+h,p)u(p)\|.$$

Since u is an almost-orbit of U, the right-hand side tends to zero as $p \to \infty$ uniformly in $h \geq 0$. Therefore $\{\zeta(p) : p \to \infty\}$ is a Cauchy net that converges strongly to a limit ζ_∞. Finally, we have for all $p, h \geq 0$

$$u(p+h) - \zeta_\infty = [u(p+h) - U(p+h,p)u(p)] + [U(p+h,p)u(p) - \zeta(p)] + [\zeta(p) - \zeta_\infty].$$

Given $\varepsilon > 0$ we can choose p large enough so that the first and third terms on the right-hand side are less than ε in norm, uniformly in h for the first term. Next for such a fixed p, we let $h \to \infty$ so that the second term $\tau-$converges to zero. Hence $u(t)$ is $\tau-$convergent to ζ_∞ as $t \to \infty$. ∎

Remark 7.10 *Consider the follwing, more general setting: Let (X,d) be a complete metric space (not even the linear structure is necessary). The Lipschitz condition in the definition of* M-LES *reads $d(U(t,s)x, U(t,s)y) \leq Md(x,y)$. The definition of almost-orbit can be rephrased as $\lim_{t\to\infty} \sup_{h\geq 0} d(u(t+h), U(t+h,t)u(t)) = 0$. It is easy to see that Lemma 7.5, Corollary 7.6 and the statement in Proposition 7.9 concerning the strong topology are still true.*

Example 18 *In [87] the authors proved the strong convergence of the orbits of semigroups under certain assumptions on the semigroup and the underlying space. More than two decades later the result was extended in [106] for the almost orbits in the same setting. This extension is straightforward using Proposition 7.9. The result was also extended in [53] to more general spaces.*

7.2.3 Convergence in average

Given a function $v \in L^\infty_{\text{loc}}(0,\infty;X)$, define

$$\overline{v}(t) = \frac{1}{t}\int_0^t v(\xi)\ d\xi.$$

We say that v converges strongly (resp. weakly) *in average* if $\overline{v}(t)$ has a strong (resp. weak) limit as $t \to \infty$. Under fairly general conditions, convergence in average is also inherited by almost-orbits.

Remark 7.11 *Given $v \in L^\infty_{loc}(0,\infty;X)$ and $h \geq 0$, we define $v_h(t) = v(h+t)$ for all $t \geq 0$. Observe that*

$$\overline{v_h}(t) = \frac{1}{t}\int_0^t v(h+\xi)\ d\xi = \left(\frac{t+h}{t}\right)\frac{1}{t+h}\int_0^{t+h} v(\eta)\ d\eta - \frac{1}{t}\int_0^h v(\xi)\ d\xi.$$

Therefore, if v converges strongly (resp. weakly) in average to L, the same holds for v_h, for each $h \geq 0$.

Proposition 7.12 *Let U be an* M-LES. *If every orbit of U converges strongly (resp. weakly) in average, so does every almost-orbit.*

Proof. Let u be an almost-orbit of U. For $p, h \geq 0$ and t sufficiently large, define

$$\sigma_h(t,p) = \frac{1}{t}\int_0^t U(p+h+\xi,p)u(p)\ d\xi \tag{7.2}$$

and set $\zeta(p) = \tau - \lim_{t\to\infty} \sigma_0(t,p)$, where τ stands for either the strong or the weak topology according to the hypothesis. Notice that

$$[\sigma_0(t,p+h) - \sigma_0(t+h,p)] - [\sigma_h(t,p) - \sigma_0(t+h,p)] = [\sigma_0(t,p+h) - \sigma_h(t,p)]\,. \tag{7.3}$$

By virtue of Remark 7.11,

$$\tau - \lim_{t\to\infty} \sigma_h(t,p) = \tau - \lim_{t\to\infty} \sigma_0(t,p) = \tau - \lim_{t\to\infty} \sigma_0(t+h,p)$$

for each $h \geq 0$. We let $t \to \infty$ in equation (7.3) and use the weak lower-semicontinuity of the norm to obtain

$$\|\zeta(p+h) - \zeta(p)\| \leq \|\sigma_0(t,p+h) - \sigma_h(t,p)\| \leq M\ \|u(p+h) - U(p+h,p)u(p)\|,$$

which in turn tends to zero as $p \to \infty$ uniformly in $h \geq 0$. As a consequence, $\zeta(p)$ converges strongly to some ζ_∞ as $p \to \infty$.

Finally, for any $p, h \geq 0$ we write

$$\begin{aligned}\overline{u}(p+h) - \zeta_\infty \quad &= \quad \frac{1}{p+h}\int_0^p u(\xi)\ d\xi + \frac{1}{p+h}\int_0^h [u(p+\xi) - U(p+\xi,p)u(p)]\ d\xi \\ &\quad + \left[\frac{h}{p+h}\sigma(h,p) - \zeta(p)\right] + [\zeta(p) - \zeta_\infty].\end{aligned}$$

The second term on the right-hand side is bounded by $\sup_{k\geq 0}\|u(p+k)-U(p+k,p)u(p)\|$, which is independent of h and tends to zero as $p\to\infty$. The last term converges strongly to zero as $p\to\infty$. Thus, given any $\varepsilon>0$, we can choose p_ε large enough so that the second and forth terms are both less than ε. Having fixed p_ε, the first term converges strongly to zero as $h\to\infty$ while the third term τ-converges to zero. As a consequence $\overline{u}(t)$ is τ-convergent to ζ_∞ as $t\to\infty$. ■

7.2.4 Almost-convergence

We say that a function $v\in L^\infty_{\text{loc}}(0,\infty;X)$ is strongly (resp. weakly) *almost-convergent* (in the sense of Lorentz, [79]) if there exists $y\in X$ such that $\overline{v_h}(t)$ converges strongly (resp. weakly) to y as $t\to\infty$ uniformly in $h\geq 0$. Clearly almost-convergence implies convergence in average. Conversely, according to Remark 7.11, if v converges in average then $\overline{v_h}$ converges for each $h\geq 0$, so the uniformity in $h\geq 0$ is what makes the difference.

This type of convergence is interesting because a trajectory $v(t)$ is convergent if and only if it is almost-convergent and *asymptotically regular* (the difference $v(t+h)-v(t)$ converges to zero as $t\to\infty$ for each $h\geq 0$) for the corresponding topology (see [79]). This fact − or method of proof − has been used, for instance in [20]. The following result shows that almost-convergence of the orbits is also inherited by the almost-orbits.

Proposition 7.13 *Let U be an* M-LES. *If every bounded orbit of U is strongly (resp. weakly) almost-convergent, so is every bounded almost-orbit of U.*

Proof. Let u be a bounded almost-orbit of U and define

$$\sigma_h(t,p)=\frac{1}{t}\int_0^t U(p+h+\xi,p)u(p)\;d\xi.$$

According to Remark 7.8, there exists $p_0\geq 0$ such that for all $p\geq p_0$ and $h\geq 0$ we have $\|U(p+h,p)u(p)\|\leq 1+\|u\|_\infty$. Therefore, by virtue of the hypothesis, for every $p\geq p_0$ there exists $\zeta(p)\in X$ such that for all $h\geq 0$, $\zeta(p)=\tau-\lim_{t\to\infty}\sigma_h(t,p)$, and the convergence is uniform in $h\geq 0$.

Next, we prove that $\{\zeta(p):p\geq 0\}$ is a Cauchy net. For every $p,h\geq 0$ and $t\geq p+h$ We have

$$\begin{aligned}\|\sigma_0(t,p+h)-\sigma_h(t,p)\| &\leq \frac{1}{t}\int_0^t \|U(p+h+\xi,p+h)u(p+h)-U(p+h+\xi,p)u(p)\|\;d\xi\\ &\leq M\|u(p+h)-U(p+h,p)u(p)\|.\end{aligned}$$

Let $t\to\infty$ to get

$$\|\zeta(p+h)-\zeta(p)\|\leq M\|u(p+h)-U(p+h,p)u(p)\|,$$

which tends to 0 as $p\to\infty$ uniformly in $h\geq 0$. Hence $\zeta(p)\to\zeta_\infty$ as $p\to\infty$, for some ζ_∞.

For any $p,h,k\geq 0$ we write

$$\begin{aligned}\overline{u_k}(p+h)-\zeta_\infty &= \frac{1}{p+h}\int_0^p u(k+\xi)\;d\xi+\frac{1}{p+h}\int_0^h[u(p+k+\xi)-U(p+k+\xi,p)u(p)]\;d\xi\\ &+\left[\frac{h}{p+h}\sigma_k(h,p)-\zeta(p)\right]+[\zeta(p)-\zeta_\infty].\end{aligned}$$

The first term on the right-hand side is bounded by $p/(p+h)\|u\|_\infty$, independently of k. The second term is bounded by $\sup_{q\geq 0}\|u(p+q)-U(p+q,p)u(p)\|$, which is independent of

h and k, and tends to zero as $p \to \infty$. The last term converges strongly to zero as $p \to \infty$. Thus, given any $\varepsilon > 0$, we can choose p_ε large enough so that the second and forth terms are both less than ε. Then, for such p_ε, the first term converges strongly to zero as $h \to \infty$ while the third term τ-converges to zero, both uniformly in k. As a consequence $\overline{u_k}(t)$ is τ-convergent to ζ_∞ as $t \to \infty$ uniformly in k. ∎

Remark 7.14 *Proposition 7.13 was proved in [84] under the following additional assumptions: i) U is an autonomous and strongly continuous* CES, *ii) the set of stationary points is nonempty, and iii) for the weak topology, the space X is assumed to be* weakly complete. *The latter condition means that every weak Cauchy net converges weakly to an element in X. 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 [77]).*

Remark 7.15 *Observe that, compared with Propositions 7.9 and 7.12, the hypotheses and conclusion in Proposition 7.13 are weaker. However, according to Remark 7.7, the two formulations are equivalent whenever the evolution system has bounded almost-orbits. Therefore, for Proposition 7.13 to be useful, one has to prove that the system has at least one bounded almost-orbit. The good news is that, in practice, this step tends to be useful for proving that the orbits are convergent. In fact, in many applications one has to do it anyway.*

7.2.5 Unifying framework: convergence of means with respect to probability measures

In order to present the results of the previous sections in a unified manner and cover standard convergence, convergence in average as well as almost-convergence, we will introduce some general notions of convergence with respect to time-dependent probability measures.

