

Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators

Hedy Attouch¹  · Juan Peypouquet^{2,3}

Received: 3 May 2017 / Accepted: 16 February 2018 / Published online: 7 March 2018
© Springer-Verlag GmbH Germany, part of Springer Nature and Mathematical Optimization Society 2018

Abstract We study the behavior of the trajectories of a second-order differential equation with vanishing damping, governed by the Yosida regularization of a maximally monotone operator with time-varying index, along with a new *Regularized Inertial Proximal Algorithm* obtained by means of a convenient finite-difference discretization. These systems are the counterpart to accelerated forward–backward algorithms in the context of maximally monotone operators. A proper tuning of the parameters allows us to prove the weak convergence of the trajectories to zeroes of the operator. Moreover, it is possible to estimate the rate at which the speed and acceleration vanish. We also study the effect of perturbations or computational errors that leave the convergence properties unchanged. We also analyze a growth condition under which strong convergence can be guaranteed. A simple example shows the criticality of the assumptions on the Yosida approximation parameter, and allows us to illustrate the behavior of these systems compared with some of their close relatives.

H. Attouch: Effort sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under Grant Number F49550-1 5-1-0500. J. Peypouquet: supported by Fondecyt Grant 1140829, Millenium Nucleus ICM/FIC RC130003 and Basal Project CMM Universidad de Chile.

✉ Hedy Attouch
hedy.attouch@univ-montp2.fr

Juan Peypouquet
juan.peypouquet@usm.cl

¹ Institut Montpelliérain Alexander Grothendieck, UMR 5149 CNRS, Université Montpellier, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France

² Departamento de Matemática, Universidad Técnica Federico Santa María, Av España 1680, Valparaíso, Chile

³ Departamento de Ingeniería Matemática & Centro de Modelamiento Matemático (CNRS UMI2807), FCFM, Universidad de Chile, Beauchef 851, Santiago, Chile

Keywords Asymptotic stabilization · Large step proximal method · Damped inertial dynamics · Lyapunov analysis · Maximally monotone operators · Time-dependent viscosity · Vanishing viscosity · Yosida regularization

Mathematics Subject Classification 37N40 · 46N10 · 49M30 · 65K05 · 65K10 · 90B50 · 90C25

1 Introduction

Let \mathcal{H} be a real Hilbert space endowed with the scalar product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. Given a maximally monotone operator $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$, we study the asymptotic behavior, as time goes to $+\infty$, of the trajectories of the second-order differential equation

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + A_{\lambda(t)}(x(t)) = 0, \quad t > t_0 > 0, \quad (1)$$

where

$$A_\lambda = \frac{1}{\lambda} \left(I - (I + \lambda A)^{-1} \right)$$

stands for the Yosida regularization of A of index $\lambda > 0$ (see “Appendix A.1” for its main properties), along with a new *Regularized Inertial Proximal Algorithm* obtained by means of a convenient finite-difference discretization of (1). The design of rapidly convergent dynamics and algorithms to solve monotone inclusions of the form

$$0 \in Ax \quad (2)$$

is a difficult problem of fundamental importance in optimization, equilibrium theory, economics and game theory, partial differential equations, statistics, among other subjects. We shall come back to this point shortly. The dynamical systems studied here are closely related to Nesterov’s acceleration scheme, whose rate of convergence for the function values is worst-case optimal. We shall see that, properly tuned, these systems converge to solutions of (2), and do so in a robust manner. We hope to open a broad avenue for further studies concerning related stochastic approximation and optimization methods. Let us begin by putting this in some more context. In all that follows, the set of solutions to (2) will be denoted by S .

1.1 From the heavy ball to fast optimization

Let $\Phi : \mathcal{H} \rightarrow \mathbb{R}$ be a continuously differentiable convex function. The *heavy ball with friction* system

$$(HBF) \quad \ddot{x}(t) + \gamma \dot{x}(t) + \nabla \Phi(x(t)) = 0 \quad (3)$$

was first introduced, from an optimization perspective, by Polyak [31]. The convergence of the trajectories of (HBF) in the case of a convex function Φ has been obtained

by Álvarez in [1]. In recent years, several studies have been devoted to the study of the Inertial Gradient System $(\text{IGS})_\gamma$, with a time-dependent damping coefficient $\gamma(\cdot)$

$$(\text{IGS})_\gamma \quad \ddot{x}(t) + \gamma(t)\dot{x}(t) + \nabla\Phi(x(t)) = 0. \quad (4)$$

A particularly interesting situation concerns the case $\gamma(t) \rightarrow 0$ of a vanishing damping coefficient. Indeed, as pointed out by Su et al. in [37], the $(\text{IGS})_\gamma$ system with $\gamma(t) = \frac{\alpha}{t}$, namely:

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla\Phi(x(t)) = 0, \quad (5)$$

can be seen as a continuous version of the fast gradient method of Nesterov (see [26, 27]), and its widely used successors, such as the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) of Beck and Teboulle [15]. When $\alpha \geq 3$, the rate of convergence of these methods is $\Phi(x_k) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{k^2}\right)$, where k is the number of iterations. Convergence of the trajectories generated by (5), and of the sequences generated by Nesterov's method, has been an elusive question for decades. However, when considering (5) with $\alpha > 3$, it was shown by Attouch et al. [6] and May [24], that each trajectory converges weakly to an optimal solution, with the improved rate of convergence $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o\left(\frac{1}{t^2}\right)$. Corresponding results for the algorithmic case have been obtained by Chambolle and Dossal [20] and Attouch and Peypouquet [8]. Independently, and mainly motivated by applications to partial differential equations and control problems, Jendoubi and May [22] and Attouch and Cabot [3, 4] consider more general time-dependent damping coefficient $\gamma(\cdot)$. The latter includes the corresponding forward–backward algorithms, and unifies previous results. In the case of a convex lower semicontinuous proper function $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup +\{\infty\}$, the $(\text{IGS})_\gamma$ dynamic (4) is not well-posed, see [5]. A natural idea is to replace Φ by its Moreau envelope Φ_λ , and consider

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla\Phi_{\lambda(t)}(x(t)) = 0 \quad (6)$$

(see [14, 29] for details, including the fact that $\nabla\Phi_\lambda = (\partial\Phi)_\lambda$). Fast minimization and convergence properties for the trajectories of (6) and their related algorithms have been recently obtained by Attouch et al. [9]. This also furnishes a viable path towards its extension to the maximally monotone operator setting. It is both useful and natural to consider a time-dependent regularization parameter, as we shall explain now.

1.2 Inertial dynamics and cocoercive operators

In analogy with (3), Álvarez and Attouch [2] and Attouch and Maingé [7] studied the equation

$$\ddot{x}(t) + \gamma\dot{x}(t) + A(x(t)) = 0, \quad (7)$$

where A is a cocoercive operator. Cocoercivity plays an important role, not only to ensure the existence of solutions, but also in analyzing their long-term behavior. They discovered that it was possible to prove weak convergence to a solution of (2) if the

cocoercivity parameter λ and the damping coefficient γ satisfy $\lambda\gamma^2 > 1$. Taking into account that for $\lambda > 0$, the operator A_λ is λ -cocoercive and that $A_\lambda^{-1}(0) = A^{-1}(0)$ (see “Appendix A.1”), we immediately deduce that, under the condition $\lambda\gamma^2 > 1$, each trajectory of

$$\ddot{x}(t) + \gamma \dot{x}(t) + A_\lambda(x(t)) = 0$$

converges weakly to a zero of A . In the quest for a faster convergence, our analysis of Eq. (1), led us to introduce a time-dependent regularizing parameter $\lambda(\cdot)$ satisfying

$$\lambda(t) \times \frac{\alpha^2}{t^2} > 1$$

for $t \geq t_0$. A similar condition appears in the study of the corresponding algorithms. We mention that a condition of the type $\lambda(t) \sim t^2$ appears to be critical to obtain fast convergence of the values in the case $A = \partial\Phi$, according to the results in [9]. But system (1) is not just an interesting extension of the heavy ball dynamic. It also arises naturally in stochastic variational analysis.

1.3 Connections with stochastic optimization

Let us present some links between our approach and stochastic optimization.

1.3.1 Stochastic approximation algorithms

A close relationship between stochastic approximation algorithms and inertial dynamics with vanishing damping was established by Cabot et al. in [19]. Let us briefly (and skipping the technicalities) comment on this link and see how it naturally extends to our setting.

When A is a sufficiently smooth operator, stochastic approximation algorithms are frequently used to approximate, with a random version of the Euler’s scheme, the behavior of the ordinary differential equation $\dot{x}(t) + A(x(t)) = 0$. If A is a general maximally monotone operator, a natural idea is to apply this method to the regularized equation $\dot{x}(t) + A_\lambda(x(t)) = 0$, which has the same equilibrium points, since A and A_λ have the same set of zeros. Consider first the case where $\lambda > 0$ is a fixed parameter. If we denote by $(X_n)_{n \in \mathbb{N}}$ the random approximants, $(w_n)_{n \geq 1}$ and $(\eta_n)_{n \geq 1}$ two auxiliary stochastic processes, the recursive approximation is written as

$$X_{n+1} = X_n - \epsilon_{n+1} A_\lambda(X_n, w_{n+1}) + \epsilon_{n+1} \eta_{n+1}, \quad (8)$$

where ϵ_n is a sequence of positive real numbers, and η_n is a small residual perturbation. Under appropriate conditions (see [19]), solutions of (8) asymptotically behave like those of the deterministic differential equation $\dot{x}(t) + A_\lambda(x(t)) = 0$. A very common case occurs when $(w_n)_{n \geq 1}$ is a sequence of independent identically distributed variables with distribution μ and

$$A_\lambda(x) = \int A_\lambda(x, \omega) \mu(d\omega).$$

When $A = \partial\Phi$ derives from a potential, this gives a stochastic optimization algorithm (see [32]). When the random variable $A_\lambda(X_n, \cdot)$ has a large variance, the stochastic approximation of $A_\lambda(X_n)$ by $A_\lambda(X_n, w_{n+1})$ can be numerically improved by using the following modified recursive definition:

$$X_{n+1} = X_n - \frac{\epsilon_{n+1}}{\sum_i \epsilon_i} \sum_i \epsilon_i A_\lambda(X_i, w_{i+1}). \quad (9)$$

As proved in [19, Appendix A], the limit ordinary differential equation has the form

$$\ddot{X}(t) + \frac{1}{t+b} \dot{X}(t) + \frac{1}{t+b} A_\lambda(X(t)) = 0, \quad (10)$$

where b is a positive parameter. After the successive changes of the time variable $t \mapsto t - b$, $t \mapsto 2\sqrt{t}$, we finally obtain

$$\ddot{X}(t) + \frac{1}{t} \dot{X}(t) + A_\lambda(X(t)) = 0, \quad (11)$$

which is precisely (1) with $\alpha = 1$ and $\lambda(t) \equiv \lambda$.

This opens a number of possible lines of research: First, it would be interesting to see how the coefficient α will appear in the stochastic case. Next, it appears important to understand the connection between (10) and (11) with a time-dependent coefficient $\lambda(\cdot)$. Combining these two developments, we could be able to extend the stochastic approximation method to a wide range of equilibrium problems, and expect that the fast convergence results to hold, considering that (1) is an accelerated method in the case of convex minimization when $\alpha \geq 3$. In the case of a gradient operator, the above results also suggest a natural link with the epigraphic law of large numbers of Attouch and Wets [11]. The analysis can be enriched using the equivalence between epi-convergence of a sequence of functions and the convergence of the associated evolutionary semigroups.

1.3.2 Robust optimization, regularization

Suppose we are interested in finding a zero of an operator A , which is not exactly known (a situation that is commonly encountered in inverse problems, for instance). If the uncertainty follows a *known* distribution, then the stochastic approximation approach described above may be applicable. Otherwise, if A is known to be sufficiently close to some model \hat{A} (in a sense to be precised below), an alternative is to use robust analysis techniques, interpreting A as a perturbation of \hat{A} . Regularization techniques (like Tikhonov's, where the operator A is replaced by a regularized operator $A + \rho B$) follow a similar logic. We recall the notion of graph-distance bewteen two operators, introduced by Attouch and Wets in [12, 13] (see also [35]). It can be formulated using Yosida approximations as

$$d_{\lambda,\rho}(A, \hat{A}) = \sup_{\|x\| \leq \rho} \|A_\lambda(x) - \hat{A}_\lambda(x)\|,$$

where ρ is a positive parameter. These pseudo-distances are particularly well adapted to our dynamics, which is expressed using Yosida approximations. In the case of minimization problems, they are associated with the notion of epi-distance. The calculus rules developed in [12, 13] make it possible to estimate these distances in the case of operators with an additive or composite structure. In Sect. 4, the convergence of the algorithm is examined in the case of errors that go asymptotically to zero. A more general situation, where approximation and iterative methods are coupled in the presence of noise, is an ongoing research topic. See [25, 36, 38] for recent results in this direction.

1.4 Organization of the paper

In Sect. 2, we study the asymptotic convergence properties of the trajectories of the continuous dynamics (1). We prove that each trajectory converges weakly, as $t \rightarrow +\infty$, to a solution of (2), with $\|\dot{x}(t)\| = O(1/t)$ and $\|\ddot{x}(t)\| = O(1/t^2)$. Then, in Sect. 3, we consider the corresponding proximal-based inertial algorithms and prove parallel convergence results. The effect of external perturbations on both the continuous-time system and the corresponding algorithms is analyzed in Sect. 4, yielding the robustness of both methods. In Sect. 5, we describe how convergence is improved under a quadratic growth condition. Finally, in Sect. 6, we give an example that shows the sharpness of the results and illustrates the behavior of (1) compared to other related systems. Some auxiliary technical results are gathered in an “Appendix”.

