Accelerated Proximal Algorithms with a Correction Term for Monotone Inclusions

Paul-Emile Maingé[1]

Abstract

In this paper, we propose an accelerated variant of the proximal point algorithm for computing a zero of an arbitrary maximally monotone operator A. The method incorporates an inertial term (based upon the acceleration techniques introduced by Nesterov), relaxation factors and a correction term. In a general Hilbert space setting, we obtain the weak convergence of the iterates (x_n) to equilibria, with the fast rate $\|x_{n+1} - x_n\| = o(n^{-1})$ (for the discrete velocity). In particular, when using constant proximal indexes, we establish the accuracy of (x_n) to a zero of A with the worst-case rate $\|A_\lambda(x_n)\| = o(n^{-1})$ (A_λ being the Yosida regularization of A with some index λ). Furthermore, given any positive integer p and using an appropriate adjustment of increasing proximal indexes, we obtain the worst-case convergence rate $\|A_\lambda(x_n)\| = o(n^{-(p+1)})$. This acceleration process can be applied to various proximal-type algorithms such as the augmented Lagrangian method, the alternating direction method of multipliers, operator splitting methods and so on. Our methodology relies on a Lyapunov analysis combined with a suitable reformulation of the considered algorithm. Numerical experiments are also performed.

Keywords Nesterov-type algorithm · Inertial-type algorithm · Global rate of convergence · Fast first-order method · Relaxation factors · Correction term · Accelerated proximal algorithm

✉ Paul-Emile Maingé
   Paul-Emile.Mainge@univ-antilles.fr

[1]   F.W.I., MEMIAD, Université des Antilles, Campus de Schoelcher, 97233 Schoelcher Cedex, Martinique

1 Introduction

1.1 The considered problem

Let $\mathcal{H}$ be a real Hilbert space endowed with inner product and induced norm denoted by $\langle ., . \rangle$ and $\|.\|$, respectively. Our goal is to propose and study a rapidly converging method for solving the monotone inclusion problem

$$\text{find } \bar{x} \in \mathcal{H} \text{ such that } 0 \in A\bar{x}, \tag{1.1}$$

under the following conditions:

$$A : \mathcal{H} \to 2^{\mathcal{H}} \text{ is maximally monotone on } \mathcal{H}, \tag{1.2a}$$
$$S := A^{-1}(0) \neq \emptyset. \tag{1.2b}$$

This problem finds many important applications in scientific fields such as image processing, computer vision, machine learning, signal processing, optimization, equilibrium theory, economics, game theory, partial differential equations, statistics, and so on (see, e.g., [7,9,17,31,33,42,46,47]). It includes, as special cases, variational inequalities, convex-concave saddle-point problems. In particular, we recall that (1.1)–(1.2) encompasses the non-smooth convex minimization problem

$$\min_{\mathcal{H}} g, \tag{1.3}$$

where g verify the following conditions:

$$g : \mathcal{H} \to (-\infty, \infty] \text{ is proper, convex and lower semi-continuous,} \tag{1.4a}$$
$$\text{argmin}_{\mathcal{H}} g \neq \emptyset. \tag{1.4b}$$

A typical method for computing zeroes of a maximally monotone operator $A : \mathcal{H} \to 2^{\mathcal{H}}$ is the so-called proximal point algorithm, PPA for short (see Martinet [36], Rockafellar [44,45]), which consists of the iteration

$$x_{n+1} = J_{\mu A}(x_n), \tag{1.5}$$

where $J_{\mu A} := (I + \mu A)^{-1}$ denotes the proximal mapping of A (with index μ) which is well-known to be single-valued, everywhere defined (see, e.g., [13,24,30] for more details). Many celebrated algorithms can be recast as specific cases of PPA. As examples we mention the augmented Lagrangian method [26,41,43], the alternating direction method of multipliers (ADMM for short, see [1,12,18,22]), the split inexact Uzawa method [47], as well as operator splitting methods (including the Douglas-Rachford splitting method [19,34] and its generalized variant in [18] (also inspired from the Peaceman-Racheford splitting method [40]). Thus any enhancement of PPA has a wide range of applications.

It is well-known that PPA generates weakly convergent sequences (x_n) with worst-case rates $\|x_{n+1} - x_n\| = \mathcal{O}(n^{-\frac{1}{2}})$ and $\|A_\mu(x_n)\| = \mathcal{O}(n^{-\frac{1}{2}})$ (see, e.g. [15,18,23]), where $A_\mu = \mu^{-1}(I - J_{\mu A})$ denotes the Yosida regularization of A. This latter operator enjoys numerous nice properties (see, for instance, [13,14]). In particular, it satisfies $A_\mu^{-1}(0) = S := A^{-1}(0)$ and the quantity $\|A_\mu(x)\|$ (for $x \in \mathcal{H}$) can be used to measure the accuracy of x to a zero of A (see [18]).

The minimization problem (1.3)–(1.4), as a special instance of (1.1)–(1.2) when $A = \partial g$ (where ∂g is the Fenchel sub-differential of g) can be solved by PPA in which the resolvent operator of A reduces to

$$J_{\mu \partial g} = \text{prox}_{\mu g}(x) := \text{argmin}_{y \in \mathcal{H}} \left(g(y) + (2\mu)^{-1}\|x - y\|^2 \right). \qquad (1.6)$$

Recall that the corresponding PPA has been considerably enhanced, by means of extrapolation processes based upon Nesterov's and Güler's acceleration techniques. The latter variant generates convergent sequences (x_n) with worst-case convergence rates $g(x_n) - \inf_{\mathcal{H}} g = o(n^{-2})$ (for the function values), $\|x_{n+1} - x_n\| = o(n^{-1})$ (for the discrete velocity) and $\|x_n^*\| = o(n^{-1})$ (for the sub-gradients $x_n^* \in \partial g(x_n)$), instead of the worst-case rates $g(x_n) - \inf_{\mathcal{H}} g = \mathcal{O}(n^{-1})$, $\|x_{n+1} - x_n\| = \mathcal{O}(n^{-\frac{1}{2}})$ and $\|x_n^*\| = o(n^{-\frac{1}{2}})$ established for PPA.