Let μ be a probability measure on $[0,\infty)$. Given $v \in L^\infty_{\text{loc}}(0,\infty;X)$ such that

$$\mu(v) = \int_0^\infty v(\xi)\ d\mu(\xi) = \lim_{T\to\infty} \int_0^T v(\xi)\ d\mu(\xi)$$

exists, we say that v *is μ-integrable.* In such a case, we call $\mu(v)$ the *μ-mean of* v on $[0,\infty)$. More generally, given a family $\{\mu_t\}_{t\geq 0}$ of probability measures on $[0,\infty)$, we say that a function $v \in L^\infty_{\text{loc}}(0,\infty;X)$ is *$\{\mu_t\}$-integrable* if $\mu_t(v)$ exists for all $t \geq 0$.

If there exists $y \in X$ such that the μ_t-mean of v converges strongly (resp. weakly) to y as $t \to \infty$, *i.e.* $y = \tau - \lim_{t\to\infty} \mu_t(v)$ for τ the strong (resp. weak) topology, then we say that v converges strongly (resp. weakly) to y *in μ_t-mean.*

Example 19 *Let $v \in L^\infty_{\text{loc}}(0,\infty;X)$. If $\mu_t = \delta_t$ is the Dirac mass concentrated at t then $\mu_t(v) = v(t)$, and convergence in μ_t-mean is just standard convergence. Now suppose $d\mu_t(\xi) = \frac{1}{t}\chi_{[0,t]}(\xi)d\xi$, where χ_A is the characteristic function of the set A. Then $\mu_t(v) = \frac{1}{t}\int_0^t v(\xi)\ d\xi = \overline{v}(t)$; in this case, convergence in μ_t-mean is precisely convergence in average as defined in Section 7.2.3.*

From now on, we assume that:

$$\text{For all } p \geq 0,\ \mu_t([0,p]) \to 0 \text{ as } t \to \infty. \tag{7.4}$$

The following result can be proved by a direct adaptation of the proof of Proposition 7.12. We leave the details to the reader.

Proposition 7.16 *Let U be an* M-LES *and $\{\mu_t\}_{t\geq 0}$ a family of probability measures satisfying* (7.4). *If the μ_t-mean of every orbit of U converges strongly (resp. weakly) as $t \to \infty$, so does the μ_t-mean of every almost-orbit.*

Next, given $v \in L^\infty_{\text{loc}}(0,\infty;X)$ and $h \geq 0$, we define $v_h(t) = v(h+t)$ for all $t \geq 0$. If there exists $y \in X$ such that

$$y = \tau - \lim_{t\to\infty} \mu_t(v_h) = \tau - \lim_{t\to\infty} \int_0^\infty v(h+\xi)\, d\mu_t(\xi) \quad \text{uniformly in } h \geq 0,$$

for τ the strong (resp. weak) topology, then we say that v converges strongly (resp. weakly) to y *in μ_t-mean, uniformly with respect to translations.*

Example 20 *If μ_t is the Dirac mass at t, then $\mu_t(v_h) = v(t+h)$, and in this case the convergence in μ_t-mean recovers standard convergence which is automatically uniform with respect to translations. Take now $d\mu_t(\xi) = \frac{1}{t}\chi_{[0,t]}(\xi)d\xi$, so that $\mu_t(v_h) = \frac{1}{t}\int_0^t v(h+\xi)\, d\xi$. In this case, convergence in μ_t-mean uniformly with respect to translations is precisely almost-convergence, as defined in Section 7.2.4.*

Proposition 7.17 *Let U be an* M-LES *and $\{\mu_t\}_{t\geq 0}$ a family of probability measures satisfying* (7.4). *If every bounded orbit of U converges strongly (resp. weakly) in μ_t-mean uniformly with respect to translations as $t \to \infty$, so does every bounded almost-orbit.*

This result is a direct extension of Proposition 7.13 whose proof can be easily adapted to this more general setting. We leave the details to the reader.

The previous results can also be stated in terms of almost-equivalence of evolution systems as follows:

Theorem 7.18 *Let U and V be two* M-LES *which are asymptotically almost-equivalent. Consider a family $\{\mu_t\}_{t\geq 0}$ of probability measures on $[0,\infty)$ satisfying* (7.4). *If every orbit of U converges strongly (resp. weakly) in μ_t-mean as $t \to \infty$, so does every orbit of V. If in addition the convergence is uniform with respect to translations for all bounded orbits of U, the same holds true for the bounded orbits of V.*

7.3 Further results on Lipschitz evolution systems

7.3.1 Almost-stationary points

Let U be an ES on C. Let $SP(U) := \{x \in C \mid U(t,s)x = x, \forall t \geq s\}$ be the (possibly empty) set of all the *stationary points* of U. Similarly, denote by $\text{ASP}(U)$ the set of all the *almost-stationary points* of U, that is,

$$\text{ASP}(U) = \left\{ x \in C \;\middle|\; \lim_{t\to\infty} \sup_{h\geq 0} \|U(t+h,t)x - x\| = 0 \right\}. \tag{7.5}$$

Clearly $\text{ASP}(U) \supseteq SP(U)$. If U is autonomous then it is easy to see that $\text{ASP}(U) = SP(U)$, but this is not the case in general even for a *CES* (take, for instance, $U(t,s)x = x + e^{-s} - e^{-t}$, $x \in \mathbf{R}$, for which $\text{ASP}(U) = \mathbf{R}$ and $SP(U) = \emptyset$).

Remark 7.19 *An interesting characterization of $SP(U)$ is given in [102] when U is given by a semigroup T of contractions on a weakly compact subset of a Banach space having the Opial property: The authors prove that $z \in SP(U)$ if, and only if there is a sequence $t_n \to \infty$ such that $w\text{-}\lim_{n\to\infty} \frac{1}{t_n}\int_0^{t_n} T(s)z dz = z$. A similar result using strong limits is given in [103]. A challenging task is to extend this characterization to nonautonomous systems.*

Observe that if $x^* \in \mathrm{ASP}(U)$, the constant function $u(t) \equiv x^*$ is a bounded almost-orbit of U. Therefore, according to Remark 7.7, if $\mathrm{ASP}(U) \neq \emptyset$ every almost-orbit of U is bounded.

Next, we turn our attention to some closedness and convexity properties of $\mathrm{ASP}(U)$.

Lemma 7.20 *Suppose C is closed for the strong topology. If U is an* M-LES *on C then $ASP(U)$ is closed for the strong topology.*

Proof. If $\{x_n\}$ is a sequence in $\mathrm{ASP}(U)$ converging to $x \in X$ then $x \in C$ and

$$\|U(t+h,t)x - x\| \leq (M+1)\|x_n - x\| + \|U(t+h,t)x_n - x_n\|.$$

Therefore

$$\limsup_{t\to\infty} \left[\sup_{h\geq 0} \|U(t+h,t)x - x\|\right] \leq (M+1)\|x_n - x\|.$$

Letting $n \to \infty$, we conclude that $x \in \mathrm{ASP}(U)$. ■

By virtue of Remark 7.3, if U is an *M-LES* then $\mathrm{ASP}(U)$ contains all the limits of the strongly convergent almost-orbits of U, if there is any. For weak limits we have the following:

Corollary 7.21 *Let C be closed for the strong topology. If all the weak limits of the orbits of an* M-LES *U lie in $ASP(U)$, the same is true for all the weak limits of the almost-orbits of U.*

Proof. With the notation introduced in the proof of Proposition 7.9, as $t \to \infty$, $u(t)$ converges to ζ_∞, which is the strong limit of weak limits of orbits of U. ■

Remark 7.22 *If U is an autonomous* M-LES *whose orbits converge weakly to stationary points, then the weak limits of almost-orbits are also stationary points of U.*

Given a convex nonempty subset K of X and a continuous strictly increasing function $\gamma : \mathbf{R}_+ \to \mathbf{R}_+$ such that $\gamma(0) = 0$, we say that a function $F : K \to X$ is *of type* γ if for all $x, y \in K$ and $\lambda \in [0,1]$ we have

$$\gamma(\|F(\lambda x + (1-\lambda)y) - \lambda F(x) - (1-\lambda)F(y)\|) \leq \|x - y\| - \|F(x) - F(y)\|.$$

Proposition 7.23 *Let U be a* CES *on a convex set C and suppose that there exists γ such that for each $t \geq s$, $U(t,s)$ is of type γ on a convex set K containing $ASP(U)$. Then $ASP(U)$ is convex.*

Proof. Take $x_1, x_2 \in \mathrm{ASP}(U)$ and define $\psi_i(t) = \sup_{h\geq 0} \|U(t+h,t)x_i - x_i\|$ for $i = 1, 2$. Now take $\lambda \in (0,1)$ and set $z = \lambda x_1 + (1-\lambda)x_2$. We have

$$\begin{aligned}
\|U(t+h,t)z - z\| &\leq \|U(t+h,t)z - \lambda U(t+h,t)x_1 - (1-\lambda)U(t+h,t)x_2\| \\
&\quad + \lambda\psi_1(t) + (1-\lambda)\psi_2(t) \\
&\leq \gamma^{-1}\left(\|x_1 - x_2\| - \|U(t+h,t)x_1 - U(t+h,t)x_2\|\right) \\
&\quad + \lambda\psi_1(t) + (1-\lambda)\psi_2(t) \\
&\leq \gamma^{-1}\left(\psi_1(t) + \psi_2(t)\right) + \lambda\psi_1(t) + (1-\lambda)\psi_2(t)
\end{aligned}$$

Letting $t \to \infty$ we get the result. ■

If X is a uniformly convex Banach space and K is bounded and convex, then there exists γ such that every nonexpansive function $F : K \to X$ is of type γ (see [39, Lemma 1.1]). In particular, if U is a *CES* on a convex set C then for every bounded subset $A \subset C$, there exists γ such that for each $t \geq s$, $U(t,s)$ is of type γ on $\text{co}(A)$, the convex hull of A. Hence, we deduce the following:

Corollary 7.24 *Let U be a* CES *on a convex subset C of a uniformly convex Banach space X. Then ASP(U) is convex.*

Proof. Let $x_1, x_2 \in \text{ASP}(U)$. The line segment K joining x_1 and x_2 is bounded and convex, so there is a function γ such that the restriction of $U(t,s)$ to K is of type γ for all $t \geq s \geq 0$. The argument in Proposition 7.23 gives

$$\|U(t+h,t)z - z\| \leq \gamma^{-1}\left(\psi_1(t) + \psi_2(t)\right) + \lambda\psi_1(t) + (1-\lambda)\psi_2(t)$$

and so $z \in \text{ASP}(U)$. ■

Remark 7.25 *Proposition 7.23 is valid if we replace ASP(U) with the set of all almost-orbits. The proof is quite similar and uses Corollary 7.6 iii) for concluding. As a consequence, if X is uniformly convex and U is a* CES *on a bounded convex set C, then the set of all almost-orbits is convex.*

7.3.2 On strongly contracting evolution systems

Consider a family $\{M(t,s) \ : \ t \geq s \geq 0 \ \}$ of positive numbers satisfying $\lim\limits_{t\to\infty} M(t,s) = 0$ for each s. If an evolution system U on C satisfies $\|U(t,s)x - U(t,s)y\| \leq M(t,s)\|x-y\|$ for all $x, y \in C$ and $t \geq s \geq 0$ we say it is a *strongly contracting evolution system* (*SCES*).