2 Convergence results for the continuous-time system

In this section, we shall study the asymptotic properties of the continuous-time system:

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + A_{\lambda(t)}(x(t)) = 0, \quad t > t_0 > 0. \quad (12)$$

Although some of the results presented in this paper are valid when $\lambda(\cdot)$ is locally integrable, we shall assume $\lambda(\cdot)$ to be continuous, for simplicity. The function $(t, x) \mapsto A_{\lambda(t)}(x)$ is continuous in (t, x) , and uniformly Lipschitz-continuous with respect to x . This makes (12) a classical differential equation, whose solutions are unique, and defined for all $t \geq t_0$, for every initial condition $x(t_0) = x_0, \dot{x}(t_0) = v_0$.¹ See Lemma (A.1) for the corresponding result in the case where $\lambda(\cdot)$ is just locally integrable.

The main result of this section is the following:

¹ The idea consisting in regularizing with the help of the Moreau envelopes an inertial dynamic governed by a nonsmooth operator was already used in the modeling of elastic shocks in [5].

Theorem 2.1 Let $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ be a maximally monotone operator such that $S = A^{-1}(0) \neq \emptyset$. Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution of (12), where

$$\alpha > 2 \quad \text{and} \quad \lambda(t) = (1 + \epsilon) \frac{t^2}{\alpha^2} \quad \text{for some } \epsilon > \frac{2}{\alpha - 2}.$$

Then, $x(t)$ converges weakly, as $t \rightarrow +\infty$, to an element of S . Moreover $\lim_{t \rightarrow +\infty} \|\dot{x}(t)\| = \lim_{t \rightarrow +\infty} \|\ddot{x}(t)\| = 0$.

2.1 Anchor

As a fundamental tool we will use the distance to equilibria in order to anchor the trajectory to the solution set $S = A^{-1}(0)$. To this end, given a solution trajectory $x : [t_0, +\infty[\rightarrow \mathcal{H}$ of (12), and a point $z \in \mathcal{H}$, we define $h_z : [t_0, +\infty[\rightarrow \mathbb{R}$ by

$$h_z(t) = \frac{1}{2} \|x(t) - z\|^2. \quad (13)$$

We have the following:

Lemma 2.2 Given $z \in S = A^{-1}(0)$, define h_z by (13). For all $t \geq t_0$, we have

$$\ddot{h}_z(t) + \frac{\alpha}{t} \dot{h}_z(t) + \lambda(t) \|A_{\lambda(t)}(x(t))\|^2 \leq \|\dot{x}(t)\|^2. \quad (14)$$

Proof First observe that

$$\dot{h}_z(t) = \langle x(t) - z, \dot{x}(t) \rangle \quad \text{and} \quad \ddot{h}_z(t) = \langle x(t) - z, \ddot{x}(t) \rangle + \|\dot{x}(t)\|^2.$$

By (12), it ensues that

$$\ddot{h}_z(t) + \frac{\alpha}{t} \dot{h}_z(t) + \langle A_{\lambda(t)}(x(t)), x(t) - z \rangle = \|\dot{x}(t)\|^2. \quad (15)$$

Since $z \in S = A^{-1}(0)$, we have $A_{\lambda(t)}(z) = 0$, and the $\lambda(t)$ -cocoercivity of $A_{\lambda(t)}$ gives

$$\langle A_{\lambda(t)}(x(t)), x(t) - z \rangle \geq \lambda(t) \|A_{\lambda(t)}(x(t))\|^2. \quad (16)$$

It suffices to combine (15) and (16) to conclude. \square

As suggested in 1.2, we are likely to need a growth assumption on $\lambda(t)$ –related to $\lambda(t) \times \left(\frac{\alpha}{t}\right)^2 > 1$ – in order to go further. Whence, the following result:

Lemma 2.3 Given $z \in S = A^{-1}(0)$, define h_z by (13) and let

$$\lambda(t) \frac{\alpha^2}{t^2} \geq 1 + \epsilon \quad (17)$$

for some $\epsilon > 0$. Then, for each $z \in S = A^{-1}(0)$ and all $t \geq t_0$, we have

$$\ddot{h}_z(t) + \frac{\alpha}{t} \dot{h}_z(t) + \epsilon \|\dot{x}(t)\|^2 + \frac{\alpha \lambda(t)}{t} \frac{d}{dt} \|\dot{x}(t)\|^2 + \lambda(t) \|\ddot{x}(t)\|^2 \leq 0. \quad (18)$$

Proof By (12), we have

$$A_{\lambda(t)}(x(t)) = -\ddot{x}(t) - \frac{\alpha}{t} \dot{x}(t).$$

Replacing $A_{\lambda(t)}(x(t))$ by this expression in (14), we obtain

$$\ddot{h}_z(t) + \frac{\alpha}{t} \dot{h}_z(t) + \lambda(t) \left\| \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) \right\|^2 \leq \|\dot{x}(t)\|^2. \quad (19)$$

After expanding the third term on the left-hand side of this expression, we obtain

$$\ddot{h}_z(t) + \frac{\alpha}{t} \dot{h}_z(t) + \left(\lambda(t) \frac{\alpha^2}{t^2} - 1 \right) \|\dot{x}(t)\|^2 + \alpha \frac{\lambda(t)}{t} \frac{d}{dt} \|\dot{x}(t)\|^2 + \lambda(t) \|\ddot{x}(t)\|^2 \leq 0.$$

The result follows from (17). \square

2.2 Speed and acceleration decay

We now focus on the long-term behavior of the speed and acceleration. We have the following:

Proposition 2.4 *Let $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ be a maximally monotone operator such that $S = A^{-1}(0) \neq \emptyset$. Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution of (12), where*

$$\alpha > 2 \text{ and } \lambda(t) = (1 + \epsilon) \frac{t^2}{\alpha^2} \text{ for some } \epsilon > \frac{2}{\alpha - 2}.$$

Then, the trajectory $x(\cdot)$ is bounded, $\|\dot{x}(t)\| = \mathcal{O}(1/t)$, $\|\ddot{x}(t)\| = \mathcal{O}(1/t^2)$ and

$$\int_{t_0}^{+\infty} t \|\dot{x}(t)\|^2 dt < +\infty.$$

Proof As before, take $z \in S = A^{-1}(0)$ and define h_z by (13). First we simplify the writing of Eq. (18), given in Lemma 2.3. By setting $g(t) = \|\dot{x}(t)\|^2$, and $\beta = \frac{1+\epsilon}{\alpha}$, we have

$$\ddot{h}_z(t) + \frac{\alpha}{t} \dot{h}_z(t) + \epsilon g(t) + \beta t \dot{g}(t) + \lambda(t) \|\ddot{x}(t)\|^2 \leq 0. \quad (20)$$

Neglecting the positive term $\lambda(t) \|\ddot{x}(t)\|^2$ and multiplying by t , we obtain

$$t \ddot{h}_z(t) + \alpha \dot{h}_z(t) + \epsilon t g(t) + \beta t^2 \dot{g}(t) \leq 0. \quad (21)$$

Integrate this inequality from t_0 to t . We have

$$\begin{aligned} \int_{t_0}^t (\dot{s}\ddot{h}_z(s) + \alpha\dot{h}_z(s)) ds &= t\dot{h}_z(t) - t_0\dot{h}_z(t_0) - \int_{t_0}^t \dot{h}_z(s)ds + \alpha h_z(t) - \alpha h_z(t_0) \\ &= t\dot{h}_z(t) + (\alpha - 1)h_z(t) - t_0\dot{h}_z(t_0) - (\alpha - 1)h_z(t_0) \\ \int_{t_0}^t (\epsilon s g(s) + \beta s^2 \dot{g}(s)) ds &= \beta t^2 g(t) - \beta t_0^2 g(t_0) - 2\beta \int_{t_0}^t s g(s)ds + \epsilon \int_{t_0}^t s g(s)ds \\ &= \beta t^2 g(t) - \beta t_0^2 g(t_0) + (\epsilon - 2\beta) \int_{t_0}^t s g(s)ds. \end{aligned}$$

Adding the two above expressions, we obtain

$$t\dot{h}_z(t) + (\alpha - 1)h_z(t) + \beta t^2 g(t) + (\epsilon - 2\beta) \int_{t_0}^t s g(s)ds \leq C \quad (22)$$

for some positive constant C that depends only on the initial data. By the definition of $\beta = \frac{1+\epsilon}{\alpha}$, we have $\epsilon - 2\beta = \frac{\alpha-2}{\alpha} \left(\epsilon - \frac{2}{\alpha-2} \right)$. Since $\alpha > 2$ and $\epsilon > \frac{2}{\alpha-2}$, we observe that $\epsilon - 2\beta > 0$. Hence, (22) gives

$$t\dot{h}_z(t) + (\alpha - 1)h_z(t) \leq C. \quad (23)$$

Multiply this expression by $t^{\alpha-2}$ to obtain

$$\frac{d}{dt} t^{\alpha-1} h_z(t) \leq C t^{\alpha-2}.$$

Integrating from t_0 to t , we obtain

$$h_z(t) \leq \frac{C}{\alpha-1} + \frac{D}{t^{\alpha-1}},$$

for some other positive constant D . As a consequence, $h_z(\cdot)$ is bounded, and so is the trajectory $x(\cdot)$. Set

$$M := \sup_{t \geq t_0} \|x(t)\| < +\infty,$$

and note that

$$|\dot{h}_z(t)| = |\langle x(t) - z, \dot{x}(t) \rangle| \leq \|x(t) - z\| \|\dot{x}(t)\| \leq (M + \|z\|) \|\dot{x}(t)\|. \quad (24)$$

Combining (22) with (24) we deduce that

$$\beta (\| \dot{x}(t) \|)^2 \leq C + (M + \|z\|) (t \|\dot{x}(t)\|).$$

This immediately implies that

$$\sup_{t \geq t_0} t \|\dot{x}(t)\| < +\infty. \quad (25)$$

This gives

$$\|\dot{x}(t)\| = \mathcal{O}(1/t). \quad (26)$$

From (12), we have

$$\|\ddot{x}(t)\| \leq \frac{\alpha}{t} \|\dot{x}(t)\| + \|A_{\lambda(t)}(x(t))\| \leq \frac{\alpha}{t} \|\dot{x}(t)\| + \frac{M + \|z\|}{\lambda(t)}. \quad (27)$$

Using (26) and the definition of $\lambda(t)$, we conclude that

$$\|\ddot{x}(t)\| = \mathcal{O}(1/t^2). \quad (28)$$

Finally, returning to (22), and using (24) and (25) we infer that

$$\int_{t_0}^{+\infty} t \|\dot{x}(t)\|^2 dt < +\infty, \quad (29)$$

which completes the proof. \square

2.3 Proof of Theorem 2.1

We are now in a position to prove the convergence of the trajectories to equilibria. According to Opial's Lemma A.2, it suffices to verify that $\lim_{t \rightarrow +\infty} \|x(t) - z\|$ exists for each $z \in S$, and that every weak limit point of $x(t)$, as $t \rightarrow +\infty$, belongs to S .

We begin by proving that $\lim_{t \rightarrow +\infty} \|x(t) - z\|$ exists for every $z \in S$. By Lemma 2.2, we have

$$t\ddot{h}_z(t) + \alpha\dot{h}_z(t) + t\lambda(t)\|A_{\lambda(t)}(x(t))\|^2 \leq t\|\dot{x}(t)\|^2. \quad (30)$$

Since h_z is nonnegative, and in view of (29), Lemma A.6 shows that

$$\lim_{t \rightarrow +\infty} h_z(t)$$

exists. It follows that $\lim_{t \rightarrow +\infty} \|x(t) - z\|$ exists for every $z \in S$.

From Lemma A.6 and (30), we also have

$$\int_{t_0}^{+\infty} t\lambda(t)\|A_{\lambda(t)}(x(t))\|^2 dt < +\infty. \quad (31)$$

Since $\lambda(t) = ct^2$ for some positive constant c we infer

$$\int_{t_0}^{+\infty} \|\lambda(t)A_{\lambda(t)}(x(t))\|^2 \frac{1}{t} dt < +\infty. \quad (32)$$

The central point of the proof is to show that this property implies

$$\lim_{t \rightarrow +\infty} \|\lambda(t)A_{\lambda(t)}(x(t))\| = 0. \quad (33)$$

Suppose, for a moment, that this property holds. Then the end of the proof follows easily: let $t_n \rightarrow +\infty$ such that $x(t_n) \rightarrow \bar{x}$ weakly. We have $\lambda(t_n)A_{\lambda(t_n)}(x(t_n)) \rightarrow 0$ strongly. Since $\lambda(t_n) \rightarrow +\infty$, we also have $A_{\lambda(t_n)}(x(t_n)) \rightarrow 0$ strongly. Passing to the limit in

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

and using the demi-closedness of the graph of A , we obtain

$$0 \in A(\bar{x}).$$

In other words, $\bar{x} \in S$, and we conclude by Opial's Lemma.

As a consequence, it suffices to prove (33). To obtain this result, we shall estimate the variation of the function $t \mapsto \lambda(t)A_{\lambda(t)}$. By applying Lemma A.4 with $\gamma = \lambda(t)$, $\delta = \lambda(s)$, $x = x(t)$ and $y = y(s)$ with $s, t \geq t_0$, we obtain

$$\|\lambda(t)A_{\lambda(t)}x(t) - \lambda(s)A_{\lambda(s)}x(s)\| \leq 2\|x(t) - x(s)\| + 2\|x(t) - z\| \frac{|\lambda(t) - \lambda(s)|}{\lambda(t)} \quad (34)$$

for each fixed $z \in S$. Dividing by $t - s$ with $t \neq s$, and letting s tend to t , we deduce that

$$\left\| \frac{d}{dt} (\lambda(t)A_{\lambda(t)}x(t)) \right\| \leq 2\|\dot{x}(t)\| + 2\|x(t) - z\| \frac{|\dot{\lambda}(t)|}{\lambda(t)}$$

for almost every $t > t_0$. According to Proposition 2.4, the trajectory $x(\cdot)$ is bounded by some $M \geq 0$. In turn,

$$\|\dot{x}(t)\| \leq \frac{C}{t},$$

for some $C \geq 0$ and all $t \geq t_0$, by (25). Finally, the definition of $\lambda(t)$ implies

$$\frac{|\dot{\lambda}(t)|}{\lambda(t)} = \frac{2}{t}$$

for all $t \geq t_0$. As a consequence,

$$\left\| \frac{d}{dt} (\lambda(t)A_{\lambda(t)}x(t)) \right\| \leq \frac{2C + 4(M + \|z\|)}{t}.$$

This property, along with the boundedness of $\lambda(t)A_{\lambda(t)}x(t)$, and estimation (32), together imply that the nonnegative function $w(t) := \|\lambda(t)A_{\lambda(t)}x(t)\|^2$ satisfies

$$\left| \frac{d}{dt}w(t) \right| \leq \eta(t)$$

for almost every $t > t_0$, and

$$\int_{t_0}^{+\infty} w(t) \eta(t) dt < +\infty,$$

where

$$\eta(t) = \frac{2C + 4(M + \|z\|)}{t}.$$

Noting that $\eta \notin L^1(t_0, +\infty)$, we conclude thanks to Lemma A.5.