Example 21 *Let $\{A(t) : t \geq 0\}$ be a family of maximal monotone operators on a subset C of a Hilbert space H such that for $[x,x^*], [y,y^*] \in A(t)$ we have*

$$\langle x^* - y^*, x - y\rangle \geq \alpha(t)\|x-y\|^2, \tag{7.6}$$

where $\alpha \in L^1_{loc}$. Set $\mu(t) = \exp\left(\int_0^t \alpha(\tau)\ d\tau\right)$. The function μ is positive and nondecreasing. Assume the differential inclusion

$$u'(t) + A(t)u(t) \ni 0$$

has a solution for every initial condition $u(s) = u_s \in C$. This defines an evolution system U that satisfies

$$\|U(t,s)x - U(t,s)y\| \leq \mu(t)^{-1}\mu(s)\|x-y\|.$$

U is a SCES *if, and only if, $\alpha \notin L^1$. If $A(t) = \partial f_t$, where $f_t(x) = f(x) + \alpha(t)\|x\|^2/2$ this is the Tikhonov approximation with a "slow" parameterization (see [6]).*

A *SCES* is clearly a *CES* so Corollary 7.6 holds. However, much more can be said:

Proposition 7.26 *Let U be a* SCES. *We have the following:*

i) If u_1 and u_2 are almost-orbits of U then $\lim\limits_{t\to\infty} \|u_1(t) - u_2(t)\| = 0$;

ii) The set ASP(U) has at most one element; and

iii) If $ASP(U) \neq \emptyset$, then every almost-orbit of U converges strongly to the unique $x^* \in ASP(U)$.

Proof. For i), the argument in Lemma 7.5 gives

$$\|u_1(t+s) - u_2(t+s)\| \leq \psi_1(t) + \psi_2(t) + M(t+s,t)\|u_1(t) - u_2(t)\|.$$

Letting $s \to \infty$ we have $\limsup_{s\to\infty} \|u_1(s) - u_2(s)\| \leq \psi_1(t) + \psi_2(t)$ and as $t \to \infty$ we get $\lim_{s\to\infty} \|u_1(s) - u_2(s)\| = 0$. Parts ii) and iii) are a trivial consequence. ■

7.4 Non-Lipschitz evolution systems

For general evolution systems, having a bounded almost-orbit does not imply that all the almost-orbits will be bounded as well. For example, consider the system U given by the solutions to the differential equation: $u' = u$ with initial condition $u(s) = x$; i.e., $U(t,s)x = e^{(t-s)}x$. It is clearly an evolution system and the only bounded orbits are those with $x = 0$. However, other qualitative properties of the orbits also hold for almost-orbits.

The first work that contains equivalence results for evolution systems which are not contracting seems to be [83]. However, they only consider strongly continuous semigroups which are "asymptotically nonexpansive in the intermediate sense". We shall not enter into the details but just underscore the fact that all the existing results concerning asymptotic equivalence hold for evolution systems having some additional (and strong!) regularity with respect both to time and space.

7.4.1 Convergence in $\{\mu_t\}$-mean

In the following, we use the terminology and notation introduced in Section 7.2.5. Let $\{\mu_t\}_{t\geq 0}$ be a family of probability measures on $[0,\infty)$. Suppose:

Hypothesis **H** For each $\{\mu_t\}$-integrable function g satisfying $\lim_{t\to\infty} \int_0^\infty g(\xi)\, d\mu_t(\xi) = L$, each $\varepsilon > 0$ and $K > 0$ there exists $T > 0$ such that for all $t \geq T$ one has $\left\| \int_0^\infty g(\xi)\, d\mu_t(\xi + K) - L \right\| < \varepsilon$.

Observe that the families described in Example 19 do satisfy Hypothesis **H**: This is trivial if μ_t is the Dirac mass at t. If $d\mu_t(\xi) = \frac{1}{t}\chi_{[0,t]}(\xi)$, then for t large enough

$$\int_0^\infty g(\xi)\, d\mu_t(\xi + K) = \left(\frac{t-K}{t}\right) \frac{1}{t-K} \int_0^{t-K} g(\xi)\, d\xi,$$

which tends to L as $t \to \infty$.

The fact that $\mu_t(B) \to 0$ as $t \to \infty$ for every bounded set B does not guarantee that Hypothesis **H** will hold, as is shown by the following example:

Example 22 *Define $n(\xi) = \sum_{k\geq 0} \chi_{[2k,2k+1)}(\xi)$ and $\hat{n}(\xi) = n(\xi+1)$ so that $n^2 \equiv n$ and $n\hat{n} \equiv 0$. Let $d\mu_t(\xi) = \alpha^{-1}(t)n(\xi)\chi_{[0,t]}(\xi)d\xi$, where $\alpha(t) = \int_0^t n(\xi)d\xi$. The family $\{\mu_t\}$ satisfies $\mu_t(B) \to 0$ for every bounded set B (this is obvious) but does not fulfill Hypothesis* **H**. *To see this, simply notice that $\int_0^\infty n(\xi)d\mu_t(\xi) = 1$ while $\int_0^\infty n(\xi)d\mu_t(\xi+1) = \alpha^{-1}\int_1^{t-1} \hat{n}(\xi)n(\xi)d\xi = 0$ for all t.*

Theorem 7.27 *Let U be an ES and assume that the family $\{\mu_t\}$ satisfies Hypothesis* **H**. *If $U(t,s)x$ converges strongly in μ_t-mean for all x and s, then so does every $\{\mu_t\}$-integrable almost-orbit.*

Proof. Suppose u is a $\{\mu_t\}$-integrable almost-orbit of U and let $\varepsilon > 0$. Choose $S > 0$ such that

$$\sup_{h\geq 0} \|u(t+h) - U(t+h,t)u(t)\| < \varepsilon/6$$

for all $t \geq S$. Next define

$$\zeta(S) = \lim_{t\to\infty} \int_0^\infty U(S+\xi,S)u(S)\; d\mu_t(\xi).$$

By hypothesis, there is T_1 such that $\left\|\zeta(S) - \int_0^\infty U(S+\xi,S)u(S)\; d\mu_t(\xi)\right\| < \varepsilon/6$ for all $t \geq T_1$.

We have

$$\begin{aligned}\|\mu_t(u) - \zeta(S)\| \;\leq\; & \int_0^S \|u(\xi)\|\; d\mu_t(\xi) + \int_S^\infty \|u(\xi) - U(\xi,S)u(S)\|\; d\mu_t(\xi)\\ & + \left\|\zeta(S) - \int_0^\infty U(S+\xi,S)u(S)\; d\mu_t(\xi+S)\right\|.\end{aligned}$$

For the first term, since $\mu_t([0,S]) \to 0$ as $t\to\infty$, we can take T_2 such that $\mu_t([0,S]) < \varepsilon/6C$ for all $t \geq T_2$, where $C = \sup_{0\leq\xi\leq S}\|u(\xi)\|$. The second term is always less than $\varepsilon/6$. Finally, use Hypothesis **H** to find T_3 such that the last term is less than $\varepsilon/6$ whenever $t \geq T_3$. Hence if $t \geq T = \max\{\, T_1,\; T_2,\; T_3 \,\}$, we have $\|\mu_t(u) - \zeta(S)\| < \varepsilon/2$ for all $h \geq 0$.

We have found $T > 0$ such that $\|\mu_t(u) - \mu_s(u)\| < \varepsilon$ for all $t, s \geq T$ and therefore $\mu_t(u)$ converges to some y as $t \to \infty$. ■

Remark 7.28 *If $\mu_t = \delta_t$ the preceding argument gives an alternative proof of the assertion concerning the strong topology in Proposition 7.9 without the Lipschitz assumption. Moreover, this proof is much simpler because we do not make use of the intermediate step of proving the convergence of the net $\zeta(p)$. However, this argument fails when dealing with the weak topology in arbitrary Banach spaces (see Theorem 7.29 below).*

For the corresponding result in the weak topology we need the following assumption, which is also satisfied by the families described in Example 19:

Hypothesis **w-H**: For each $\{\mu_t\}$-integrable function g satisfying $\text{w-}\lim_{t\to\infty} \int_0^\infty g(\xi)\; d\mu_t(\xi) = L$, each $\varepsilon > 0$, $K > 0$ and $f \in X^*$ there exists $T > 0$ such that for all $t \geq T$ one has $\left|\langle\; \int_0^\infty g(\xi)\; d\mu_t(\xi+K) - L, f\;\rangle\right| < \varepsilon$.

One can easily see that under Hypothesis **w-H** the argument in the proof of Theorem 7.27 implies that $\mu_t(u)$ has the Cauchy property for the weak topology. More precisely,

$$\lim_{t,s\to\infty} \langle \mu_t(u) - \mu_s(u), \phi\rangle = 0 \text{ for each } \phi \in X^*.$$

If the space is weakly complete (see Remark 7.14) the net $\{\mu_t(u)\}$ must converge weakly. We have the following:

Theorem 7.29 *Let U be an ES on a subset of a weakly complete Banach space X and assume that the family $\{\mu_t\}$ satisfies Hypothesis* **w-H**. *If $U(t,s)x$ converges weakly in μ_t-mean for all x and s, then so does every $\{\mu_t\}$-integrable almost-orbit.*

Remark 7.30 *For bounded almost-orbits it suffices to assume that the space is weakly quasi-complete, which means that every bounded Cauchy net is weakly convergent.*

7.4.2 Convergence in μ_t-mean, uniform with respect to translations

Recall that a $\{\mu_t\}$-integrable function v converges strongly (resp. weakly) to $y \in X$ in μ_t-mean, uniformly with respect to translations if $\mu_t(v_h)$ converges strongly (resp. weakly) to y as $t \to \infty$ uniformly in $h \geq 0$. Recall also that this notion includes convergence in the traditional sense of the term and almost-convergence, when the family $\{\mu_t\}$ is one of those presented in Example 19. The uniformity in $h \geq 0$ requires a slightly stronger assumption on the family of measures in order to prove the equivalence results. The reader may easily verify that the following hypotheses hold for the families described in Example 19:

Hypothesis $\mathbf{H_u}$: For each $\{\mu_t\}$-integrable function g satisfying $\lim_{t\to\infty} \int_0^\infty g(\xi)\, d\mu_t(\xi) = L$, each $\varepsilon > 0$ and $K > 0$ there exists $T > 0$ such that for all $t \geq T$ and $k \in [0, K]$ one has $\left\| \int_0^\infty g(\xi)\, d\mu_t(\xi + k) - L \right\| < \varepsilon$.

Hypothesis $\mathbf{w\text{-}H_u}$: For each $\{\mu_t\}$-integrable function g such that $\text{w-}\lim_{t\to\infty} \int_0^\infty g(\xi)\, d\mu_t(\xi) = L$, each $\varepsilon > 0$, $K > 0$ and $f \in X^*$ there exists $T > 0$ such that for all $t \geq T$ and $k \in [0, K]$ one has $\left|\langle \int_0^\infty g(\xi)\, d\mu_t(\xi + k) - L, f \rangle\right| < \varepsilon$.

Theorem 7.31 *Let U be an ES on a subset of a Banach space X*

i) Assume the family $\{\mu_t\}$ satisfies Hypothesis $\mathbf{H_u}$. *If $U(t,s)x$ converges strongly in μ_t-mean, uniformly with respect to translations for all x and s, then so does every $\{\mu_t\}$-integrable almost-orbit.*

ii) Assume X is weakly complete and the family $\{\mu_t\}$ satisfies Hypothesis $\mathbf{w\text{-}H_u}$. *If $U(t,s)x$ converges weakly in μ_t-mean uniformly with respect to translations for all x and s, then so does every $\{\mu_t\}$-integrable almost-orbit.*

Proof. We do the strong case here and leave the rest to the reader. Suppose u is a $\{\mu_t\}$-integrable almost-orbit of U and let $\varepsilon > 0$. Choose $S > 0$ such that

$$\sup_{h\geq 0} \|u(t+h) - U(t+h,t)u(t)\| < \varepsilon/6$$

for all $t \geq S$. Next define

$$\zeta(S) = \lim_{t\to\infty} \int_0^\infty U(S+\xi, S)u(S)\, d\mu_t(\xi).$$

By hypothesis, there is T_1 such that $\left\|\zeta(S) - \int_0^\infty U(S+h+\xi, S)u(S)\, d\mu_t(\xi)\right\| < \varepsilon/6$ for all $t \geq T_1$ and $h \geq 0$ (recall that the convergence is uniform in $h \geq 0$). We divide the rest of the proof in two parts:

$\underline{0 \le h \le S}$: The argument is similar to the one we presented in the proof of Theorem 7.27. We have

$$\begin{aligned}\|\mu_t(u_h) - \zeta(S)\| \;&\le\; \int_0^{S-h} \|u(h+\xi)\|\ d\mu_t(\xi) + \int_{S-h}^{\infty} \|u(h+\xi) - U(h+\xi, S)u(S)\|\ d\mu_t(\xi) \\ &\quad + \left\|\zeta(S) - \int_0^{\infty} U(S+\xi, S)u(S)\ d\mu_t(\xi + (S-h))\right\|.\end{aligned}$$

For the first term, since $\mu_t([0,S]) \to 0$ as $t \to \infty$, we can take T_2 such that $\mu_t([0,S]) < \varepsilon/6C$ for all $t \ge T_2$, where $C = \sup_{0\le\xi\le S} \|u(\xi)\|$. The second term is always less than $\varepsilon/6$. Finally, use Hypothesis $\mathbf{H_u}$ to find T_3 such that the last term is less than $\varepsilon/6$ whenever $t \ge T_3$. Hence if $t \ge T = \max\{\ T_1,\ T_2,\ T_3\ \}$, we have $\|\mu_t(u_h) - \zeta(S)\| < \varepsilon/2$ for all $h \ge 0$.

$\underline{h \ge S}$: If $t \ge T_1$ we have

$$\begin{aligned}\|\mu_t(u_h) - \zeta(S)\| \;&\le\; \int_0^{\infty} \|u(h+\xi) - U(h+\xi, S)u(S)\|\ d\mu_t(\xi) \\ &\quad + \left\|\zeta(S) - \int_0^{\infty} U(h+\xi, S)u(S)\ d\mu_t(\xi)\right\|.\end{aligned}$$

Since each term is less than $\varepsilon/6$ we have $\|\mu_t(u_h) - \zeta(S)\| < \varepsilon/3 < \varepsilon/2$ for all $t \ge T_1$ and all $h \ge S$.

Finally, $\|\mu_t(u_h) - \zeta(S)\| < \varepsilon/2$ for all $t \ge T$ and $h \ge 0$. This implies $\|\mu_t(u_h) - \mu_s(u_k)\| < \varepsilon$ for all $t, s \ge T$ and $h, k \ge 0$ and so u is strongly convergent in μ_t-mean, uniformly with respect to translations. ■

7.5 Further results on general evolution systems

7.5.1 Uniform continuity

If U is an autonomous *CES*, then every orbit is uniformly continuous. According to [84], so is every continuous almost-orbit. Their proof uses the contracting property, which we show to be unnecessary. The key lies on the time-dependence.

Proposition 7.32 *Let U be an evolution system. If every orbit is uniformly continuous, so is every continuous almost-orbit.*

Proof. Let u be a continuous almost-orbit of U and $\varepsilon > 0$. First, take $T > 0$ such that $\|u(\tau) - U(\tau, T)u(T)\| < \varepsilon/3$ for all $\tau \ge T$. Since u is continuous, it is uniformly continuous on $[0, T+1]$. Hence there is $\delta_1 > 0$ such that for every $t, s \in [0, T+1]$ satisfying $|t-s| < \delta_1$ one has $\|u(t) - u(s)\| < \varepsilon$. Now consider the function $\tau \mapsto U(\tau, T)u(T)$ defined for $\tau \ge T$. By hypothesis it is uniformly continuous, so there is $\delta_2 > 0$ such that for every $t, s \ge T$ such that $|t-s| < \delta_2$ one has $\|v(t, T)u(T) - U(s, T)u(T)\| < \varepsilon/3$. Therefore, if $t, s \ge T$ and $|t-s| < \delta_2$ we have

$$\begin{aligned}\|u(t) - u(s)\| \;&\le\; \|u(t) - U(t,T)u(T)\| + \|U(s,T)u(T) - u(s)\| \\ &\quad + \|U(t,T)u(T) - U(s,T)u(T)\| \\ &<\; \varepsilon.\end{aligned}$$

Finally, for $t, s \in [0, \infty)$ with $|t-s| < \min\{\delta_1, \delta_2, 1\}$ we have $\|u(t) - u(s)\| < \varepsilon$. ■

7.5.2 On cluster points

The following result reveals some kind of asymptotic invariance under U for the ω-limit sets of almost-orbits. Observe, in particular, that if U is autonomous then the strong ω-limit set of any almost-orbit is invariant under U.

Proposition 7.33 *Let u be an almost-orbit of an evolution system U and suppose $\{s_n\}$ is a strictly increasing sequence such that $\lim_{n\to\infty} s_n = \infty$ and $\tau\text{-}\lim_{n\to\infty} u(s_n) = x^*$. We have the following:*

i) There exist a sequence $\{t_n\}$ of positive numbers and a sequence $\{x_n\}$ in C such that $\tau\text{-}\lim_{n\to\infty} x_n = x^*$, $t_n > s_n$ and $\tau\text{-}\lim_{n\to\infty} U(t_n, s_n)x_n = x^*$. The sequence $\{t_n\}$ can be chosen such that $\lim_{n\to\infty}(t_n - s_n) = \infty$.

ii) If U is M-Lipschitz and τ is the strong topology, then $\lim_{n\to\infty} U(t_n, s_n)x^* = x^*$.

Proof. For the fisrt part, let $\varphi : \mathbf{N} \to \mathbf{N}$ be any positive function and set $h_n = s_{n+\varphi(n)} - s_n$. Write $t_n = s_n + h_n$ and $x_n = u(s_n)$. Since u is an almost orbit of V, for every $\varepsilon > 0$ there is $N \geq 0$ such that $\|u(s_n + h_n) - U(s_n + h_n, s_n)u(s_n)\| < \varepsilon$ for all $n \geq N$. Therefore

$$U(t_n, s_n)x_n - x^* = U(s_n + h_n, s_n)u(s_n) - u(s_n + h_n) + u(s_{n+\varphi(n)}) - x^*$$

tends to zero for the topology τ. Clearly φ can be chosen so that $s_{n+\varphi(n)} - s_n$ tends to ∞ as $n \to \infty$.
For the second part, notice that

$$\begin{aligned} \|U(t_n, s_n)x^* - x^*\| &\leq \|U(t_n, s_n)x^* - U(t_n, s_n)x_n\| + \|U(t_n, s_n)u(s_n) - u(t_n)\| \\ &\quad + \|u(t_n) - x^*\| \\ &\leq M\|u(s_n) - x^*\| + \|U(t_n, s_n)u(s_n) - u(t_n)\| + \|u(t_n) - x^*\|, \end{aligned}$$

which tends to zero as $n \to \infty$. ■

Chapter 8

Applications of the asymptotic almost-equivalence theory

8.1 Autonomous evolution systems

Let $(X, \|\cdot\|)$ be a Banach space and denote by X^* its dual. The duality product $\langle\cdot,\cdot\rangle : X \times X^* \to \mathcal{R}$ is defined by $\langle u, f\rangle = f(u)$ for all $u \in X$ and $f \in X^*$. The duality mapping $\mathcal{J} : X \rightrightarrows X^*$ is defined by $\mathcal{J}(u) = \{\ f \in X^* \ :\ \|f\|_* = \|u\| \text{ and } \langle u, f\rangle = \|u\|^2\ \}$.

Given a set-valued operator $A : X \rightrightarrows X$, its domain is given by $D(A) = \{u \in X \mid Au \neq \emptyset\}$. Sometimes we will identify A with its graph by writing $[u, v] \in A$ for $v \in Au$. For $u \in D(A)$ we set

$$\|Au\| = \inf_{[u,v]\in A} \|v\|.$$

For $\lambda > 0$, the *resolvent* of A is defined by $J_\lambda = (I + \lambda A)^{-1}$. If J_λ is nonexpansive for every λ, A is said to be *accretive*. If moreover J_λ is everywhere defined, A is m-accretive. Accretivity can be characterized as follows (see [63]): For all $[u_1, v_1], [u_2, v_2] \in A$ there exists $f \in \mathcal{J}(u_1 - u_2)$ such that $\langle v_1 - v_2, f\rangle \geq 0$. In Hilbert spaces, this property is called *monotonicity.*

8.1.1 Differential inclusion

Every m-accretive operator A defines a strongly continuous contraction semigroup S via $S(t)x = u(t)$, where u is the solution to the differential inclusion

$$\begin{cases} u' + Au \ni 0 & \text{a.e. on } (0,\infty) \\ u(0) = x \end{cases} \tag{8.1}$$

(see [47]). As we mentioned in Example 15, this defines a *CES* that we denote by $U(t, s)$.