3 Proximal-based inertial algorithms with regularized operator

In this section, we introduce the inertial proximal algorithm which results from the discretization with respect to time of the continuous system

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + A_{\lambda(t)}(x(t)) = 0, \quad (35)$$

where $A_{\lambda(t)}$ is the Yosida approximation of index $\lambda(t)$ of the maximally monotone operator A . Further insight into the relationship between continuous- and discrete-time systems in variational analysis can be found in [30].

3.1 A regularized inertial proximal algorithm

We shall obtain an implementable algorithm by means of an *implicit* discretization of (35) with respect to time. Note that, in view of the Lipschitz continuity property of the operator A_λ , the explicit discretization might work well too. We choose to discretize it implicitly for two reasons: implicit discretizations tend to follow the continuous-time trajectories more closely; and the explicit discretization has the same iteration complexity (they each need one resolvent computation per iteration). Taking a fixed time step $h > 0$, and setting $t_k = kh$, $x_k = x(t_k)$, $\lambda_k = \lambda(t_k)$, an implicit finite-difference scheme for (35) with centered second-order variation gives

$$\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\alpha}{kh^2}(x_k - x_{k-1}) + A_{\lambda_k}(x_{k+1}) = 0. \quad (36)$$

After expanding (36), we obtain

$$x_{k+1} + h^2 A_{\lambda_k}(x_{k+1}) = x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}). \quad (37)$$

Setting $s = h^2$, we have

$$x_{k+1} = (I + s A_{\lambda_k})^{-1} \left(x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) \right), \quad (38)$$

where $(I + s A_{\lambda_k})^{-1}$ is the resolvent of index $s > 0$ of the maximally monotone operator A_{λ_k} . This gives the following algorithm

$$\begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) \\ x_{k+1} = (I + s A_{\lambda_k})^{-1} (y_k). \end{cases} \quad (39)$$

Using equality (84):

$$(A_\lambda)_s = A_{\lambda+s},$$

we can reformulate this last equation as

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

Hence, (39) can be rewritten as

$$\text{(RIPA)} \quad \begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) \\ x_{k+1} = \frac{\lambda_k}{\lambda_k + s} y_k + \frac{s}{\lambda_k + s} (I + (\lambda_k + s) A)^{-1} (y_k), \end{cases}$$

where (RIPA) stands for the Regularized Inertial Proximal Algorithm. Let us reformulate (RIPA) in a compact way. Observe that

$$\begin{aligned} x_{k+1} &= \frac{\lambda_k}{\lambda_k + s} y_k + \frac{s}{\lambda_k + s} (I + (\lambda_k + s) A)^{-1} (y_k) \\ &= \left(1 - \frac{s}{\lambda_k + s}\right) y_k + \frac{s}{\lambda_k + s} (I + (\lambda_k + s) A)^{-1} (y_k) \\ &= y_k - s \frac{1}{\lambda_k + s} \left(y_k - (I + (\lambda_k + s) A)^{-1} (y_k) \right). \end{aligned}$$

By the definition of A_{λ_k+s} , this is

$$x_{k+1} = y_k - s A_{\lambda_k+s} (y_k). \quad (40)$$

Thus, setting $\alpha_k = 1 - \frac{\alpha}{k}$, (RIPA) is just

$$\begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = y_k - sA_{\lambda_k+s}(y_k), \end{cases} \quad (41)$$

Remark 3.1 Letting $\lambda_k \rightarrow 0$ in (41), we obtain the classical form of the inertial proximal algorithm

$$\begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = (I + sA)^{-1}(y_k). \end{cases} \quad (42)$$

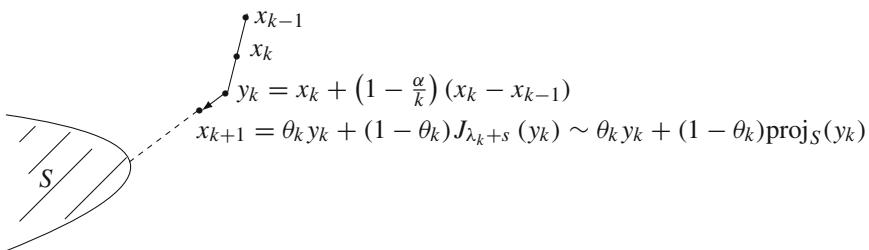
The case $0 \leq \alpha_k \leq \bar{\alpha} < 1$ has been considered by Álvarez and Attouch in [2]. The case $\alpha_k \rightarrow 1$, which is the most interesting for obtaining fast methods (in the line of Nesterov's accelerated methods), and the one that we are concerned with, was recently studied by Attouch and Cabot [3,4]. In these two papers, the convergence is obtained under the restrictive assumption

$$\sum_k \alpha_k \|x_{k+1} - x_k\|^2 < +\infty.$$

By contrast, our approach, which supposes that $\lambda_k \rightarrow +\infty$, provides convergence of the trajectories, without any restrictive assumption on the trajectories. Let us give a geometrical interpretation of (RIPA). As a classical property of the resolvents ([14, Theorem 23.44]), for any $x \in \mathcal{H}$, $J_\lambda x \rightarrow \text{proj}_S(x)$ as $\lambda \rightarrow +\infty$. Thus the algorithm writes

$$x_{k+1} = \theta_k y_k + (1 - \theta_k)J_{\lambda_k+s}(y_k)$$

with $\lambda_k \sim +\infty$, $\theta_k = \frac{\lambda_k}{\lambda_k+s} \sim 1$, and $J_{\lambda_k+s}(y_k) \sim \text{proj}_S(y_k)$. This is illustrated in the following picture.



Remark 3.2 As a main difference with (42), (RIPA) contains an additional momentum term $\frac{\lambda_k}{\lambda_k+s} y_k$, which enters the definition of x_{k+1} . Although there is some similarity, this is different from the algorithm introduced by Kim and Fessler in [23], which also contains an additional momentum term, but which comes within the definition of y_k .

It is important to mention that, in both cases, the introduction of this momentum term implies additional operations whose cost is negligible.

3.2 Preliminary estimations

Given $z \in \mathcal{H}$ and $k \geq 1$, write

$$h_{z,k} := \frac{1}{2} \|x_k - z\|^2. \quad (43)$$

Since there will be no risk of confusion, we shall write h_k for $h_{z,k}$ to simplify the notation. The following result is valid for an arbitrary sequence $\alpha_k \geq 0$:

Lemma 3.3 *Let $\alpha_k \geq 0$. For any $z \in \mathcal{H}$, the following holds for all $k \geq 1$:*

$$\begin{aligned} & (h_{k+1} - h_k) - \alpha_k(h_k - h_{k-1}) - \langle x_{k+1} - y_k, x_{k+1} - z \rangle \\ & + \frac{1}{2} \|x_{k+1} - y_k\|^2 = \frac{1}{2} (\alpha_k + \alpha_k^2) \|x_k - x_{k-1}\|^2. \end{aligned} \quad (44)$$

Proof Since $y_k = x_k + \alpha_k(x_k - x_{k-1})$, we have

$$\begin{aligned} \|y_k - z\|^2 &= \|(x_k - z) + \alpha_k(x_k - x_{k-1})\|^2 \\ &= \|x_k - z\|^2 + \alpha_k^2 \|x_k - x_{k-1}\|^2 + 2\alpha_k \langle x_k - z, x_k - x_{k-1} \rangle, \\ &= \|x_k - z\|^2 + \alpha_k^2 \|x_k - x_{k-1}\|^2 + \alpha_k \|x_k - z\|^2 \\ &\quad + \alpha_k \|x_k - x_{k-1}\|^2 - \alpha_k \|x_{k-1} - z\|^2 \\ &= \|x_k - z\|^2 + \alpha_k (\|x_k - z\|^2 - \|x_{k-1} - z\|^2) + (\alpha_k + \alpha_k^2) \|x_k - x_{k-1}\|^2 \\ &= 2(h_k + \alpha_k(h_k - h_{k-1})) + (\alpha_k + \alpha_k^2) \|x_k - x_{k-1}\|^2. \end{aligned} \quad (45)$$

Combining (45) with the elementary equality

$$(h_{k+1} - h_k) - \alpha_k(h_k - h_{k-1}) = \frac{1}{2} \|x_{k+1} - z\|^2 - (h_k + \alpha_k(h_k - h_{k-1})),$$

we deduce that

$$\begin{aligned} & (h_{k+1} - h_k) - \alpha_k(h_k - h_{k-1}) \\ &= \frac{1}{2} \|x_{k+1} - z\|^2 - \frac{1}{2} \|y_k - z\|^2 + \frac{1}{2} (\alpha_k + \alpha_k^2) \|x_k - x_{k-1}\|^2 \\ &= \langle x_{k+1} - y_k, \frac{1}{2}(x_{k+1} + y_k) - z \rangle + \frac{1}{2} (\alpha_k + \alpha_k^2) \|x_k - x_{k-1}\|^2 \\ &= \langle x_{k+1} - y_k, (x_{k+1} - z) + \frac{1}{2}(y_k - x_{k+1}) \rangle + \frac{1}{2} (\alpha_k + \alpha_k^2) \|x_k - x_{k-1}\|^2 \\ &= \langle x_{k+1} - y_k, x_{k+1} - z \rangle - \frac{1}{2} \|x_{k+1} - y_k\|^2 + \frac{1}{2} (\alpha_k + \alpha_k^2) \|x_k - x_{k-1}\|^2 \end{aligned}$$

which gives the claim. \square

Let us now use the specific form of the inertial algorithm (RIPA) and the cocoercivity of the operator A_λ . The following result is the discrete counterpart of Lemma 2.2:

Lemma 3.4 *Let $S = A^{-1}(0) \neq \emptyset$, and $0 \leq \alpha_k \leq 1$. For any $z \in S$, the following holds for all $k \geq 1$*

$$(h_{k+1} - h_k) - \alpha_k(h_k - h_{k-1}) + s\lambda_k \|A_{\lambda_k+s}(y_k)\|^2 \leq \alpha_k \|x_k - x_{k-1}\|^2. \quad (46)$$

Proof By Lemma 3.3 and the fact that $x_{k+1} = y_k - sA_{\lambda_k+s}(y_k)$ (see (40)), we have

$$\begin{aligned} & (h_{k+1} - h_k) - \alpha_k(h_k - h_{k-1}) + s\langle A_{\lambda_k+s}(y_k), y_k - sA_{\lambda_k+s}(y_k) - z \rangle \\ & + \frac{s^2}{2} \|A_{\lambda_k+s}(y_k)\|^2 = \frac{1}{2}(\alpha_k + \alpha_k^2) \|x_k - x_{k-1}\|^2. \end{aligned}$$

Since $z \in S$, we have $A_{\lambda_k+s}(z) = 0$. By the $(\lambda_k + s)$ -cocoercivity property of A_{λ_k+s} , we deduce that

$$\langle A_{\lambda_k+s}(y_k), y_k - z \rangle \geq (\lambda_k + s) \|A_{\lambda_k+s}(y_k)\|^2.$$

As a consequence,

$$\begin{aligned} & (h_{k+1} - h_k) - \alpha_k(h_k - h_{k-1}) + s\left(\lambda_k + \frac{s}{2}\right) \|A_{\lambda_k+s}(y_k)\|^2 \\ & \leq \frac{1}{2}(\alpha_k + \alpha_k^2) \|x_k - x_{k-1}\|^2. \end{aligned}$$

But $\frac{1}{2}(\alpha_k + \alpha_k^2) \leq \alpha_k$ because $0 \leq \alpha_k \leq 1$. Since $s \geq 0$, the result follows immediately. \square

It would be possible to continue the analysis assuming the right-hand side of (46) is summable. The main disadvantage of this hypothesis is that it involves the trajectory (x_k) , which is unknown. In the following lemma, we show that the two antagonistic terms $s(\lambda_k + \frac{s}{2}) \|A_{\lambda_k+s}(y_k)\|^2$ and $\alpha_k \|x_k - x_{k-1}\|^2$ can be balanced, provided λ_k is taken large enough.