Remark 8.1 *The asymptotic behavior of* (8.1) *has been studied by many authors. Assume that A is a m-monotone operator in Hilbert space and that $\mathcal{S} = A^{-1}0 \neq \emptyset$.*

1. *Every orbit of U* **converges weakly** *to a point in $\mathcal{S}$ if A is demipositive ([36]). This is not the case in general, the $\pi/2$-rotation in the plain gives a counterexample.*

2. *Every orbit of U* **converges strongly** *to a point in $\mathcal{S}$ if*

   (a) A is strongly monotone; or

   (b) $\mathcal{S}$ has nonempty interior; or

(c) $A = \partial f$, whenever f is proper, lower-semicontinuous, convex and even ([36]).

3. *Every orbit of U* **converges weakly in average** *to a point in $\mathcal{S}$ ([19]).*

8.1.2 The proximal point algorithm

Let A be a m-accretive operator, x_0 a point in X and $\{\lambda_n\}$ a bounded sequence of positive numbers. A *proximal sequence* is a sequence $\{x_n\}$ satisfying

$$x_{n-1} - x_n \in \lambda_n A x_n \tag{8.2}$$

Equivalently, $x_n = J_{\lambda_n}(x_{n-1})$. We define a *CES* $V(t,s)$ as in Example 16.

Remark 8.2 *The system* (8.2) *has also been studied extensively. We assume that A is a maximal monotone operator in Hilbert space and that $\mathcal{S} \neq \emptyset$.*

1. *Every orbit of U* **converges weakly** *to a point in $\mathcal{S}$ ([33]) if*
   - *(a) $\{\lambda_n\} \notin \ell^2$; or*
   - *(b) A is demipositive and $\{\lambda_n\} \notin \ell^1$ (again, it is false in general).*
2. *Every orbit of U* **converges strongly** *to a point in $\mathcal{S}$ if*
   - *(a) A is strongly monotone; or*
   - *(b) $\mathcal{S}$ has nonempty interior; or*
   - *(c) $A = \partial f$, whenever f is proper, lower-semicontinuous, convex and even ([33]).*
3. *Every orbit of U* **converges weakly in average** *to a point in $\mathcal{S}$ if $\{\lambda_n\} \in \ell^2 \setminus \ell^1$ ([78]).*
4. *Every orbit of U* **converges strongly in average** *to a point in $\mathcal{S}$ if A is odd and $\{\lambda_n\} \notin \ell^1$ ([78]).*

8.1.3 Results on asymptotic equivalence

The first work that relates the convergence properties of (8.1) and (8.2) (in any Banach space) seems to be [90]. The author first proves that if U and V are equivalent, then the orbits of U converge strongly or weakly if, and only if, the orbits of V do. This is a particular case of Theorem 7.9 above. Next he proves that the systems *are* equivalent in two special cases:

i) The operator A is single-valued and Lipschitz and $\{\lambda_n\} \in \ell^2 \setminus \ell^1$; or

ii) The sequence $\{\lambda_n\}$ satisfies the following sophisticated[1] summability condition:

   - $\{\lambda_n\} \notin \ell^1$; and
   - There exists a subsequence $\{k_i\}_i$ of $\mathbf{N}$ such that $\sum \sqrt{b(k_i, k_{i+1})} < \infty$, where $b(n,m) = \max\{\lambda_k \mid n < k \leq m\} \cdot \sum_{k=n+1}^{m} \lambda_k$.[2]

[1] This condition may seem unnatural at a first glance but it was motivated by an older result from Theorem A.1 in [46].

[2] This condition implies $\{\lambda_n\} \in \ell^2 \setminus \ell^1$. Take $N \in \mathbf{N}$ and set i_N such that $k_{i_N} \geq N$. We have

$$\sum_{k=1}^{N} \lambda_k^2 \leq \sum_{i=1}^{i_N - 1} \sum_{j=k_i+1}^{k_{i+1}} \lambda_j^2 \leq \sum_{i=1}^{i_N-1} b(k_i, k_{i+1}).$$

A few years later, the systems are proved to be equivalent for any m-accretive operator A whenever $\{\lambda_n\} \in \ell^2 \setminus \ell^1$ (see [70]). This implies that in Hilbert space we can obtain some of the weak and strong convergence results mentioned in Remark 8.2 as a consequence of those in Remark 8.1 and viceversa. The same can be done with the results concerning strong and weak convergence in average by using our Proposition 7.12 and so on.

Finally in [56] the author proves that if A is the subdifferential of a proper, lower-semicontinuous convex function then the condition $\{\lambda_n\} \notin \ell^1$ suffices to guarantee the equivalence of the systems. A remarkable consequence of this particular result is the existence of a proper, lower-semicontinuous convex funtion in a Hilbert space such that the proximal point algorithm converges weakly but not strongly, thus giving an answer to a question posed earlier in [97]. This was done by using the well-known counterexample in [17] for the continuous-time scheme and translating the result to the discretization by means of the equivalence theorem.

8.1.4 Euler's scheme

No results of this kind were found in the literature for the explicit discretization. Let A be a m-accretive operator and $\{\lambda_n\}$ a sequence of positive numbers bounded by 1. Starting from $z_0 \in X$ define the sequence $\{z_n\}$ by

$$z_{n+1} \in z_n - \lambda_n A z_n \tag{8.3}$$

for $n \geq 0$. Now suppose $A = I - T$, where T is a nonexpansive mapping. If $\{\lambda_n\}$ and $\{\widehat{\lambda}_n\}$ are sequences in $(0, 1]$, $\{z_n\}$ and $\{\widehat{z}_n\}$ are defined by (8.3), then for any $u \in X$ and $n, m \in \mathbf{N}$ we have

$$\|z_n - \widehat{z}_m\| \leq \|z_0 - u\| + \|\widehat{z}_0 - u\| + \|Au\|\sqrt{(\sigma_n - \widehat{\sigma}_m)^2 + \tau_n + \widehat{\tau}_m}. \tag{8.4}$$

Here $\sigma_n = \sum_1^n \lambda_k$, $\tau_n = \sum_1^n \lambda_k^2$ and respectively for $\widehat{\sigma}_m$ and $\widehat{\tau}_m$. The previous inequality was pointed out by [105]. It is a Kobayashi-type inequality for this explicit discretization and the proof is essentially the same as in Lemma 2.1 in [66].

Let U be the evolution system given by the solutions of the differential inclusion (8.1), which now takes the form $u + u' = Tu$ with initial condition $u(s) = x \in X$ and let W be the one defined by Euler's scheme as in Example 16.

Lemma 8.3 *Suppose $\{\lambda_n\} \in \ell^2 \setminus \ell^1$. Every bounded orbit of U (rep. W) is an almost-orbit of W (rep. U).*

Proof. If $n \geq k$ and $t \geq s$, by passing to the limit in inequality (8.4) we get

$$\|z_n - u(t)\| \leq \|z_k - u(s)\| + \|Ay\|\sqrt{[(\sigma_n - \sigma_k) - (t - s)]^2 + \tau_n - \tau_k}, \tag{8.5}$$

where y can be replaced by z_k or $u(s)$ and $u(\cdot)$ is a solution of (8.1) (the convergence to the continuous-time trajectory u follows as in the proof of Proposition 2.5). Assume u is bounded. Then so is $\|Au(\cdot)\|$ and

$$\begin{aligned} \|W(t+h,t)u(t) - u(t+h)\| &\leq C\sqrt{\left[\sigma_{\nu(t+h)} - \sigma_{\nu(t)} - h\right]^2 + \tau_{\nu(t+h)} - \tau_{\nu(t)}} \\ &\leq C\sqrt{3\sum_{j\geq\nu(t)} \lambda_j^2}, \end{aligned}$$

which tends to 0 as $t \to \infty$ uniformly in $h \geq 0$. The converse is similar. ■

Theorem 8.4 *Let $A = I - T$, with T nonexpansive. Suppose $\{\lambda_n\}$ is a sequence in $(0, 1]$ such that $\{\lambda_n\} \in \ell^2 \setminus \ell^1$. Then the solutions of (8.1) converge weakly (or strongly) as $t \to \infty$ for each initial condition $x \in X$ if, and only if, the sequences z_n defined by (8.3) converge weakly (or strongly) as $n \to \infty$ for all $z_0 \in X$.*

Proof. Let $u(\cdot)$ be a solution of (8.1) and suppose it converges as $t \to \infty$ for some initial condition $x \in X$: Then it is bounded and so it is an almost-orbit of W by Lemma 8.3. By Corollary 7.6, every orbit of W is bounded as well. Applying Lemma 8.3 one more time one deduces that every orbit of W must be an almost-orbit of U. A similar argument holds in the opposite direction and so, U and W are equivalent. ■

Corollary 8.5 *Let T be a nonexpansive mapping on a Hilbert space having a fixed point. Suppose $\{\lambda_n\}$ is a sequence in $(0, 1]$ such that $\{\lambda_n\} \in \ell^2 \setminus \ell^1$. Then the sequence $\{x_n\}$ defined by*

$$x_n = \lambda_n T(x_{n-1}) + (1 - \lambda_n)x_{n-1} \tag{8.6}$$

converges weakly to a fixed point of T.

Proof. Since $I - T$ is demipositive it suffices to use Theorems 3.11 and 8.4. ■

Sequences satisfying (8.6) are often referred to as *Mann's iterations*. In [80] the author proved convergence in $\mathbf{R}$ when $\lambda_n = \frac{1}{n+1}$. A great number of extensions have been studied thereafter.

8.2 Nonautonomous evolution systems

8.2.1 Tikhonov's regularization in a nonautonomous setting

Let $A(t)$ be a family of maximal monotone operators on a Hilbert space H and $\varepsilon \in L^1(0, \infty; \mathbf{R})$.[3] Consider the differential inclusions

$$u'(t) + A(t)u(t) \ni 0 \qquad \text{and} \qquad v'(t) + A(t)v(t) + \varepsilon(t)v(t) \ni 0.$$

We are not interested in existence results here, so we shall assume that for each initial condition in H both problems have solutions and they are bounded. The interested reader may consult [49], [67], [8] or [64]. Denote respectively by U and V the corresponding evolution systems. Assume also that for every $R > 0$ there exists $M > 0$ such that $\|x\| \leq R$ implies $\|U(t, s)x\| \leq M$. This occurs, for instance, if $A(t) \equiv A$ and $\mathcal{S} \neq \emptyset$.

Proposition 8.6 *Every bounded orbit of V is an almost-orbit of U.*

Proof. Let $y(t) = V(t, t_0)v_0$ be an orbit of V. Define $X_s(t) = U(t, s)y(s)$. We shall prove that $\|X_t(t + h) - y(t + h)\|$ tends to 0 as $t \to \infty$ uniformly in $h \geq 0$. Since y is bounded by some $R > 0$, $X_s(t) \leq M$ for all $s \leq t$ by hypothesis. Fix t and define $\psi(h) = \frac{1}{2}\|X_t(t+h) - y(t+h)\|^2$. From the properties of the inner product we have $\langle \zeta, \zeta - \xi \rangle \geq$

[3]Observe that $\varepsilon(\cdot)$ need not be nonnegative.

$-\frac{1}{4}\|\xi\|^2$ for all $\zeta, \xi \in H$. Therefore, $\psi'(h) \leq \frac{1}{4}|\varepsilon(t+h)|\|X_t(t+h)\|^2$ for almost every $h > 0$. Integrating from 0 to h and observing that $\psi(0) = 0$ we obtain

$$\|X_t(t+h) - y(t+h)\|^2 \leq \frac{1}{4}\int_t^{t+h} |\varepsilon(\tau)|\ \|X_t(\tau)\|^2\ d\tau \leq \frac{M}{4}\int_t^{t+h} |\varepsilon(\tau)|\ d\tau.$$

Since ε is in L^1 the right-hand tends to 0 as $t \to \infty$ uniformly in $h \geq 0$. ■

If $A(t) \equiv A$ we have the following:

Corollary 8.7 *Let A be a maximal monotone operator in Hilbert space with $\mathcal{S} \neq \emptyset$. Take $\varepsilon \in L^1$ and consider a solution v of the differential inclusion $v'(t) + Av(t) + \varepsilon(t)v(t) \ni 0$. We have*

1. *v converges weakly in average to a point in $\mathcal{S}$.*
2. *If A is odd, v converges strongly in average to a point in $\mathcal{S}$.*
3. *If A is demipositive, v converges to weakly to a point in $\mathcal{S}$.*
4. *v converges strongly to a point in $\mathcal{S}$ if*
   (a) *A is strongly monotone;*
   (b) *$\mathcal{S}$ has nonempty interior; or*
   (c) *$A = \partial f$, whenever $f \in \Gamma_0(H)$ is even.*

Remark 8.8 *Part (3) in the previous corollary is a consequence of Theorem 2 in [52] in the case where $A = \partial f$.*

8.2.2 Non-autonomous Lipschitz dynamics and diagonal prox

Let $A(t)$ be a family of (m-)accretive maps on a Banach space X satisfying

$$\|A(t)x - A(s)y\| \leq \alpha\|x-y\| + \Theta(x,y)|t-s| \quad \text{and} \tag{8.7}$$
$$\|A(t)x\| \leq \Phi(x), \tag{8.8}$$

where Θ and Φ are locally bounded functions.

Example 23 *Let $A, B : X \longrightarrow X$ and $\varepsilon : [0,\infty) \longrightarrow (0,\infty)$ be Lipschitz functions with constants a, b and η respectively and suppose ε is bounded by a constant E. Then $A(t) = A + \varepsilon(t)B$ satisfies* (8.7).

Example 24 *Let $T : X \longrightarrow X$ be nonexpansive and ε as in the preceding example with $E = 1$. Then the family $\{A(t)\}$ defined by $A(t)z = z - T((1-\varepsilon(t))z)$ satisfies* (8.7).

Given $x \in X$ and $s \in \mathbf{R}$, the equation

$$\begin{cases} u'(t) &= -A(t)u(t) \\ u(s) &= x \end{cases}$$

has a unique solution by virtue of the Cauchy-Lipschitz-Picard Theorem, which we denote by $U(t,s)x$. The two-parameter family $\{U(t,s)\}$ is a *CES* and we shall assume there is a locally bounded function $\Psi : X \longrightarrow \mathbf{R}_+$ such that

$$\|U(t,t_0)z - U(s,s_0)z\| \leq \Psi(z)\left[|t-s| + |t_0 - s_0|\right]. \tag{8.9}$$

For any $\lambda > 0$ and $t \geq 0$ the resolvent $J_\lambda(t)$ and the Yosida approximation $A_\lambda(t)$ of $A(t)$ are defined by

$$J_\lambda(t) = (I + \lambda A(t))^{-1} \qquad \text{and} \qquad A_\lambda(t) = \tfrac{1}{\lambda}(I - J_\lambda(t)),$$

respectively. From an initial point $x_0 \in X$ and a sequence $\{\lambda_n\}$ in $(0, \Lambda]$ we define a *diagonal proximal sequence* $\{x_n\}$ by $x_n = J_{\lambda_n}(\sigma_n)x_{n-1}$, where $\sigma_n = \sum_{k=1}^n \lambda_k$. In order to simplify notation we shall write $P_m^n = \prod_{k=m+1}^n J_{\lambda_k}(\sigma_k)$ so that $x_n = P_m^n x_m$ if $m < n$. The following lemma was motivated by Lemma 5.2 in [56]. It can be easily proved by induction:

Lemma 8.9 *For all $n, p \geq 1$ all $x \in X$ we have*

i) $\|U(\sigma_{n+p}, \sigma_n)x - P_n^{n+p}x\| \leq \sum_{m=n+1}^{n+p} \|U(\sigma_m, \sigma_n)x - P_{m-1}^m U(\sigma_{m-1}, \sigma_n)x\|$*; and*

ii) $\|U(\sigma_{n+p}, \sigma_n)x - P_n^{n+p}x\| \leq \sum_{m=n+1}^{n+p} \|P_n^m x - U(\sigma_m, \sigma_{m-1})P_n^{m-1}x\|.$

Now we provide an estimate for the distance between the two processes after one iteration:

Lemma 8.10 *There is a locally bounded function Ξ such that for all $k \in \mathbf{N}$ and $z \in X$ we have*

$$\|U(\sigma_k, \sigma_{k-1})z - J_{\lambda_k}(\sigma_k)z\| \leq \Xi(z)\lambda_k^2.$$

Proof. First observe that

$$\begin{aligned}
U(\sigma_k, \sigma_{k-1})z - J_{\lambda_k}(\sigma_k)z &= [z - J_{\lambda_k}(\sigma_k)z] + [U(\sigma_k, \sigma_{k-1})z - z] \\
&= \lambda_k A_{\lambda_k}(\sigma_k)z - \int_{\sigma_{k-1}}^{\sigma_k} A(\tau)U(\tau, \sigma_{k-1})z \, d\tau \\
&= \int_{\sigma_{k-1}}^{\sigma_k} A(\sigma_k)J_{\lambda_k}(\sigma_k)z \, d\tau - \int_{\sigma_{k-1}}^{\sigma_k} A(\tau)U(\tau, \sigma_{k-1})z \, d\tau.
\end{aligned}$$

From the preceding argument and (8.7) we get

$$\begin{aligned}
\|U(\sigma_k, \sigma_{k-1})z - J_{\lambda_k}(\sigma_k)z\| &\leq \int_{\sigma_{k-1}}^{\sigma_k} \|A(\tau)U(\tau, \sigma_{k-1})z - A(\sigma_k)J_{\lambda_k}(\sigma_k)z\| \, d\tau \\
&\leq \alpha \int_{\sigma_{k-1}}^{\sigma_k} \|J_{\lambda_k}(\sigma_k)z - z\| \, d\tau \\
&\quad + \alpha \int_{\sigma_{k-1}}^{\sigma_k} \|z - U(\tau, \sigma_{k-1})z\| \, d\tau \\
&\quad + \int_{\sigma_{k-1}}^{\sigma_k} \Theta(J_{\lambda_k}(\sigma_k)z, U(\tau, \sigma_{k-1})z)(\sigma_k - \tau) \, d\tau.
\end{aligned}$$

The first term can be bounded using the Yosida approximation and (8.7):

$$\|J_{\lambda_k}(\sigma_k)z - z\| = \lambda_k \|A_{\lambda_k}(\sigma_k)z\| \leq \lambda_k \|A(\sigma_k)z\| \leq \lambda_k \Phi(z).$$

For the remaining terms, use (8.7) and (8.9) in a similar fashion and then integrate. This procedure yields

$$\|U(\sigma_k,\sigma_{k-1})z - J_{\lambda_k}(\sigma_k)z\| \le \Xi(z)\lambda_k^2$$

for the locally bounded function

$$\Xi(z) = \alpha(\Phi(z)+\Psi(z)) + \tfrac{1}{2}\Theta(\|z\| + \Lambda\Phi(z), \|z\| + \Lambda\Psi(z)).$$

■

As before, let $\nu(t) = \max\{\, k \in \mathbf{N} \mid \sigma_k \le t \,\}$ and define a *CES* by $V(t,s) = P_{\nu(s)}^{\nu(t)}$.

Proposition 8.11 *Take $s \in \mathbf{R}$ and $x \in X$. If $\sum_{k=1}^{\infty} \lambda_k^2 < \infty$ then*

1. *If the function $t \mapsto V(t,s)x$ is bounded then it is an almost-orbit of U; and*
2. *If the function $t \mapsto U(t,s)x$ is bounded then it is an almost-orbit of V.*

Proof. Without loss of generality we assume $s = 0$. Take $n, p \in \mathbf{N}$ such that $n = \nu(t)$ and $n + p = \nu(t+h)$. We have

$$\begin{aligned}
\|V(t+h,0)x - U(t+h,t)V(t,0)x\| &\le \|P_n^{n+p}P_0^n x - U(\sigma_{n+p},\sigma_n)P_0^n x\| \\
&\quad + \|U(\sigma_{n+p},\sigma_n)P_0^n x - U(t+h,t)P_0^n x\| \\
&\le \sum_{m=n+1}^{n+p} \|U(\sigma_m,\sigma_{m-1})P_0^{m-1}x - P_{m-1}^m P_0^{m-1}x\| \\
&\quad + \Psi(P_0^n x)[\lambda_{n+p} + \lambda_n] \\
&\le \sum_{m=n+1}^{n+p} \Xi(P_0^{m-1}x)\lambda_m^2 + 2\Psi(P_0^n x)\max_{k\ge n}\{\lambda_k\} \\
&\le K_1\left[\max_{m\ge n}\{\lambda_m\} + \sum_{m=n+1}^{\infty}\lambda_m^2\right].
\end{aligned}$$

The right-hand side tends to 0 uniformly in p as $n \to \infty$.