Lemma 3.5 *Let $S = A^{-1}(0) \neq \emptyset$, and take $\alpha_k = 1 - \frac{\alpha}{k}$ with $\alpha > 2$. For each $k \geq 1$, set*

$$\lambda_k = (1 + \epsilon) \frac{s}{\alpha^2} k^2, \quad (47)$$

for some $\epsilon > 0$, and write $\beta = \frac{1+\epsilon}{\alpha}$. Then, for each $z \in S$ and all $k \geq 1$, we have

$$(h_{k+1} - h_k) - \alpha_k(h_k - h_{k-1}) + \epsilon \|x_k - x_{k-1}\|^2 + \beta k \left(\|x_{k+1} - x_k\|^2 - \|x_k - x_{k-1}\|^2 \right) \leq 0. \quad (48)$$

Proof First, rewrite (46) as

$$(h_{k+1} - h_k) - \alpha_k(h_k - h_{k-1}) + \frac{\lambda_k}{s} \|x_{k+1} - y_k\|^2 \leq \alpha_k \|x_k - x_{k-1}\|^2. \quad (49)$$

Let us write $\|x_{k+1} - y_k\|^2$ in a recursive form. To this end, we use the specific form of $\alpha_k = 1 - \frac{\alpha}{k}$ to obtain

$$\begin{aligned} \|x_{k+1} - y_k\|^2 &= \|(x_{k+1} - x_k) - \alpha_k(x_k - x_{k-1})\|^2 \\ &= \|(x_{k+1} - x_k) - (x_k - x_{k-1}) + (1 - \alpha_k)(x_k - x_{k-1})\|^2 \\ &= \left\| (x_{k+1} - x_k) - (x_k - x_{k-1}) + \frac{\alpha}{k}(x_k - x_{k-1}) \right\|^2 \\ &= \|(x_{k+1} - x_k) - (x_k - x_{k-1})\|^2 + \frac{\alpha^2}{k^2} \|x_k - x_{k-1}\|^2 \\ &\quad + 2 \frac{\alpha}{k} \langle (x_{k+1} - x_k) - (x_k - x_{k-1}), x_k - x_{k-1} \rangle. \end{aligned}$$

But

$$\begin{aligned} \frac{1}{2} \|x_{k+1} - x_k\|^2 &= \frac{1}{2} \|x_k - x_{k-1}\|^2 + \langle (x_{k+1} - x_k) - (x_k - x_{k-1}), x_k - x_{k-1} \rangle \\ &\quad + \frac{1}{2} \|(x_{k+1} - x_k) - (x_k - x_{k-1})\|^2. \end{aligned}$$

By combining the above equalities, we get

$$\begin{aligned} \|x_{k+1} - y_k\|^2 &= \left(1 - \frac{\alpha}{k}\right) \|(x_{k+1} - x_k) - (x_k - x_{k-1})\|^2 + \frac{\alpha^2}{k^2} \|x_k - x_{k-1}\|^2 \\ &\quad + \frac{\alpha}{k} \left(\|x_{k+1} - x_k\|^2 - \|x_k - x_{k-1}\|^2 \right). \end{aligned}$$

Using this equality in (49), and neglecting the nonnegative term $(1 - \frac{\alpha}{k}) \|(x_{k+1} - x_k) - (x_k - x_{k-1})\|^2$, we obtain

$$\begin{aligned} &(h_{k+1} - h_k) - \alpha_k(h_k - h_{k-1}) \\ &+ \frac{\lambda_k \alpha^2}{s k^2} \|x_k - x_{k-1}\|^2 + \frac{\lambda_k \alpha}{s k} \left(\|x_{k+1} - x_k\|^2 - \|x_k - x_{k-1}\|^2 \right) \leq \|x_k - x_{k-1}\|^2. \end{aligned} \quad (50)$$

Using (47) and the definition $\beta = \frac{1+\epsilon}{\alpha}$, inequality (50) becomes (48). \square

3.3 Main convergence result

We are now in position to prove the main result of this section, namely:

Theorem 3.6 Let $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ be a maximally monotone operator such that $S = A^{-1}(0) \neq \emptyset$. Let (x_k) be a sequence generated by the Regularized Inertial Proximal Algorithm

$$(RIPA) \quad \begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}) \\ x_{k+1} = \frac{\lambda_k}{\lambda_k + s}y_k + \frac{s}{\lambda_k + s}(I + (\lambda_k + s)A)^{-1}(y_k), \end{cases}$$

where $\alpha > 2$ and

$$\lambda_k = (1 + \epsilon)\frac{s}{\alpha^2}k^2$$

for some $\epsilon > \frac{2}{\alpha-2}$ and all $k \geq 1$. Then,

- (i) The speed tends to zero. More precisely, $\|x_{k+1} - x_k\| = \mathcal{O}(\frac{1}{k})$ and $\sum_k k\|x_k - x_{k-1}\|^2 < +\infty$.
- (ii) The sequence (x_k) converges weakly, as $k \rightarrow +\infty$, to some $\hat{x} \in S$.
- (iii) The sequence (y_k) converges weakly, as $k \rightarrow +\infty$, to \hat{x} .

Proof First, we simplify the writing of the Eq. (48) given in Lemma 3.5. Setting $g_k := \|x_k - x_{k-1}\|^2$, we have

$$(h_{k+1} - h_k) - \alpha_k(h_k - h_{k-1}) + \epsilon g_k + \beta k(g_{k+1} - g_k) \leq 0.$$

Then, we multiply by k to obtain

$$k(h_{k+1} - h_k) - (k - \alpha)(h_k - h_{k-1}) + \epsilon kg_k + \beta k^2(g_{k+1} - g_k) \leq 0.$$

We now write these inequalities in a recursive form, in order to simplify their summation. We have

$$\begin{aligned} k(h_{k+1} - h_k) - (k - 1)(h_k - h_{k-1}) + (\alpha - 1)(h_k - h_{k-1}) + \epsilon kg_k \\ + \beta k^2 g_{k+1} - \beta(k - 1)^2 g_k - \beta(2k - 1)g_k \leq 0. \end{aligned}$$

Summing for $p = 1, \dots, k$, we obtain

$$k(h_{k+1} - h_k) + (\alpha - 1)h_k + (\epsilon - 2\beta) \sum_1^k pg_p + \beta k^2 g_{k+1} + \beta \sum_1^k g_p \leq C \quad (51)$$

for some positive constant C that depends only on the initial data. Since $\beta = \frac{1+\epsilon}{\alpha}$ with $\alpha > 2$ and $\epsilon > \frac{2}{\alpha-2}$, we have $\epsilon - 2\beta = \frac{\alpha-2}{\alpha}\left(\epsilon - \frac{2}{\alpha-2}\right) > 0$. From (51) we infer that

$$k(h_{k+1} - h_k) + (\alpha - 1)h_k \leq C \quad (52)$$

for all $k \geq 1$. Since $\alpha > 2$ and $h_k \geq 0$, (52) implies

$$k(h_{k+1} - h_k) + h_k \leq C.$$

Equivalently,

$$kh_{k+1} - (k-1)h_k \leq C.$$

Applying this fact recursively, we deduce that

$$kh_{k+1} \leq Ck,$$

which immediately gives $\sup_k h_k < +\infty$. Therefore, the sequence (x_k) is bounded. Set

$$M := \sup_k \|x_k\| < +\infty.$$

Now, (51) also implies that

$$k(h_{k+1} - h_k) + \beta k^2 g_{k+1} \leq C \quad (53)$$

But

$$\begin{aligned} |h_{k+1} - h_k| &= \frac{1}{2} \left| \|x_{k+1} - z\|^2 - \|x_k - z\|^2 \right| = \left| \left\langle x_{k+1} - x_k, \frac{1}{2}(x_{k+1} + x_k) - z \right\rangle \right| \\ &\leq (M + \|z\|) \|x_{k+1} - x_k\|. \end{aligned} \quad (54)$$

Combining this inequality with (53), and recalling the definition $g_k = \|x_k - x_{k-1}\|^2$, we deduce that

$$\beta [k \|x_{k+1} - x_k\|]^2 - (M + \|z\|) [k \|x_{k+1} - x_k\|] - C \leq 0.$$

This immediately implies that

$$\sup_k k \|x_{k+1} - x_k\| < +\infty. \quad (55)$$

In other words, $\|x_{k+1} - x_k\| = \mathcal{O}(\frac{1}{k})$. Another consequence of (51) is that

$$(\epsilon - 2\beta) \sum_{p=1}^k p \|x_p - x_{p-1}\|^2 \leq C - k(h_{k+1} - h_k).$$

By (54), we deduce that

$$(\epsilon - 2\beta) \sum_{p=1}^k p \|x_p - x_{p-1}\|^2 \leq C + k(M + \|z\|) \|x_{k+1} - x_k\|.$$

Then, (55) gives

$$\sum_k k \|x_k - x_{k-1}\|^2 < +\infty, \quad (56)$$

which completes the proof of item i).

For the convergence of the sequence (x_k) , we use Opial's Lemma A.3. First, since $\alpha_k = 1 - \frac{\alpha}{k} \leq 1$, Lemma 3.4 gives

$$(h_{k+1} - h_k) + s \lambda_k \|A_{\lambda_k+s}(y_k)\|^2 \leq \left(1 - \frac{\alpha}{k}\right) (h_k - h_{k-1}) + \|x_k - x_{k-1}\|^2$$

for all $k \geq 1$. Using (56) and invoking Lemma A.7, we deduce that

$$\sum_k k \lambda_k \|A_{\lambda_k+s}(y_k)\|^2 < +\infty. \quad (57)$$

and

$$\sum_k [h_k - h_{k-1}]_+ < +\infty.$$

Since h_k is nonnegative, this implies the existence of $\lim_{k \rightarrow +\infty} h_k$, and also that of $\lim_{k \rightarrow +\infty} \|x_k - z\|$.

In order to conclude using Opial's Lemma A.3, it remains to show that every weak limit point of the sequence (x_k) , as $k \rightarrow +\infty$, belongs to S . We begin by expressing (57) with respect to x_k , instead of y_k . We have

$$\begin{aligned} \|A_{\lambda_k+s}(x_k)\|^2 &\leq 2\|A_{\lambda_k+s}(y_k)\|^2 + 2\|A_{\lambda_k+s}(x_k) - A_{\lambda_k+s}(y_k)\|^2 \\ &\leq 2\|A_{\lambda_k+s}(y_k)\|^2 + \frac{2}{(\lambda_k + s)^2} \|y_k - x_k\|^2 \\ &\leq 2\|A_{\lambda_k+s}(y_k)\|^2 + \frac{2}{\lambda_k^2} \|x_k - x_{k-1}\|^2, \end{aligned}$$

where the last inequality follows from the definition of y_k given in (39). Using (55) and the definition of λ_k , we may find a constant $D \geq 0$ such that

$$\|A_{\lambda_k+s}(x_k)\|^2 \leq 2\|A_{\lambda_k+s}(y_k)\|^2 + \frac{D}{k^2 \lambda_k^2}.$$

Hence,

$$\sum_k k\lambda_k \|A_{\lambda_k+s}(x_k)\|^2 \leq 2 \sum_k k\lambda_k \|A_{\lambda_k+s}(y_k)\|^2 + \sum_k \frac{D}{k\lambda_k}.$$

Now, using (57) and the definition of λ_k , we conclude that

$$\sum_k k\lambda_k \|A_{\lambda_k+s}(x_k)\|^2 < +\infty.$$

Since λ_k tends to infinity, this immediately implies

$$\sum_k k(\lambda_k + s) \|A_{\lambda_k+s}(x_k)\|^2 < +\infty.$$

To simplify the notations, set $\mu_k = \lambda_k + s$, so that

$$\sum_k k\mu_k \|A_{\mu_k}(x_k)\|^2 < +\infty. \quad (58)$$

As we shall see, this fact implies

$$\lim_{k \rightarrow +\infty} \|\mu_k A_{\mu_k}(x_k)\| = 0. \quad (59)$$

Suppose, for a moment, that this is true. Let $(x_{k_n})_n$ be a subsequence of (x_k) which converges weakly to some \bar{x} . We want to prove that $\bar{x} \in S = A^{-1}(0)$. Since μ_k tends to infinity, we also have

$$\lim_{k \rightarrow +\infty} \|A_{\mu_k}(x_k)\| = 0.$$

Passing to the limit in

$$A_{\mu_{k_n}}(x_{k_n}) \in A(x_{k_n} - \mu_{k_n} A_{\mu_{k_n}}(x_{k_n})),$$

and using the demi-closedness of the graph of A , we obtain

$$0 \in A(\bar{x}).$$

In other words, $\bar{x} \in S$. As a consequence, it only remains to prove (59) in order to obtain ii) by Opial's Lemma. To this end, define

$$\omega_k := \|\mu_k A_{\mu_k}(x_k)\|^2.$$

We intend to prove that $\lim_{k \rightarrow +\infty} \omega_k = 0$. Using (58) and the definition of μ_k , we deduce that

$$\sum_k \frac{1}{k} \omega_k < +\infty. \quad (60)$$

Therefore, if $\lim_{k \rightarrow +\infty} \omega_k$ exists, it must be zero. Since

$$\|\mu_k A_{\mu_k}(x_k)\| = \|x_k - J_{\mu_k A}(x_k)\| \leq \|x_k - z\| \leq M + \|z\|,$$

we have

$$\begin{aligned} |\omega_{k+1} - \omega_k| &= \left| \|\mu_{k+1} A_{\mu_{k+1}}(x_{k+1})\|^2 - \|\mu_k A_{\mu_k}(x_k)\|^2 \right| \\ &\leq 2(M + \|z\|) \|\mu_{k+1} A_{\mu_{k+1}}(x_{k+1}) - \mu_k A_{\mu_k}(x_k)\|. \end{aligned} \quad (61)$$

On the other hand, by Lemma A.4 we have

$$\begin{aligned} \|\mu_{k+1} A_{\mu_{k+1}}(x_{k+1}) - \mu_k A_{\mu_k}(x_k)\| &\leq 2\|x_{k+1} - x_k\| + 2\|x_{k+1} - z\| \frac{|\mu_{k+1} - \mu_k|}{\mu_{k+1}} \\ &\leq 2\|x_{k+1} - x_k\| + 2\|x_{k+1} - z\| \frac{|\lambda_{k+1} - \lambda_k|}{\lambda_{k+1}} \\ &\leq 2\|x_{k+1} - x_k\| + \frac{4(M + \|z\|)}{k+1}, \end{aligned}$$

by the definition of λ_k . Using (55) and replacing the resulting inequality in (61), we deduce that there is a constant $E \geq 0$ such that

$$|\omega_{k+1} - \omega_k| \leq \frac{E}{k}.$$

But then

$$\sum_k |(\omega_{k+1})^2 - (\omega_k)^2| \leq E \sum_k \frac{1}{k} (\omega_{k+1} + \omega_k) < +\infty$$

by (60). It follows that $\lim_{k \rightarrow +\infty} \omega_k^2$ exists and, since $\omega_k \geq 0$, $\lim_{k \rightarrow +\infty} \omega_k$ exists as well. This completes the proof of item ii). Finally, item iii) follows from the fact that $\lim_k \|x_{k+1} - y_k\| = \lim_k \|sA_{\lambda_k+s}(y_k)\| = 0$. \square

3.4 An application to convex-concave saddle value problems

As shown by Rockafellar [33], to each closed convex-concave function $L : X \times Y \rightarrow \bar{\mathbb{R}}$ acting on the product of two Hilbert spaces X and Y is associated a maximally monotone operator $M : X \times Y \rightrightarrows X \times Y$ which is given by $M = (\partial_x L, -\partial_y L)$. This makes it possible to convert convex-concave saddle value problems into the search

for the zeros of a maximally monotone operator, and thus to apply our results. Let's illustrate it in the case of the convex constrained structured minimization problem

$$(P) \quad \min \{f(x) + g(y) : Ax - By = 0\},$$

where data satisfy the following assumptions:

- X, Y, Z are real Hilbert spaces
- $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ and $g : Y \rightarrow \mathbb{R} \cup \{+\infty\}$ are closed convex proper functions.
- $A : X \rightarrow Z$ and $B : Y \rightarrow Z$ are linear continuous operators.