For the second part, we have

$$\begin{aligned}
\|U(t+h,0)x - V(t+h,t)U(t,0)x\| &\le \|U(t+h,t)U(t,0)x - U(\sigma_{n+p},\sigma_n)U(\sigma_n,0)x\| \\
&\quad + \|U(\sigma_{n+p},\sigma_n)U(\sigma_n,0)x - V(t+h,t)U(\sigma_n,0)x\| \\
&\quad + \|V(t+h,t)U(\sigma_n,0)x - V(t+h,t)U(t,0)x\| \\
&\le \sum_{m=n+1}^{n+p} \Xi(U(\sigma_{m-1},0)x)\lambda_m^2 \\
&\quad + 2\Psi(U(\sigma_n,0)x)\max_{k\ge n}\{\lambda_k\} \\
&\quad + \Psi(x)\lambda_n \\
&\le K_2\left[\lambda_n + \max_{m\ge n}\{\lambda_m\} + \sum_{m=n+1}^{\infty}\lambda_m^2\right]
\end{aligned}$$

and the result follows as in the first part. ■

8.2.3 Comparing the trajectories of two differential inclusions

In this section we prove an equivalence result for trajectories defined by the differential inclusions governed by two families, $\{A(t)\}_{t\geq t_0}$ and $\{\widehat{A}(t)\}_{t\geq t_0}$, of (m)-accretive operators on X defined on a common t-independent domain.

Let U and $\widehat{U}$ be the evolution systems defined by A and $\widehat{A}$ as in (6.9a) and (6.9b), which we assume to exist. They are DS-limit solutions of the differential inclusion (6.1) with A and $\widehat{A}$, respectively. Notice also that U and $\widehat{U}$ are *CES*. Under (6.11), Theorem 6.8 gives

$$\|U(t,s)x - \widehat{U}(t,s)x\| \leq \sqrt{2}\left|\int_s^t \|A(\tau)x\| - \|\widehat{A}(\tau)x\|\ d\tau\right| + \int_s^t \Theta(\tau,\tau)\ d\tau \tag{8.10}$$

for all $t \geq t_0$.

If we assume that for each $r > 0$ there is a function $F_r \in L^1(t_0,\infty;\mathbf{R})$ such that for every $x \in B(0,r)$ one has

$$\Theta(t,t) + \Big|\|A(t)x\| - \|\widehat{A}(t)x\|\Big| \leq F_r(t) \tag{8.11}$$

almost everywhere on $[t_0,\infty)$, we get the following:

Lemma 8.12 *Under* (6.11) *and* (8.11)*, every bounded orbit of U is an almost-orbit of $\widehat{U}$.*

Proof. Let $U(\cdot,t_0)x_0$ be bounded in norm by $r > 0$. According to inequalities (8.10) and (8.11) we have

$$\|U(t+h,t_0)x_0 - \widehat{U}(t+h,t)U(t,t_0)x_0\| \leq \sqrt{2}\int_t^{t+h} F_r(\xi)\ d\xi$$

and so, $U(\cdot,t_0)x_0$ is an almost-orbit of $\widehat{U}$. ■

Example 25 *Hypotheses* (6.11) *and* (8.11) *hold, for instance, if $A(t)x = Ax + f(t,x)$, $\widehat{A}(t)x = Ax + \widehat{f}(t,x)$ whenever $\|f(t,x) - \widehat{f}(s,\widehat{x})\|$ can be bounded in the ball $B(0,r)$ by some function $\Phi_r(t,s)$ such that $\int_0^\infty \Phi_r(\tau,\tau)d\tau < \infty$ for each r. In this case, one can also derive the following continuity result:*

$$\|U - \widehat{U}\|_\infty \leq \|x - \widehat{x}\| + 2\sqrt{2}\Phi_r(\tau,\tau)d\tau < \infty, \tag{8.12}$$

when U and $\widehat{U}$ start at x and $\widehat{x}$, respectively. This is similar to the bound given in [50, Theorem 4], where A is the gradient of a $\mathcal{C}^1$ convex function having minimizers. However, inequality (8.12) *holds for any maximal monotone operator.*

Theorem 8.13 *Assume* (6.11) *and* (8.11) *hold. If $U(t,s)x$ converges weakly (strongly) as $t \to \infty$ for all s and x, then $\widehat{U}(t,s)x$ does.*

Proof. This is an immediate consequence of Theorem 7.9 and Lemma 8.12. ■

Remark 8.14 *Theorems 7.5 and 6.8 may be used to establish discrete-discrete and discrete-continuous versions of Lemma 8.12 but this will not be done here.*

8.2.4 Second order: The nonlinear oscillator with damping

In [2], the author studies the problem

$$u''(t) + \gamma u'(t) + \nabla\Phi(u(t)) = 0, \tag{8.13}$$

in Hilbert space H, where Φ is a $\mathcal{C}^1$ convex function which is bounded from below. He proves the following result:

Theorem 8.15 *If $u \in \mathcal{C}^2$ satisfies (8.13), then $u' \in L^2$, $\lim_{t\to\infty} u'(t) = 0$ and $\lim_{t\to\infty} \Phi(u(t)) = \inf \Phi = \Phi^*$. Moreover, if Argmin($\Phi$) $\neq \emptyset$, then $u(t)$ converges weakly to a minimizer of Φ as $t \to \infty$. Strong convergence occurs in the following cases:*

1. *$\nabla\Phi$ is strongly monotone,*
2. *Φ is even,*
3. *Argmin(Φ) has nonempty interior.*

Later, in [7] the authors study a similar differential equation:

$$u''(t) + \gamma u'(t) + \nabla\Phi(u(t)) + \varepsilon(t)u(t) = 0. \tag{8.14}$$

The control ε is assumed to be positive, $\mathcal{C}^2$ and nonincreasing (weaker hypotheses are considered in the cited article and can be treated by our method too). The authors establish convergence results which depend on whether ε is in L^1 or not. For $\varepsilon \notin L^1$, the trajectory always converges strongly to the least norm element of Argmin(Φ).

For $\varepsilon \in L^1$, the trajectory converges weakly to a minimizer that may depend on the initial conditions. The authors also obtain strong convergence in the first two cases mentioned in Theorem 8.15 above.[4] We see that the evolution systems defined by the solutions to (8.13) and (8.14) have similar asymptotic properties when $\varepsilon \in L^1$. In fact, as we shall see below, they are equivalent. This immediately provides an alternative proof for some of the results obtained in [7], namely Theorem 2.4 and Corollaries 2.2 and 2.3. This procedure also generalizes the strong convergence result when Argmin(Φ) has nonempty interior. It is easy to prove that, since Φ has minimizers, all the solutions of (8.13) and (8.14) are bounded.

For $t \geq t_0 \geq 0$ and $u_0, v_0 \in H$ we define $\mathcal{U}(t,t_0)[u_0,v_0] = [u(t),u'(t)]$ as the solution of (8.13) with initial condition $u(t_0) = u_0$ and $u'(t_0) = v_0$. In a similar way define $\hat{\mathcal{U}}(t,t_0)[\hat{u}_0,\hat{v}_0]$ for (8.14).

Proposition 8.16 *Take $\varepsilon \in L^1$ and let $\mathcal{S} \neq \emptyset$. Every orbit of $\hat{\mathcal{U}}$ is an almost orbit of $\mathcal{U}$.*

Proof. Let $[y(t),y'(t)] = \hat{\mathcal{U}}(t,t_0)[\hat{u}_0,\hat{v}_0]$ be an orbit of $\hat{\mathcal{U}}$ and define $[X_s(t),X_s'(t)] = \mathcal{U}(t,s)[y(s),y'(s)]$. We must prove that

$$\alpha(t,h) = \|X_t(t+h) - y(t+h)\| \quad \text{and} \quad \beta(t,h) = \|X_t'(t+h) - y'(t+h)\|$$

both tend to zero as $t \to \infty$ uniformly in $h \geq 0$.

Fix t and set $E(h) = \frac{1}{2}\|X_t'(t+h)\|^2 + \Phi(X_t(t+h))$ and $\hat{E}(h) = \frac{1}{2}\|y'(t+h)\|^2 + \Phi(y(t+h)) + \frac{1}{2}\varepsilon(t+h)\|y(t+h)\|^2$. Both functions are decreasing, so we have

$$\|X_t'(t+h)\|^2 \leq \|y'(t)\|^2 + 2(\Phi(y(t)) - \Phi^*)$$

[4] They give a slightly more general definition of strong monotonicity, but the proof in [2] may be easily adapted accordingly.

and

$$\|y'(t+h)\|^2 \leq \|y'(t)\|^2 + 2(\Phi(y(t)) - \Phi^*) + \varepsilon(t)\|y(t)\|^2.$$

It is clear that $\beta(t,h) \to 0$ as $t \to \infty$ uniformly in $h \geq 0$.

Now define $\psi(h) = \frac{1}{2}\alpha(t,h)^2$. Observing that $\langle \zeta, \zeta - \xi \rangle \geq -\frac{1}{4}\|\xi\|^2$ for all $\zeta, \xi \in H$ we get

$$\psi''(h) + \gamma\psi'(h) \leq \beta(t,h)^2 + \frac{1}{4}\varepsilon(t+h)\|X_t(t+h)\|^2. \tag{8.15}$$

The factor $\|X_t(t+h)\|^2$ can be bounded uniformly in t and h by a constant C. Observe that $|\psi'(h)| \leq 2\sqrt{\beta(t,h)\psi(h)}$. Integrating from 0 to h we get

$$\gamma\psi(h) \leq \beta(t,h)\sqrt{2\psi(h)} + \int_0^h \beta(t,\eta)^2 \, d\eta + \frac{C}{4}\int_0^h \varepsilon(t+\eta) \, d\eta.$$

This is a second-order inequality in $\sqrt{\psi(h)}$ that can be reduced to

$$2\gamma\sqrt{\psi(h)} \leq \sqrt{2}\beta(t,h) + \sqrt{2\beta(t,h)^2 + 4\gamma\int_0^h \beta(t,\eta)^2 \, d\eta + \gamma C\int_0^h \varepsilon(t+\eta) \, d\eta}.$$

We get

$$\sqrt{2}\gamma\alpha(t,h) \leq 2\sqrt{2}\beta(t,h) + \sqrt{4\gamma\int_0^h \beta(t,\eta)^2 \, d\eta} + \sqrt{\gamma C\int_t^\infty \varepsilon(\eta) \, d\eta}.$$

The first and last terms on the right-hand side of the previous inequality tend to 0 as $t \to \infty$ uniformly in $h \geq 0$. It remains to show that the second term has the same behavior. Recall that $y' \in L^2$. By the definition of β it suffices to show that $\int_t^{t+h} \|X_t'(\eta)\|^2 \, d\eta$ tends to zero as $t \to \infty$ uniformly in $h \geq 0$. To see this, recall first that $E(h) = \frac{1}{2}\|X_t'(t+h)\|^2 + \Phi(X_t(t+h))$, so that $E'(h) = -\gamma\|X_t'(t+h)\|^2$. Integrating we get

$$\gamma\int_t^{t+h} \|X_t(\eta)\|^2 \, d\eta = E(t) - E(t+h) \leq \frac{1}{2}\|y'(t)\|^2 + (\Phi(y(t)) - \Phi^*),$$

which tends to 0 as $t \to \infty$ uniformly in $h \geq 0$. ∎

Bibliography

[1] Alart P, Lemaire B, *Penalization in non-classical convex programming via variational convergence.* Math. Programming, **51** (1991), 307-331.