Let us first reformulate (P) as a saddle value problem

$$\min_{(x,y) \in X \times Y} \max_{z \in Z} \{f(x) + g(y) + \langle z, Ax - By \rangle\}. \quad (62)$$

The Lagrangian $L : X \times Y \times Z \rightarrow \mathbb{R} \cup \{+\infty\}$ associated to (62)

$$L(x, y, z) = f(x) + g(y) + \langle z, Ax - By \rangle$$

is a convex-concave extended-real-valued function. The maximal monotone operator $M : X \times Y \times Z \rightrightarrows X \times Y \times Z$ that is associated to L is given by

$$M(x, y, z) = (\partial_{x,y} L, -\partial_z L)(x, y, z) = (\partial f(x) + A^t z, \partial g(y) - B^t z, By - Ax). \quad (63)$$

When the proximal algorithm is applied to the maximally monotone operator M , we obtain the so-called proximal method of multipliers. This method was initiated by Rockafellar [34]. By combining this method with the alternating proximal minimization algorithm for weakly coupled minimization problems, a fully split method is obtained. This approach was successfully developed by Attouch and Soueycatt in [10]. Introducing inertial terms in this algorithm, as given by (RIPA), is a subject of further study, which is part of the active research on the acceleration of the (ADMM) algorithms.

4 Stability with respect to perturbations, errors

In this section, we discuss the stability of the convergence results with respect to external perturbations. We first consider the continuous case, then the corresponding algorithmic results.

4.1 The continuous case

The continuous dynamics is now written in the following form

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + A_{\lambda(t)}(x(t)) = f(t), \quad (64)$$

where, depending on the context, f can be interpreted as a source term, a perturbation, or an error. We suppose that f is locally integrable to ensure existence and uniqueness for the corresponding Cauchy problem (see Lemma A.1 in the “Appendix”). Assuming that $f(t)$ tends to zero fast enough as $t \rightarrow +\infty$, we will see that the convergence results proved in Sect. 2 are still valid for the perturbed dynamics (64). Due to its similarity to the unperturbed case, we give the main lines of the proof, just highlighting the differences.

Theorem 4.1 *Let $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ be a maximally monotone operator such that $S = A^{-1}(0) \neq \emptyset$. Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution of the continuous dynamic (64), where $\alpha > 2$ and $\lambda(t) = (1 + \epsilon) \frac{t^2}{\alpha^2}$ with $\epsilon > \frac{2}{\alpha - 2}$. Assume also that $\int_{t_0}^{+\infty} t^3 \|f(t)\|^2 dt < +\infty$ and $\int_{t_0}^{+\infty} t \|f(t)\| dt < +\infty$. Then, $x(t)$ converges weakly, as $t \rightarrow +\infty$, to an element of S . Moreover $\|\dot{x}(t)\| = \mathcal{O}(1/t)$.*

Proof First, a similar computation as in Lemma 2.2 gives

$$\ddot{h}_z(t) + \frac{\alpha}{t} \dot{h}_z(t) + \lambda(t) \|A_{\lambda(t)}(x(t))\|^2 \leq \|\dot{x}(t)\|^2 + \|x(t) - z\| \|f(t)\|. \quad (65)$$

Following the arguments in the proof of Lemma 2.3, we use (64), then develop and simplify (65) to obtain

$$\ddot{h}_z(t) + \frac{\alpha}{t} \dot{h}_z(t) + \epsilon g(t) + \beta t \dot{g}(t) \leq \|x(t) - z\| \|f(t)\| + 2\beta t \|\dot{x}(t)\| \|f(t)\|, \quad (66)$$

where, as in the proof of Proposition 2.4, we have set $g(t) = \|\dot{x}(t)\|^2$. Using the fact that for any $0 < \theta < 1$

$$2\beta t \|\dot{x}(t)\| \|f(t)\| \leq \epsilon \theta \|\dot{x}(t)\|^2 + \frac{\beta^2}{\epsilon \theta} t^2 \|f(t)\|^2,$$

and multiplying by t , we obtain

$$t \ddot{h}_z(t) + \alpha \dot{h}_z(t) + \epsilon(1 - \theta) t g(t) + \beta t^2 \dot{g}(t) \leq t \|x(t) - z\| \|f(t)\| + \frac{\beta^2}{\epsilon \theta} t^3 \|f(t)\|^2. \quad (67)$$

Integration from t_0 to t yields

$$\begin{aligned} t \dot{h}_z(t) + (\alpha - 1) h_z(t) + \beta t^2 g(t) + (\epsilon(1 - \theta) - 2\beta) \int_{t_0}^t s g(s) ds &\leq C \\ &+ \int_{t_0}^t s \|x(s) - z\| \|f(s)\| ds + \frac{\beta^2}{\epsilon \theta} \int_{t_0}^t s^3 \|f(s)\|^2 ds \end{aligned} \quad (68)$$

for some positive constant C that depends only on the initial data. In all that follows, C is a generic notation for a constant. By the definition $\beta = \frac{1+\epsilon}{\alpha}$ and the assumptions and

the parameters α and ϵ , we can choose θ positive small enough to get $\epsilon(1-\theta)-2\beta > 0$. Taking into account also the hypothesis $\int_{t_0}^{+\infty} t^3 \|f(t)\|^2 dt < +\infty$, we deduce that

$$t \dot{h}_z(t) + (\alpha - 1) h_z(t) \leq C + \int_{t_0}^t s \|x(s) - z\| \|f(s)\| ds.$$

Multiply this expression by $t^{\alpha-2}$, integrate from t_0 to t , and use Fubini's Theorem to obtain

$$\frac{1}{2} \|x(t) - z\|^2 \leq C + \frac{1}{\alpha - 1} \int_{t_0}^t t \|x(t) - z\| \|f(t)\| dt.$$

The main difference with Sect. 2 is here. We apply Gronwall's Lemma (see [16, Lemma A.5]) to get

$$\|x(t) - z\| \leq C + \frac{1}{\alpha - 1} \int_{t_0}^t t \|f(t)\| dt.$$

Since $\int_{t_0}^{+\infty} t \|f(t)\| dt < +\infty$, we deduce that the trajectory $x(\cdot)$ is bounded. The rest of the proof is essentially the same. First, we obtain

$$\sup_{t \geq t_0} t \|\dot{x}(t)\| < +\infty,$$

by bounding that quantity between the roots of a quadratic expression. Then, we go back to (68) to get that

$$\int_{t_0}^{+\infty} t \|\dot{x}(t)\|^2 dt < +\infty.$$

We use Lemma A.6 to deduce that $\lim_{t \rightarrow +\infty} h_z(t)$ exists and

$$\int_{t_0}^{+\infty} t \lambda(t) \|A_{\lambda(t)}(x(t))\|^2 dt < +\infty.$$

The latter implies $\lim_{t \rightarrow +\infty} \lambda(t) \|A_{\lambda(t)}(x(t))\| = 0$, and we conclude by means of Opial's Lemma A.2. \square

4.2 The algorithmic case

Let us first consider how the introduction of the external perturbation f into the continuous dynamics modifies the corresponding algorithm. Setting $f_k = f(kh)$, a discretization similar to that of the unperturbed case gives

$$\frac{1}{h^2} (x_{k+1} - 2x_k + x_{k-1}) + \frac{\alpha}{kh^2} (x_k - x_{k-1}) + A_{\lambda_k}(x_{k+1}) = f_k.$$

After expanding this expression, and setting $s = h^2$, we obtain

$$x_{k+1} + sA_{\lambda_k}(x_{k+1}) = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}) + sf_k,$$

which gives

$$\begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}) \\ x_{k+1} = (I + sA_{\lambda_k})^{-1}(y_k + sf_k). \end{cases}$$

Using the resolvent Eq. (84), we obtain the Regularized Inertial Proximal Algorithm with perturbation

$$\text{(RIPA-pert)} \quad \begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}) \\ x_{k+1} = \frac{\lambda_k}{\lambda_k + s}(y_k + sf_k) + \frac{s}{\lambda_k + s}J_{(\lambda_k+s)A}(y_k + sf_k). \end{cases}$$

Setting $\alpha_k = 1 - \frac{\alpha}{k}$, and with help of the Yosida approximation, this can be written in a compact way as

$$\begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = (y_k + sf_k) - sA_{\lambda_k+s}(y_k + sf_k). \end{cases} \quad (69)$$

When $f_k = 0$ we recover (RIPA). The convergence of (RIPA-pert) algorithm is analyzed in the following theorem.

Theorem 4.2 *Let $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ be a maximally monotone operator such that $S = A^{-1}(0) \neq \emptyset$. Let (x_k) be a sequence generated by the algorithm (RIPA-pert) where $\alpha > 2$ and*

$$\lambda_k = \left(1 + \frac{s}{2} + \epsilon\right) \frac{2s}{\alpha^2} k^2$$

for some $\epsilon > \frac{2+s}{\alpha-2}$ and all $k \geq 1$. Suppose that $\sum_k k\|f_k\| < +\infty$ and $\sum_k k^3\|f_k\|^2 < +\infty$. Then,

- (i) *The speed tends to zero. More precisely, $\|x_{k+1} - x_k\| = \mathcal{O}(\frac{1}{k})$ and $\sum_k k\|x_k - x_{k-1}\|^2 < +\infty$.*
- (ii) *The sequence (x_k) converges weakly, as $k \rightarrow +\infty$, to some $\hat{x} \in S$.*
- (iii) *The sequence (y_k) converges weakly, as $k \rightarrow +\infty$, to \hat{x} .*

Proof Let us observe that the definitions of y_k and h_k are the same as in the unperturbed case. Hence, it is only when using the constitutive equation $x_{k+1} = (y_k + sf_k) - sA_{\lambda_k+s}(y_k + sf_k)$, which contains the perturbation term, that changes occur in the proof. Thus, the beginning of the proof and Lemma 3.3 is still valid, which gives

$$(h_{k+1} - h_k) - \alpha_k(h_k - h_{k-1}) - \langle x_{k+1} - y_k, x_{k+1} - z \rangle$$

$$+\frac{1}{2}\|x_{k+1}-y_k\|^2=\frac{1}{2}(\alpha_k+\alpha_k^2)\|x_k-x_{k-1}\|^2. \quad (70)$$

The next step, which corresponds to Lemma 3.4, uses the constitutive equation. Let us adapt it to our situation. By (70) and $x_{k+1}-y_k=s(f_k-A_{\lambda_k+s}(y_k+sf_k))$, it follows that

$$\begin{aligned} & (h_{k+1}-h_k)-\alpha_k(h_k-h_{k-1})+s\langle A_{\lambda_k+s}(y_k+sf_k) \\ & -f_k, (y_k+sf_k-z)-sA_{\lambda_k+s}(y_k+sf_k) \rangle \\ & +\frac{s^2}{2}\|A_{\lambda_k+s}(y_k+sf_k)-f_k\|^2=\frac{1}{2}(\alpha_k+\alpha_k^2)\|x_k-x_{k-1}\|^2. \end{aligned} \quad (71)$$

Since $z \in S$, we have $A_{\lambda_k+s}(z)=0$. By the (λ_k+s) -cocoercivity property of A_{λ_k+s} , we deduce that

$$\langle A_{\lambda_k+s}(y_k+sf_k), y_k+sf_k-z \rangle \geq (\lambda_k+s)\|A_{\lambda_k+s}(y_k+sf_k)\|^2.$$

Using the above inequality in (71), and after development and simplification, we obtain

$$\begin{aligned} & (h_{k+1}-h_k)-\alpha_k(h_k-h_{k-1})+s\left(\lambda_k+\frac{s}{2}\right)\|A_{\lambda_k+s}(y_k+sf_k)\|^2 \\ & \leq \frac{s^2}{2}\|f_k\|^2+s\langle y_k-z, f_k \rangle+\frac{1}{2}(\alpha_k+\alpha_k^2)\|x_k-x_{k-1}\|^2. \end{aligned} \quad (72)$$

From $A_{\lambda_k+s}(y_k+sf_k)=\frac{1}{s}(y_k+sf_k-x_{k+1})$ it follows that

$$\begin{aligned} & (h_{k+1}-h_k)-\alpha_k(h_k-h_{k-1})+\frac{1}{s}\left(\lambda_k+\frac{s}{2}\right)\|(y_k-x_{k+1})+sf_k\|^2 \\ & \leq \frac{s^2}{2}\|f_k\|^2+s\langle y_k-z, f_k \rangle+\frac{1}{2}(\alpha_k+\alpha_k^2)\|x_k-x_{k-1}\|^2. \end{aligned}$$

Using the elementary inequality $\|y_k-x_{k+1}\|^2 \leq 2\|(y_k-x_{k+1})+sf_k\|^2+2\|sf_k\|^2$, we deduce that

$$\begin{aligned} & (h_{k+1}-h_k)-\alpha_k(h_k-h_{k-1})+\frac{1}{2s}\left(\lambda_k+\frac{s}{2}\right)\|x_{k+1}-y_k\|^2 \\ & \leq \frac{s^2}{2}\|f_k\|^2+s\left(\lambda_k+\frac{s}{2}\right)\|f_k\|^2+s\langle y_k-z, f_k \rangle+\frac{1}{2}(\alpha_k+\alpha_k^2)\|x_k-x_{k-1}\|^2. \end{aligned}$$

Since $\frac{1}{2}(\alpha_k+\alpha_k^2) \leq \alpha_k \leq 1$, $s > 0$, and by using Cauchy–Schwarz inequality, it follows that

$$\begin{aligned} & (h_{k+1}-h_k)-\alpha_k(h_k-h_{k-1})+\frac{\lambda_k}{2s}\|x_{k+1}-y_k\|^2 \\ & \leq s(s+\lambda_k)\|f_k\|^2+s\|y_k-z\|\|f_k\|+\|x_k-x_{k-1}\|^2. \end{aligned} \quad (73)$$

By the definition of y_k and elementary inequalities we have

$$\begin{aligned} \|y_k - z\| \|f_k\| &\leq \|x_k - z\| \|f_k\| + \|x_k - x_{k-1}\| \|f_k\| \leq \|x_k - z\| \|f_k\| \\ &+ \frac{1}{2} \|f_k\|^2 + \frac{1}{2} \|x_k - x_{k-1}\|^2. \end{aligned}$$

Combining this inequality with (73) we obtain

$$\begin{aligned} (h_{k+1} - h_k) - \alpha_k(h_k - h_{k-1}) + \frac{\lambda_k}{2s} \|x_{k+1} - y_k\|^2 \\ \leq s \left(\frac{1}{2} + s + \lambda_k \right) \|f_k\|^2 + s \|x_k - z\| \|f_k\| \\ + \left(1 + \frac{s}{2} \right) \|x_k - x_{k-1}\|^2. \end{aligned} \quad (74)$$

Let us write $\|x_{k+1} - y_k\|^2$ in a recursive form. The same computation as in Lemma 3.5 gives

$$\begin{aligned} \|x_{k+1} - y_k\|^2 &= \left(1 - \frac{\alpha}{k} \right) \|(x_{k+1} - x_k) - (x_k - x_{k-1})\|^2 \\ &+ \frac{\alpha^2}{k^2} \|x_k - x_{k-1}\|^2 + \frac{\alpha}{k} \left(\|x_{k+1} - x_k\|^2 - \|x_k - x_{k-1}\|^2 \right). \end{aligned}$$

Using this equality in (74), and neglecting the nonnegative term $(1 - \frac{\alpha}{k})\|(x_{k+1} - x_k) - (x_k - x_{k-1})\|^2$, we obtain

$$\begin{aligned} (h_{k+1} - h_k) - \alpha_k(h_k - h_{k-1}) + \frac{\lambda_k \alpha^2}{2sk^2} \|x_k - x_{k-1}\|^2 \\ + \frac{\lambda_k \alpha}{2sk} \left(\|x_{k+1} - x_k\|^2 - \|x_k - x_{k-1}\|^2 \right) \\ \leq s \left(\frac{1}{2} + s + \lambda_k \right) \|f_k\|^2 + s \|x_k - z\| \|f_k\| + \left(1 + \frac{s}{2} \right) \|x_k - x_{k-1}\|^2. \end{aligned} \quad (75)$$

Using $\lambda_k = (1 + \frac{s}{2} + \epsilon) \frac{2s}{\alpha^2} k^2$ and the definition $\beta = \frac{1 + \frac{s}{2} + \epsilon}{\alpha}$, inequality (75) becomes

$$\begin{aligned} (h_{k+1} - h_k) - \alpha_k(h_k - h_{k-1}) + \epsilon \|x_k - x_{k-1}\|^2 + \beta k \left(\|x_{k+1} - x_k\|^2 - \|x_k - x_{k-1}\|^2 \right) \\ \leq s \left(\frac{1}{2} + s + \lambda_k \right) \|f_k\|^2 + s \|x_k - z\| \|f_k\|. \end{aligned} \quad (76)$$

Setting $g_k := \|x_k - x_{k-1}\|^2$, we have

$$\begin{aligned} (h_{k+1} - h_k) - \alpha_k(h_k - h_{k-1}) + \epsilon g_k + \beta k (g_{k+1} - g_k) \\ \leq s \left(\frac{1}{2} + s + \lambda_k \right) \|f_k\|^2 + s \|x_k - z\| \|f_k\|. \end{aligned}$$

Then, we multiply by k to obtain

$$\begin{aligned} k(h_{k+1} - h_k) - (k - \alpha)(h_k - h_{k-1}) + \epsilon kg_k + \beta k^2(g_{k+1} - g_k) \\ \leq sk \left(\frac{1}{2} + s + \lambda_k \right) \|f_k\|^2 + sk\|x_k - z\|\|f_k\|. \end{aligned}$$

We now write these inequalities in a recursive form, in order to simplify their summation. We have

$$\begin{aligned} k(h_{k+1} - h_k) - (k - 1)(h_k - h_{k-1}) + (\alpha - 1)(h_k - h_{k-1}) \\ + \epsilon kg_k + \beta k^2 g_{k+1} - \beta(k - 1)^2 g_k - \beta(2k - 1)g_k \\ \leq sk \left(\frac{1}{2} + s + \lambda_k \right) \|f_k\|^2 + sk\|x_k - z\|\|f_k\|. \end{aligned}$$

Summing for $p = 1, \dots, k$, we obtain

$$\begin{aligned} k(h_{k+1} - h_k) + (\alpha - 1)h_k + (\epsilon - 2\beta) \sum_1^k pg_p + \beta k^2 g_{k+1} + \beta \sum_1^k g_p \\ \leq s \sum_1^k p\|x_p - z\|\|f_p\| + s \sum_1^k p \left(\frac{1}{2} + s + \lambda_p \right) \|f_p\|^2. \end{aligned} \quad (77)$$

Since $\beta = \frac{1+\frac{s}{2}+\epsilon}{\alpha}$ with $\alpha > 2$ and $\epsilon > \frac{2+s}{\alpha-2}$, we have $\epsilon - 2\beta = \frac{\epsilon(\alpha-2)-(s+2)}{\alpha} > 0$. Moreover by the definition of λ_k and the assumption $\sum_k k^3\|f_k\|^2 < +\infty$, we have $s \sum_1^k p(\frac{1}{2} + s + \lambda_p)\|f_p\|^2 \leq C$ for some positive constant C . Whence

$$k(h_{k+1} - h_k) + (\alpha - 1)h_k + \beta k^2 g_{k+1} \leq C + \sum_1^k p\|x_p - z\|\|f_p\| \quad (78)$$

for all $k \geq 1$. Since $\alpha > 2$ and $h_k \geq 0$, (78) implies

$$kh_{k+1} - (k - 1)h_k \leq C + \sum_1^k p\|x_p - z\|\|f_p\|.$$

By summing the above inequalities, and applying Fubini's Theorem, we deduce that

$$kh_{k+1} \leq Ck + \sum_1^k p\|x_p - z\|\|f_p\|(k - p) \leq Ck + k \sum_1^k p\|x_p - z\|\|f_p\|.$$

Hence

$$\frac{1}{2} \|x_{k+1} - z\|^2 \leq C + \sum_1^k p \|x_p - z\| \|f_p\| (k-p) \leq C + \sum_1^k p \|x_p - z\| \|f_p\|.$$

Applying the discrete form of the Gronwall's Lemma A.8, and $\sum_k k \|f_k\| < +\infty$, we obtain $\sup_k h_k < +\infty$. Therefore, the sequence (x_k) is bounded. The remainder of the proof is pretty much as that of Theorem 3.6. We first derive

$$\sup_k k \|x_{k+1} - x_k\| < +\infty. \quad (79)$$

Then, we combine (77) with (79) to obtain

$$\sum_k k \|x_k - x_{k-1}\|^2 < +\infty. \quad (80)$$

Since $\alpha_k = 1 - \frac{\alpha}{k} \leq 1$, inequalities (72) and (80) give

$$(h_{k+1} - h_k) + s \left(\lambda_k + \frac{s}{2} \right) \|A_{\lambda_k+s}(y_k)\|^2 \leq \left(1 - \frac{\alpha}{k} \right) (h_k - h_{k-1}) + \theta_k$$

for all $k \geq 1$, and $\sum_{k \in \mathbb{N}} k \theta_k < +\infty$. Invoking Lemma A.7, we deduce that $\lim_{k \rightarrow +\infty} h_k$ exists and

$$\sum_k k \lambda_k \|A_{\lambda_k+s}(y_k)\|^2 < +\infty. \quad (81)$$

We conclude using Opial's Lemma A.3 as in the unperturbed case. \square

Remark 4.3 The perturbation can be interpreted either as a miscomputation of y_k from the two previous iterates, or as an error due to the fact that the resolvent can be computed at a neighboring point $y_k + sf_k$, rather than y_k . Anyway, perturbations of order less than $\frac{1}{k^2}$ are admissible and the convergence properties are preserved.

5 Quadratic growth and strong convergence

In this section, we examine the case of a maximally monotone operator A satisfying a quadratic growth property. More precisely, we assume that there is $v > 0$ such that

$$\langle x^*, x - z \rangle \geq v \operatorname{dist}(x, S)^2 \quad (82)$$

whenever $x^* \in Ax$ and $z \in S$. If A is strongly monotone, then (82) holds and S is a singleton. Another example is the subdifferential of a convex function Φ satisfying a quadratic error bound (see [18]). Indeed,

$$\langle x^*, x - z \rangle \geq \Phi(x) - \min(\Phi) \geq v \operatorname{dist}(x, S)^2$$

if $x^* \in \partial\Phi(x)$ and $z \in S = \operatorname{argmin}(\Phi)$. A particular case is when $A = M^*M$, where M is a bounded linear operator with closed range (if $\Phi(x) = \frac{1}{2}\|Mx\|^2$, then $\nabla\Phi(x) = M^*Mx$). We have,

$$\langle Ax, x - z \rangle = \|M(x - z)\|^2 \geq v \operatorname{dist}(x, \operatorname{Ker}(M))^2$$

(see [17, Exercise 2.14]).

We obtain the following convergence result:

Theorem 5.1 *Let $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ be a maximally monotone operator satisfying (82) for some $v > 0$ and all $x \in \mathcal{H}$ and $z \in S$. Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution of the continuous dynamic*

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + A_{\lambda(t)}(x(t)) = 0,$$

where $\alpha > 2$ and

$$\lambda(t) = (1 + \epsilon)\frac{t^2}{\alpha^2} \text{ with } \epsilon > \frac{2}{\alpha - 2}.$$

Then, $\lim_{t \rightarrow +\infty} \operatorname{dist}(x(t), S) = 0$. If, moreover, $S = \{\bar{z}\}$, then $x(t)$ converges strongly to \bar{z} as $t \rightarrow +\infty$.

Proof First, fix $t \geq t_0$ and observe that

$$\begin{aligned} \langle A_{\lambda(t)}(x(t)), x(t) - z \rangle &= \langle A_{\lambda(t)}(x(t)), J_{\lambda(t)A}(x(t)) - z \rangle \\ &\quad + \langle A_{\lambda(t)}(x(t)), x(t) - J_{\lambda(t)A}(x(t)) \rangle \\ &\geq v \operatorname{dist}(J_{\lambda(t)A}(x(t)), S)^2 + \frac{1}{\lambda(t)}\|x(t) - J_{\lambda(t)A}(x(t))\|^2 \\ &\geq v \operatorname{dist}(J_{\lambda(t)A}(x(t)), S)^2 \\ &\quad + \frac{1}{\lambda(t)} \left[(1 - \xi)\|x(t) - z\|^2 \right. \\ &\quad \left. + \left(1 - \frac{1}{\xi}\right)\|z - J_{\lambda(t)A}(x(t))\|^2 \right] \end{aligned}$$

for all $\xi > 0$ (we shall select a convenient value later on). In turn, the left-hand side satisfies

$$\begin{aligned} \langle A_{\lambda(t)}(x(t)), x(t) - z \rangle &\leq \|A_{\lambda(t)}(x(t))\| \|x(t) - z\| \\ &\leq \lambda(t) \|A_{\lambda(t)}(x(t))\|^2 + \|A_{\lambda(t)}(x(t))\| \|J_{\lambda(t)A}(x(t)) - z\| \\ &\leq \left(\lambda(t) + \frac{\xi}{2v} \right) \|A_{\lambda(t)}(x(t))\|^2 + \frac{v}{2\xi} \|J_{\lambda(t)A}(x(t)) - z\|^2. \end{aligned}$$

Since $\|x(t) - z\| \geq \text{dist}(x(t), S)$, by taking z as the projection of $J_{\lambda(t)A}(x(t))$ onto S , and combining the last two inequalities, we obtain

$$\begin{aligned} & \frac{(1-\zeta)}{\lambda(t)} \text{dist}(x(t), S)^2 + \left(v - \frac{v}{2\zeta} + \frac{\zeta-1}{\zeta\lambda(t)} \right) \text{dist}(J_{\lambda(t)A}(x(t)), S)^2 \\ & \leq \left(\lambda(t) + \frac{\zeta}{2v} \right) \|A_{\lambda(t)}(x(t))\|^2. \end{aligned}$$

whenever $0 < \zeta < 1$. Set

$$\zeta = \frac{2 + \lambda(t)v}{2(1 + \lambda(t)v)} = \frac{1}{2} + \frac{1}{2(1 + \lambda(t)v)},$$

so that

$$0 < \frac{1}{2} < \zeta \leq \frac{1}{2} + \frac{1}{2(1 + \lambda(t_0)v)} < 1,$$

and

$$v - \frac{v}{2\zeta} + \frac{\zeta-1}{\zeta\lambda(t)} = \frac{2\zeta\lambda(t)v - \lambda(t)v + 2\zeta - 2}{2\zeta\lambda(t)} = \frac{2\zeta(1 + \lambda(t)v) - (2 + \lambda(t)v)}{2\zeta\lambda(t)} = 0.$$

It follows that

$$\frac{(1-\zeta)}{\lambda(t)} \text{dist}(x(t), S)^2 \leq \left(\lambda(t) + \frac{\zeta}{2v} \right) \|A_{\lambda(t)}(x(t))\|^2,$$

and so

$$\text{dist}(x(t), S)^2 \leq \left[\frac{2(1 + \lambda(t_0)v)}{\lambda(t_0)v} \right] \lambda(t) \left(\lambda(t) + \frac{1}{2v} \right) \|A_{\lambda(t)}(x(t))\|^2.$$

The right-hand side goes to zero by (33). \square

A similar result holds for (RIPA), namely:

Theorem 5.2 *Let $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ be a maximally monotone operator satisfying (82) for some $v > 0$ and all $x \in \mathcal{H}$ and $z \in S$. Let (x_k) be a sequence generated by the algorithm (RIPA), where $\alpha > 2$ and $\lambda_k = (1 + \epsilon) \frac{s k^2}{\alpha^2}$ with $\epsilon > \frac{2}{\alpha-2}$. Then, $\lim_{k \rightarrow +\infty} \text{dist}(x_k, S) = 0$. If, moreover, $S = \{\bar{z}\}$, then x_k converges strongly to \bar{z} as $k \rightarrow +\infty$.*

6 Further conclusions from a keynote example

Let us illustrate our results in the case where $\mathcal{H} = \mathbb{R}^2$, and A is the counterclockwise rotation centered at the origin and with the angle $\frac{\pi}{2}$, namely $A(x, y) = (-y, x)$. This is a model situation for a maximally monotone operator that is not cocoercive. The linear operator A is antisymmetric, that is, $\langle A(x, y), (x, y) \rangle = 0$ for all $(x, y) \in \mathcal{H}$.