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

[3] Alvarez F, Carrasco M, Pichard K, *Convergence of a hybrid projection-proximal point algorithm coupled with approximation methods in convex optimization*, Math. Oper. Res., **30** (2005), 966-984.

[4] Alvarez F, Peypouquet J, *Asymptotic equivalence and Kobayashi-type estimates for nonautonomous monotone operators in Banach spaces*, submitted for publication at DCDS.

[5] Alvarez F, Peypouquet J, *Asymptotic almost-equivalence of abstract evolution systems*, submitted for publication at Proc. Amer. Math. Soc.

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

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

[8] Attouch H, Damlamian A, *Strong solutions for parabolic variational inequalities*, J. Nonlinear Anal., **2** (1978), 329-353.

[9] Aubin JP, Cellina A, "Differential inclusions. Set-valued maps and viability theory". Springer-Verlag, Berlin, 1984.

[10] Auslender A, *Numerical methods for nondifferentiable convex optimization*, Math. Program. Studies, **30** (1987), 102-127.

[11] Auslender A, Crouzeix JP, Fedit P, *Penalty-proximal methods in convex programming*, J. Optim. Theory Appl., **55** (1987), 1-21.

[12] Azuma R, *Approximation theorem for evolution operators*, Studia Math., **154** (2003), 195-206.

[13] Bahraoui MA, Lemaire B, *Diagonal bundle methods for convex minimization*, unpublished.

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

[15] Baillon JB, *Quelques propriétés de convergence asymptotique pour les semi-groupes de contractions impaires*, CRAS, **238** (1976), A75-A78.

[16] Baillon JB, *Quelques propriétés de convergence asymptotique pour les contractions impaires*, CRAS, **238** (1976), A587-A590.

[17] 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.

[18] Baillon JB, "Quelques aspects de la th'eorie des points fixes dans les espaces de Banach I, II". Séminaire d'analyse fonctionnelle, École Polytechnique (1978-1979).

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

[20] Baillon JB, Bruck RE, Reich S, *On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces*, Houston J. Math., **4** (1978), 1-9.

[21] Baillon JB, Cominetti R, *A convergence result for non-autonomous subgradient evolution equations and its application to the steepest descent exponential penalty trajectory in linear programming*, Journal of Functional Analysis, **187** (2001), 263-273.

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

[23] Benaïm M, *A dynamical sytem approach to stochastic approximation*, SIAM J. Control Optim., **34** (1996), 437-472.

[24] Benaïm M, El Karoui N, "Promenade aléatoire". Éditions de l'École Polytechnique, 2004.

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

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

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

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

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

[30] Brézis H, "Asymptotic behavior of some evolution systems". Nonlinear Evolution Equations, ed. by M. G. Crandall, Academic Press, 1978, 141-154.

[31] Brézis H, Browder FE, *Nonlinear ergodic theorems*, Bull. Amer. Math. Soc., **82** (1976), 959-961.

[32] Brézis H, Browder FE, *Remarks on nonlinear ergodic theory*, Adv. Math., **25** (1977), 165-177.

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

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

[35] Bruck RE, *An iterative solution of a variational inequality for certain monotone operators in Hilbert space*, Bull. Amer. Math. Soc., **81** (1975), 890-892.

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

[37] 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.

[38] 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.

[39] Bruck RE, *A simple proof of the mean ergodic theorem for nonlinear contractions in Banach spaces*, Israel J. Math., **32** (1979), 107-116.

[40] Bruck RE, Passty G, *Almost convergence of the infinite product of resolvents in Banach spaces*, Nonlinear Anal., **3** (1979), 279-282.

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

[42] Cominetti R, *Coupling the proximal point algorithm with approximation methods*, J. Optim. Theory Appl., **95** (1997), 581-600.

[43] Cominetti R, Alemany O, *Steepest descent evolution equations: asymptotic behavior of solutions and rate of convergence*, Trans. Amer. Math. Soc., **351** (1999), 4847- 4860.

[44] Cominetti R, Courdurier M, *Coupling general penalty schemes for convex programming with the steepest descent method and the proximal point algorithm*, SIAM J. Optim, **13** (2002), 745-765.

[45] Cominetti R, Peypouquet J, Sorin S, *Strong asymptotic convergence of evolution equations governed by maximal monotone operators with Tikhonov regularization*, work in progress.

[46] Crandall MG, "An introduction tu evolution governed by accretive operators", in Dynamical systems, **1**, Academic Press, New York, 1976.

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

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

[49] Crandall MG, Pazy A, *Nonlinear evolution equations in Banach spaces*, Israel J. Math, **11** (1972), 57-94.

[50] Czarnecki MO, *Asymptotic control and stabilization of nonlinear oscillators with non-isolated equilibria, a note: from L^1 to non L^1*, J. Diff. Equations, **217** (2005), 501-511.

[51] Ferris M, *Finite termination of the proximal point algorithm*, Math. Programming, **50** (1991), 359-366.

[52] Furuya H, Miyashiba K, Kenmochi N, *Asymptotic behavior of solutions to a class of nonlinear evolution equations*, J. Diff. Equations, **62** (1986), 73-94.

[53] García-Falset J, Reich S, Zeroes of accretive operators and the asymptotic behavior of nonlinear semigroups. *Houston J. Math.*, **32** (2006), 1197-1225.

[54] Georgescu P, Oharu S, *Generation and characterization of locally Lipschitzian semigroups associated with semilinear evolution equations*, Hiroshima Math. J., **31** (2001), 133-169.

[55] Gol'shtein EG, *A method for the modification of monotone mappings*, Ekon. i. Matem. Metody (Matecon), **11** (1975), 1144 - 1159.

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

[57] Güler O, *New proximal point algorithms for convex minimization*, SIAM J. Optim., **2** (1992), 649-664.

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

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

[60] Hirsch F, "Familles résolvantes, générateurs, cogénérateurs, potentiels". Thèse de Doctorat d'État, Orsay, 1971.

[61] Kaplan A, *On a convex programming method with internal regularization*, Soviet Mathematics Doklady, **19** (1978), 795-799.

[62] Kaplan A, Tichatschke R, *Proximal methods in wiew of interior-point strategies*, J. Optim. Theory Appl., **98** (1998), 399-429.

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

[64] Kenmochi N, Takahashi T, *Nonautonomous differential equations in Banach spaces*, Nonlinear Anal., **4** (1980), 1109-1121.

[65] Knopp K, "Theory and application of infinite series". Blackie & Son, Glasgow, 1951.

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

[67] Kobayasi K, Kobayashi Y, Oharu S, *Nonlinear evolution operators in Banach spaces*, Osaka J. Math, **21** (1984), 281-310.

[68] Kobayasi K, Miyadera I, *On the strong convergence of the Cesàro means of contractions in Banach spaces*, Proc. Japan. Acad., **56** (1980), 245-249.

[69] Kohlberg E, Neyman A, *Asymptotic behavior of nonexpansive mappings in normed linear spaces*, Israel J. Math., **38** (1981), 269-275.

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

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

[72] Lau A, Nishiura K, Takahashi W, Convergence of almost-orbits of nonexpansive semigroups in Banach spaces. *Proc. Amer. Math. Soc.*, **135** (2007), 3143-3150.

[73] Lemaire B, "About the convergence of the proximal method", Advance in optimization, Proceedings Lambrecht 1991, Lecture Notes in Economics and Mathematical Systems 382, Springer-Verlag, 1992.

[74] Lemaire B, *Bounded diagonally stationary sequences in convex optimization*, J. Convex Anal., **1** (1994), 75-86.

[75] Lemaire B, "On the convergence of some iterative methods for convex minimization. Recent developments in optimization" (Dijon, 1994), 252 - 268, Lecture Notes in Econom. and Math. Systems, **429**, Springer, Berlin, 1995.

[76] Lemaire B, *An asymptotical variational principle associated with the steepest descent method for a convex function*, Journal of Convex Analysis, **3** (1996), 63-70.

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

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

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

[80] Mann WR, *Mean value methods in iteration*, Proc. Amer. Math. Soc., **4** (1953), 506-510.

[81] Martinet B, *Régularisation d'inéquations variationnelles par approximations successives*, Rev. fr. autom. inform. rech. opér., **4** (1970), 154-158.

[82] Minty G, *Monotone (nonlinear) operators in Hilbert space*, Duke Math. J., **29** (1962), 341-346.

[83] Miyadera I, *Nonlinear ergodic theorems for semigroups of non-Lipschitzian mappings in Banach spaces*, Nonlinear Anal., **50** (2002), 27-39.

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

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

[86] Moudafi A, *Coupling proximal algorithm and Tikhonov method*, Nonlinear Times and Digest, **1** (1994), 203-210.

[87] 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.

[88] Oharu S, Tebbs D, *A time-dependent product formula and its application to an HIV infection model*, Adv. Math. Sci. Appl., **14** (2004), 251-265.

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

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

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

[92] Peypouquet J, *Asymptotic convergence to the optimal value of diagonal proximal iterations in convex minimization*, submitted for publication at J. Convex Anal.

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

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

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

[96] Rockafellar RT, *Augmented lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res., **1** (1976), 97-116.

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

[98] Rouhani BD, Asymptotic behavior of uniformly asymptotically almost nonexpansive curves in a Hilbert space. *Nonlinear Anal.*, **58** (2004), 143-157.

[99] Rouhani BD, Kim JK, Ergodic theorems for almost-orbits of semigroups of Non-Lipschitzian mappings in a Hilbert space. *J. Nonlinear Convex Anal.*, **4** (2003). 175-183.

[100] Royden HL, "Real Analysis". The Macmillan Company, Ney York, 1968.

[101] Solodov MV, Svaiter BF, *A hybrid projection-proximal point algorithm*, J. Convex Anal., **6** (1999), 59-70.

[102] Suzuki T, Some remarks on the set of common fixed points of one-parameter semigroups of nonexpansive mappings in Banach spaces with the Opial property. *Nonlinear Anal.*, **58** (2004), 441-458.

[103] Suzuki T, Takahashi W, Strong convergence of Mann's type sequences for one-parameter nonexpansive semigroups in general Banach spaces. *J. Nonlinear Convex Anal.*, **5** (2004), 209-216.

[104] Tebbs D, *On the product formula approach to a class of quasilinear evolution systems*, Tokyo J. Math., **26** (2003), 423-445.

[105] Vigeral G, Personal communication (2006).

[106] Xu HK, Strong asymptotic behavior of almost-orbits of nonlinear semigroups. *Nonlinear Anal.*, **46** (2001), 135-151.