6.1 Critical parameters

Our results are based on an appropriate tuning of the Yosida approximation parameter. Let us analyze the asymptotic behavior of the solution trajectories of the second-order differential equation

$$\ddot{u}(t) + \frac{\alpha}{t} \dot{u}(t) + A_{\lambda(t)}(u(t)) = 0, \quad (83)$$

where $u(t) = (x(t), y(t))$. Since 0 is the unique zero of A , the question is to find the conditions on $\lambda(t)$ which ensure the convergence of $u(t)$ to zero. An elementary computation gives

$$A_\lambda = \frac{1}{1 + \lambda^2} \begin{pmatrix} \lambda & -1 \\ 1 & \lambda \end{pmatrix}.$$

For easy computation, it is convenient to set $z(t) = x(t) + iy(t)$, and work with the equivalent formulation of the problem in the Hilbert space $\mathcal{H} = \mathbb{C}$, equipped with the real Hilbert structure $\langle z_1, z_2 \rangle = \operatorname{Re}(z_1 \bar{z}_2)$. So, the operator and its Yosida approximation are given respectively by $Az = iz$ and $A_\lambda z = \frac{\lambda+i}{1+\lambda^2}z$. Then (83) becomes

$$\ddot{z}(t) + \frac{\alpha}{t} \dot{z}(t) + \frac{\lambda+i}{1+\lambda^2} z(t) = 0.$$

Passing to the phase space $\mathbb{C} \times \mathbb{C}$, and setting $Z(t) = (z(t), \dot{z}(t))^T$, we obtain the first-order equivalent system

$$\dot{Z}(t) + M(t)Z(t) = 0 \quad \text{where } M(t) = \begin{pmatrix} 0 & -1 \\ \frac{\lambda(t)+i}{1+\lambda(t)^2} & \frac{\alpha}{t} \end{pmatrix}.$$

The asymptotic behavior of the trajectories of this system can be analyzed by examining the eigenvalues of the matrix $M(t)$, which are given by

$$\theta(t) = \frac{\alpha}{2t} \left\{ 1 \pm \sqrt{1 - \frac{4t^2}{\alpha^2} \frac{\lambda(t)+i}{1+\lambda(t)^2}} \right\}.$$

Let us restrict ourselves to the case $\lambda(t) \sim t^p$. If $p > 2$, the eigenvalues θ_+ and θ_- satisfy

$$\theta_+(t) \sim \frac{1}{t} \quad \text{and} \quad \theta_-(t) \sim \frac{1}{t^{p-1}}.$$

Although the solutions of the differential equation $\dot{v}(t) + \frac{1}{t}v(t) = 0$ converge to 0, those of $\dot{v}(t) + \frac{1}{t^{p-1}}v(t) = 0$ do not. Thus, to obtain the convergence results of our theorem, we are not allowed to let $\lambda(t)$ tend to infinity at a rate greater than t^2 , which shows that t^2 is a critical size for $\lambda(t)$.

Table 1 Distance to the unique equilibrium for a solution of each equation

Key	Differential equation	Distance to $(0, 0)$ at $t = 100$
(E1)	$\dot{x}(t) + Ax(t) = 0$	14.141911
(E2)	$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + Ax(t) = 0$	3.186e24
(E3)	$\dot{x}(t) + A_{\lambda(t)}(x(t)) = 0$	0.0135184
(E4)	$\dot{x}(t) + A_{\lambda}(x(t)) = 0$	0.0007827
(E5)	$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + A_{\lambda(t)}x(t) = 0$	0.000323

6.2 A comparative illustration

As an illustration, we depict solutions of some first- and second-order equations involving the rotation operator A , obtained using Scilab's `ode` solver. In all cases, the initial condition at $t_0 = 1$ is $(10, 10)$. For second-order equations, we take the initial velocity as $(0, 0)$ in order not to force the system in any direction. When relevant, we take $\lambda(t) = (1 + \epsilon)t^2/\alpha^2$ with $\alpha = 10$ and $\epsilon = 1 + 2(\alpha - 2)^{-1}$. For the constant λ , we set $\lambda = 10$. Table 1 shows the distance to the unique equilibrium $(\bar{x}, \bar{y}) = (0, 0)$ at $t = 100$.

Observe that the final position of the solution of (E5) is comparable to that of (E4), which is a first-order equation governed by the *strongly monotone* operator A_λ .

Acknowledgements The authors thank P. Redont for his careful and constructive reading of the paper.

A Auxiliary results

A.1 Yosida regularization of an operator A

Given a maximally monotone operator A and $\lambda > 0$, the *resolvent* of A with index λ and the *Yosida regularization* of A with parameter λ are defined by

$$J_{\lambda A} = (I + \lambda A)^{-1} \quad \text{and} \quad A_\lambda = \frac{1}{\lambda} (I - J_{\lambda A}),$$

respectively. The operator $J_{\lambda A} : \mathcal{H} \rightarrow \mathcal{H}$ is nonexpansive and everywhere defined (indeed it is firmly non-expansive). Moreover, A_λ is λ -cocoercive: for all $x, y \in \mathcal{H}$ we have

$$\langle A_\lambda y - A_\lambda x, y - x \rangle \geq \lambda \|A_\lambda y - A_\lambda x\|^2.$$

This property immediately implies that $A_\lambda : \mathcal{H} \rightarrow \mathcal{H}$ is $\frac{1}{\lambda}$ -Lipschitz continuous. Another property that proves useful is the *resolvent equation* (see, for example, [16, Proposition 2.6] or [14, Proposition 23.6])

$$(A_\lambda)_\mu = A_{(\lambda+\mu)}, \tag{84}$$

which is valid for any $\lambda, \mu > 0$. This property allows to compute simply the resolvent of A_λ : for any $\lambda, \mu > 0$ by

$$J_{\mu A_\lambda} = \frac{\lambda}{\lambda + \mu} I + \frac{\mu}{\lambda + \mu} J_{(\lambda + \mu)A}.$$

Also note that for any $x \in \mathcal{H}$, and any $\lambda > 0$

$$A_\lambda(x) \in A(J_{\lambda A}x) = A(x - \lambda A_\lambda(x)).$$

Finally, for any $\lambda > 0$, A and A_λ have the same solution set $S := A_\lambda^{-1}(0) = A^{-1}(0)$. For a detailed presentation of the properties of the maximally monotone operators and the Yosida approximation, the reader can consult [14] or [16].

A.2 Existence and uniqueness of solution in the presence of a source term

Let us first establish the existence and uniqueness of the solution trajectory of the Cauchy problem associated to the continuous regularized dynamic (1) with a source term.

Lemma A.1 *Take $t_0 > 0$. Let us suppose that $\lambda : [t_0, +\infty[\rightarrow \mathbb{R}^+$ is a measurable function such that $\lambda(t) \geq \underline{\lambda}$ for some $\underline{\lambda} > 0$. Suppose that $f \in L^1([t_0, T], \mathcal{H})$ for all $T \geq t_0$. Then, for any $x_0 \in \mathcal{H}$, $v_0 \in \mathcal{H}$, there exists a unique strong global solution $x : [t_0, +\infty[\rightarrow \mathcal{H}$ of the Cauchy problem*

$$\begin{cases} \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + A_{\lambda(t)}(x(t)) = f(t) \\ x(t_0) = x_0, \quad \dot{x}(t_0) = v_0. \end{cases} \quad (85)$$

Proof The argument is standard, and consists in writing (85) as a first-order system in the phase space. By setting

$$X(t) = \begin{pmatrix} x(t) \\ \dot{x}(t) \end{pmatrix}, \quad F(t, u, v) = \begin{pmatrix} v \\ -\frac{\alpha}{t}v - A_{\lambda(t)}u + f(t) \end{pmatrix} \text{ and } X_0 = \begin{pmatrix} x_0 \\ v_0 \end{pmatrix},$$

the system can be written as

$$\begin{cases} \dot{X}(t) = F(t, X(t)) \\ X(t_0) = X_0. \end{cases} \quad (86)$$

Using the $\frac{1}{\lambda}$ -Lipschitz continuity property of A_λ , one can easily verify that the conditions of the Cauchy–Lipschitz theorem are satisfied. Precisely, we can apply the non-autonomous version of this theorem given in [21, Proposition 6.2.1]. Thus, we obtain a strong solution, that is, $t \mapsto \dot{x}(t)$ is locally absolutely continuous. If, moreover, we suppose that the functions $\lambda(\cdot)$ and f are continuous, then the solution is a classical solution of class C^2 . \square

A.3 Opial's Lemma

The following results are often referred to as Opial's Lemma [28]. To our knowledge, it was first written in this form in Baillon's thesis. See [30] for a proof.

Lemma A.2 *Let S be a nonempty subset of \mathcal{H} and let $x : [0, +\infty[\rightarrow \mathcal{H}$. Assume that*

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

Then $x(t)$ converges weakly as $t \rightarrow \infty$ to a point in S .

Its discrete version is

Lemma A.3 *Let S be a non empty subset of \mathcal{H} , and (x_k) a sequence of elements of \mathcal{H} . Assume that*

- (i) *for every $z \in S$, $\lim_{k \rightarrow +\infty} \|x_k - z\|$ exists;*
- (ii) *every weak sequential limit point of (x_k) , as $k \rightarrow \infty$, belongs to S .*

Then x_k converges weakly as $k \rightarrow \infty$ to a point in S .

A.4 Variation of the function $\gamma \mapsto \gamma A_\gamma x$

Lemma A.4 *Let $\gamma, \delta > 0$, and $x, y \in \mathcal{H}$. Then, for each $z \in S = A^{-1}(0)$, and all $t \geq t_0$, we have*

$$\|\gamma A_\gamma x - \delta A_\delta y\| \leq 2\|x - y\| + 2\|x - z\| \frac{|\gamma - \delta|}{\gamma} \quad (87)$$

Proof We use successively the definition of the Yosida approximation, the resolvent identity [14, Proposition 23.28 (i)], and the nonexpansive property of the resolvent, to obtain

$$\begin{aligned} \|\gamma A_\gamma x - \delta A_\delta y\| &\leq \|x - y\| + \|J_{\gamma A} x - J_{\delta A} y\| \\ &= \|x - y\| + \|J_{\delta A} \left(\frac{\delta}{\gamma} x + \left(1 - \frac{\delta}{\gamma}\right) J_{\gamma A} x \right) - J_{\delta A} y\| \\ &\leq \|x - y\| + \left\| \frac{\delta}{\gamma} x + \left(1 - \frac{\delta}{\gamma}\right) J_{\gamma A} x - y \right\| \\ &\leq 2\|x - y\| + \left|1 - \frac{\delta}{\gamma}\right| \|J_{\gamma A} x - x\|. \end{aligned}$$

Since $J_{\gamma A} z = z$ for $z \in S$, and using again the nonexpansive property of the resolvent, we deduce that

$$\begin{aligned}\|\gamma A_\gamma x - \delta A_\delta y\| &\leq 2\|x - y\| + |1 - \frac{\delta}{\gamma}| \| (J_{\gamma A}x - J_{\gamma A}z) + (z - x) \| \\ &\leq 2\|x - y\| + 2\|x - z\| \frac{|\gamma - \delta|}{\gamma},\end{aligned}$$

which gives the claim. \square

A.5 On integration and decay

Lemma A.5 *Let $w, \eta : [t_0, +\infty[\rightarrow [0, +\infty[$ be absolutely continuous functions such that $\eta \notin L^1(t_0, +\infty)$,*

$$\int_{t_0}^{+\infty} w(t) \eta(t) dt < +\infty,$$

and $|\dot{w}(t)| \leq \eta(t)$ for almost every $t > t_0$. Then, $\lim_{t \rightarrow +\infty} w(t) = 0$.

Proof First, for almost every $t > t_0$, we have

$$\left| \frac{d}{dt} w^2(t) \right| = 2 \left| \frac{d}{dt} w(t) \right| w(t) \leq 2w(t) \eta(t).$$

Therefore, $|\frac{d}{dt} w^2|$ belongs to L^1 . This implies that $\lim_{t \rightarrow +\infty} w^2(t)$ exists. Since w is nonnegative, it follows that $\lim_{t \rightarrow +\infty} w(t)$ exists as well. But this limit is necessarily zero because $\eta \notin L^1$. \square

A.6 On boundedness and anchoring

Lemma A.6 *Let $t_0 > 0$, and let $w : [t_0, +\infty[\rightarrow \mathbb{R}$ be a continuously differentiable function which is bounded from below. Given a nonnegative function θ , let us assume that*

$$t \ddot{w}(t) + \alpha \dot{w}(t) + \theta(t) \leq k(t), \quad (88)$$

for some $\alpha > 1$, almost every $t > t_0$, and some nonnegative function $k \in L^1(t_0, +\infty)$. Then, the positive part $[\dot{w}]_+$ of \dot{w} belongs to $L^1(t_0, +\infty)$, and $\lim_{t \rightarrow +\infty} w(t)$ exists. Moreover, we have $\int_{t_0}^{+\infty} \theta(t) dt < +\infty$.

Proof Multiply (88) by $t^{\alpha-1}$ to obtain

$$\frac{d}{dt} (t^\alpha \dot{w}(t)) + t^{\alpha-1} \theta(t) \leq t^{\alpha-1} k(t).$$

By integration, we obtain

$$\dot{w}(t) + \frac{1}{t^\alpha} \int_{t_0}^t s^{\alpha-1} \theta(s) ds \leq \frac{t_0^\alpha |\dot{w}(t_0)|}{t^\alpha} + \frac{1}{t^\alpha} \int_{t_0}^t s^{\alpha-1} k(s) ds. \quad (89)$$

Hence,

$$[\dot{w}]_+(t) \leq \frac{t_0^\alpha |\dot{w}(t_0)|}{t^\alpha} + \frac{1}{t^\alpha} \int_{t_0}^t s^{\alpha-1} k(s) ds,$$

and so,

$$\int_{t_0}^\infty [\dot{w}]_+(t) dt \leq \frac{t_0^\alpha |\dot{w}(t_0)|}{(\alpha-1)t_0^{\alpha-1}} + \int_{t_0}^\infty \frac{1}{t^\alpha} \left(\int_{t_0}^t s^{\alpha-1} k(s) ds \right) dt.$$

Applying Fubini's Theorem, we deduce that

$$\int_{t_0}^\infty \frac{1}{t^\alpha} \left(\int_{t_0}^t s^{\alpha-1} k(s) ds \right) dt = \int_{t_0}^\infty \left(\int_s^\infty \frac{1}{t^\alpha} dt \right) s^{\alpha-1} k(s) ds = \frac{1}{\alpha-1} \int_{t_0}^\infty k(s) ds.$$

As a consequence,

$$\int_{t_0}^\infty [\dot{w}]_+(t) dt \leq \frac{t_0^\alpha |\dot{w}(t_0)|}{(\alpha-1)t_0^{\alpha-1}} + \frac{1}{\alpha-1} \int_{t_0}^\infty k(s) ds < +\infty.$$

This implies $\lim_{t \rightarrow +\infty} w(t)$ exists. Back to (89), integrating from t_0 to t , using Fubini's Theorem again, and then letting t tend to $+\infty$, we obtain

$$\lim_{t \rightarrow +\infty} w(t) - w(t_0) + \frac{1}{\alpha-1} \int_{t_0}^\infty \theta(s) ds \leq \frac{t_0^\alpha |\dot{w}(t_0)|}{(\alpha-1)t_0^{\alpha-1}} + \frac{1}{\alpha-1} \int_{t_0}^\infty k(s) ds < +\infty.$$

Hence $\int_{t_0}^\infty \theta(s) ds < +\infty$. □

A.7 A summability result for real sequences

Lemma A.7 *Let $\alpha > 1$, and let (h_k) be a sequence of real numbers which is bounded from below, and such that*

$$(h_{k+1} - h_k) - \left(1 - \frac{\alpha}{k}\right) (h_k - h_{k-1}) + \omega_k \leq \theta_k \quad (90)$$

for all $k \geq 1$. Suppose that (ω_k) , and (θ_k) are two sequences of nonnegative numbers, such that $\sum_k k\theta_k < +\infty$. Then

$$\sum_{k \in \mathbb{N}} [h_k - h_{k-1}]_+ < +\infty \quad \text{and} \quad \sum_{k \in \mathbb{N}} k\omega_k < +\infty.$$

Proof Since (ω_k) is nonnegative, we have

$$(h_{k+1} - h_k) - \left(1 - \frac{\alpha}{k}\right) (h_k - h_{k-1}) \leq \theta_k.$$

Setting $b_k := [h_k - h_{k-1}]_+$ the positive part of $h_k - h_{k-1}$, we immediately infer that

$$b_{k+1} \leq \left(1 - \frac{\alpha}{k}\right) b_k + \theta_k$$

for all $k \geq 1$. Multiplying by k and rearranging the terms, we obtain

$$(\alpha - 1)b_k \leq (k - 1)b_k - kb_{k+1} + k\theta_k.$$

Summing for $k = 1, \dots, K$, and using the telescopic property, along with the fact that $Kb_{K+1} \geq 0$, we deduce that

$$(\alpha - 1) \sum_{k=1}^K b_k \leq \sum_{k=1}^K k\theta_k,$$

which gives

$$\sum_{k \in \mathbb{N}} [h_k - h_{k-1}]_+ < +\infty.$$

Let us now prove that $\sum_{k \in \mathbb{N}} k\omega_k < +\infty$, which is the most delicate part of the proof. To this end, write $\delta_k = h_k - h_{k-1}$, and $\alpha_k = (1 - \frac{\alpha}{k})$, so that (90) becomes

$$\delta_{k+1} + \omega_k \leq \alpha_k \delta_k + \theta_k.$$

An immediate recurrence (it can be easily seen by induction) shows that

$$\delta_{k+1} + \sum_{i=1}^k \left[\left(\prod_{j=i+1}^k \alpha_j \right) \omega_i \right] \leq \left(\prod_{j=1}^k \alpha_j \right) \delta_1 + \sum_{i=1}^k \left[\left(\prod_{j=i+1}^k \alpha_j \right) \theta_i \right],$$

with the convention $\prod_{j=k+1}^k \alpha_j = 1$. To simplify the notation, write $A_i^k = \prod_{j=i}^k \alpha_j$. Sum the above inequality for $k = 1, \dots, K$ to deduce that

$$h_{K+1} - h_1 + \sum_{k=1}^K \sum_{i=1}^k A_{i+1}^k \omega_i \leq \delta_1 \sum_{k=1}^K A_1^k + \sum_{k=1}^K \sum_{i=1}^k A_{i+1}^k \theta_i. \quad (91)$$

Now, using Fubini's Theorem, we obtain

$$h_{K+1} - h_1 + \sum_{i=1}^K \left[\omega_i \sum_{k=i}^K A_{i+1}^k \right] \leq \delta_1 \sum_{k=1}^K A_1^k + \sum_{i=1}^K \left[\theta_i \sum_{k=i}^K A_{i+1}^k \right]. \quad (92)$$

Simple computations (using integrals in the estimations) show that

$$\left(\frac{i}{k} \right)^\alpha \leq A_{i+1}^k \leq \left(\frac{i+1}{k+1} \right)^\alpha,$$

and

$$\frac{i}{\alpha - 1} \leq \sum_{k=i}^{\infty} A_{i+1}^k \leq \frac{i}{\alpha - 1} \left(\frac{i+1}{i} \right)^{\alpha}$$

(see also [4] for further details). Letting $K \rightarrow +\infty$ in (92), we deduce that

$$\sum_{i=1}^{\infty} i \omega_i \leq C + D \sum_{i=1}^{\infty} i \theta_i < +\infty$$

for appropriate constants C and D . \square

A.8 A discrete Gronwall lemma

Lemma A.8 *Let $c \geq 0$ and let (a_k) and (β_j) be nonnegative sequences such that (β_j) is summable and*

$$a_k^2 \leq c^2 + \sum_{j=1}^k \beta_j a_j$$

for all $k \in \mathbb{N}$. Then, $a_k \leq c + \sum_{j=1}^{\infty} \beta_j$ for all $k \in \mathbb{N}$.

Proof For $k \in \mathbb{N}$, set $A_k := \max_{1 \leq m \leq k} a_m$. Then, for $1 \leq m \leq k$, we have

$$a_m^2 \leq c^2 + \sum_{j=1}^m \beta_j a_j \leq c^2 + A_k \sum_{j=1}^{\infty} \beta_j.$$

Taking the maximum over $1 \leq m \leq k$, we obtain

$$A_k^2 \leq c^2 + A_k \sum_{j=1}^{\infty} \beta_j.$$

Bounding by the roots of the corresponding quadratic equation, we obtain the result. \square

References

1. Álvarez, F.: On the minimizing property of a second-order dissipative system in Hilbert spaces. SIAM J. Control Optim. **38**(4), 1102–1119 (2000)
2. Álvarez, F., Attouch, H.: An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. Set-Valued Anal. **9**(1–2), 3–11 (2001)

3. Attouch, H., Cabot, A.: Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity. *J. Differ. Equ.* **263**(9), 5412–5458 (2017)
4. Attouch, H., Cabot, A.: Convergence rates of inertial forward-backward algorithms, HAL-01453170 (2017)
5. Attouch, H., Cabot, A., Redont, P.: The dynamics of elastic shocks via epigraphical regularization of a differential inclusion. *Adv. Math. Sci. Appl.* **12**(1), 273–306 (2002)
6. Attouch, H., Chbani, Z., Peypouquet, J., Redont, P.: Fast convergence of inertial dynamics and algorithms with asymptotic vanishing damping, to appear in *Math. Program.* <https://doi.org/10.1007/s10107-016-0992-8>
7. Attouch, H., Maingé, P.E.: Asymptotic behavior of second order dissipative evolution equations combining potential with non-potential effects. *ESAIM Control Optim. Calc. Var.* **17**(3), 836–857 (2011)
8. Attouch, H., Peypouquet, J.: The rate of convergence of Nesterov's accelerated forward–backward method is actually faster than $\frac{1}{k^2}$. *SIAM J. Optim.* **26**(3), 1824–1834 (2016)
9. Attouch, H., Peypouquet, J., Redont, P.: Fast convergence of regularized inertial dynamics for nonsmooth convex optimization, Working paper. (2017)
10. Attouch, H., Soueycatt, M.: Augmented Lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces. Applications to games, PDE's and control. *Pac. J. Optim.* **5**(1), 17–37 (2009)
11. Attouch, H., Wets, R.: Epigraphical processes: laws of large numbers for random LSC functions. *Sem. Anal. Convexe Montp.* **20**, 13–29 (1990)
12. Attouch, H., Wets, R.: Quantitative stability of variational systems: I, the epigraphical distance. *Trans. Am. Math. Soc.* **328**(2), 695–729 (1991)
13. Attouch, H., Wets, R.: Quantitative stability of variational systems: II, a framework for nonlinear conditioning. *SIAM J. Optim.* **3**, 359–381 (1993)
14. Bauschke, H., Combettes, P.: *Convex Analysis and Monotone Operator Theory in Hilbert spaces*, CMS Books in Mathematics. Springer, Berlin (2011)
15. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**(1), 183–202 (2009)
16. Brézis, H.: Opérateurs maximaux monotones dans les espaces de Hilbert et équations d'évolution, Lecture Notes 5, North Holland, (1972)
17. Brézis, H.: *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, Berlin (2011)
18. Bolte, J., Nguyen, T.P., Peypouquet, J., Suter, B.W.: From error bounds to the complexity of first-order descent methods for convex functions. *Math. Program.* **165**(2), 471–507 (2017)
19. Cabot, A., Engler, H., Gadat, S.: On the long time behavior of second order differential equations with asymptotically small dissipation. *Trans. Am. Math. Soc.* **361**, 5983–6017 (2009)
20. Chambolle, A., Dossal, Ch.: On the convergence of the iterates of the fast iterative shrinkage thresholding algorithm. *J. Optim. Theory Appl.* **166**, 968–982 (2015)
21. Haraux, A.: Systèmes dynamiques dissipatifs et applications, RMA 17, Masson, (1991)
22. Jendoubi, M.A., May, R.: Asymptotics for a second-order differential equation with nonautonomous damping and an integrable source term. *Appl. Anal.* **94**(2), 436–444 (2015)
23. Kim, D., Fessler, J.A.: Optimized first-order methods for smooth convex minimization. *Math. Program.* **159**(1–2), 81–107 (2016). Ser. A
24. May, R.: Asymptotic for a second order evolution equation with convex potential and vanishing damping term. *Turk. J. Math.* **41**(3), 681–685 (2017)
25. Matet, S., Rosasco, L., Villa, S., Vu, B.C.: Don't relax: early stopping for convex regularization. [arXiv:1707.05422v1](https://arxiv.org/abs/1707.05422v1) [math.OC] (2017)
26. Nesterov, Y.: A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Sov. Math. Dokl.* **27**, 372–376 (1983)
27. Nesterov, Y.: *Introductory Lectures on Convex Optimization: A Basic Course*, of Applied Optimization, vol. 87. Kluwer Academic Publishers, Boston (2004)
28. Opial, Z.: Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Am. Math. Soc.* **73**, 591–597 (1967)
29. Peypouquet, J.: *Convex Optimization in Normed Spaces: Theory, Methods and Examples*. With a Foreword by Hedy Attouch. Springer Briefs in Optimization, p. xiv+124. Springer, Cham (2015)
30. Peypouquet, J., Sorin, S.: Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time. *J. Convex Anal.* **17**(3–4), 1113–1163 (2010)

31. Polyak, B.T.: Introduction to Optimization. Optimization Software, New York (1987)
32. Robbins, H., Monro, S.: A stochastic approximation method. *Ann. Math. Stat.* **22**, 400–407 (1951)
33. Rockafellar, R.T.: Monotone operators associated with saddle-functions and mini-max problems. In: Nonlinear operators and nonlinear equations of evolution in Banach spaces 2. In: 18th Proceedings of Symposia in Pure Mathematics, F.E. Browder Ed., American Mathematical Society, pp. 241–250 (1976)
34. Rockafellar, R.T.: Augmented lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.* **1**, 97–116 (1976)
35. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis, Grundlehren der mathematischen Wissenschaften, vol. 317. Springer, Berlin (1998)
36. Schmidt, M., Le Roux, N., Bach, F.: Convergence rates of inexact proximal-gradient methods for convex optimization. In: Advances in Neural Information Processing Systems (NIPS), (2011)
37. Su, W., Boyd, S., Candès, E.J.: A differential equation for modeling Nesterov’s accelerated gradient method: theory and insights. *Neural Inf. Process. Syst.* **27**, 2510–2518 (2014)
38. Villa, S., Salzo, S., Baldassarres, L., Verri, A.: Accelerated and inexact forward–backward. *SIAM J. Optim.* **23**(3), 1607–1633 (2013)