Our purpose here is to extend the above results to sequences (x_n) given by some accelerated variant of PPA, dedicated to solve the more general problem (1.1)–(1.2), especially in term of the quantities $\|x_{n+1} - x_n\|$ and $\|A_\lambda(x_n)\|$ (for some positive value λ).

Specifically, for solving (1.1), we introduce CRIPA (Corrected Relaxed Inertial Proximal Algorithm) which enters the following framework of sequences $\{z_n, x_n\} \subset \mathcal{H}$ generated from starting points $\{z_{-1}, x_{-1}, x_0\}$ by

$$z_n = x_n + \theta_n(x_n - x_{n-1}) + \gamma_n(z_{n-1} - x_n), \qquad (1.7a)$$
$$x_{n+1} = (1 - w_n)z_n + w_n J_{r_n A}(z_n), \qquad (1.7b)$$

where $\{w_n, \theta_n, \gamma_n\} \subset (0, 1)$ and $(r_n) \subset (0, \infty)$. These parameters will be specified farther.

The considered algorithm combines relaxation factors (w_n), a momentum term "$\theta_n(x_n - x_{n-1})$" (based on Nesterov's acceleration techniques) and a correction term "$\gamma_n(x_n - z_{n-1})$" (similar to that introduced by Kim [27] in some variant of PPA).

Compared to PPA, CRIPA keeps the computational cost at each iteration basically unchanged, while allowing us to extend to the wide framework of maximal monotone inclusions, the interesting convergence properties obtained for the accelerated variant relative to convex minimization. More precisely, using conveniently chosen parameters $\{r_n, w_n, \theta_n, \gamma_n\}$, we establish among others (see Theorem 1), the weak convergence of (x_n) towards some zero of A, with the convergence rate $\|x_{n+1} - x_n\| = o(n^{-1})$ for the discrete velocity. In addition, when using constant proximal indexes r_n (see Theorem 2), we obtain the accuracy of (x_n) to an equilibria with the worst-case rate $\|A_\lambda(x_n)\| = o(n^{-1})$ (for some positive value λ). This latter estimate is considerably improved when using increasing proximal indexes (see Theorem 3). To the best of our

knowledge, there are no such convergence results established so far for algorithmic solution to (1.1)–(1.2).

1.2 A brief review of the state of art.

1.2.1 Convex minimization and Güuler's acceleration processes.

Let us give some reminders concerning the useful and efficient methods proposed by Güler [25], based upon ideas of Nesterov [37] (also see [38,39]), for minimizing a convex lower semi-continuous function g. As particular instances of these acceleration techniques (when using a constant proximal index μ), we mention the following two algorithms:

The first consists of the sequences $\{x_n, z_n\}$ given for $n \geq 0$ by

$$x_{n+1} = J_{\mu \partial g}(z_n), \tag{1.8a}$$

$$t_{n+1} = \frac{1}{2}\left(1 + \sqrt{1 + 4t_n^2}\right), \tag{1.8b}$$

$$z_{n+1} = x_{n+1} + \frac{t_n - 1}{t_{n+1}}(x_{n+1} - x_n). \tag{1.8c}$$

The second method consists of $\{x_n, z_n\}$ given for $n \geq 0$ by

$$x_{n+1} = J_{\mu \partial g}(z_n), \tag{1.9a}$$

$$t_{n+1} = \frac{1}{2}\left(1 + \sqrt{1 + 4t_n^2}\right), \tag{1.9b}$$

$$z_{n+1} = x_{n+1} + \frac{t_n - 1}{t_{n+1}}(x_{n+1} - x_n) + \frac{t_n}{t_{n+1}}(x_{n+1} - z_n). \tag{1.9c}$$

Note that both methods include a momentum term $\frac{t_n-1}{t_{n+1}}(x_{n+1} - x_n)$ while the second one incorporates an additional correction term $\frac{t_n}{t_{n+1}}(x_{n+1} - z_n)$.

Both the methods (1.8) and (1.9) were shown to produce iterates (x_n) that guarantee a worst-case rate $g(x_n) - \inf_{\mathcal{H}} g = \mathcal{O}(n^{-2})$ for minimizing the function values. However the convergence of the iterates has not been established.

This drawback was overame by Chambolle-Dossal [16] (also see Attouch-Peypouquet [5]), through the variant of (1.8) given by

$$\begin{aligned} z_n &= x_n + \frac{n-1}{n+\alpha-1}(x_n - x_{n-1}), \\ x_{n+1} &= J_{\mu \partial g}(z_n), \end{aligned} \tag{1.10}$$

for some positive constant α. It was proved for $\alpha > 3$ (see [5]) that (1.10) generates (weakly) convergent sequences (x_n) that minimize the function values with a complexity result of $o(n^{-2})$, instead of the rates $\mathcal{O}(n^{-1})$ and $\mathcal{O}(n^{-2})$ obtained for (1.5) and Güler's processes (1.8)–(1.9), respectively.

Note that the iterates $\{x_n, z_n\}$ generated by the second model of Güler (1.9) satisfy, for $n \geq 1$,

$$z_n = x_n + \frac{t_{n-1} - 1}{t_n}(x_n - x_{n-1}) + \frac{t_{n-1}}{t_n}(x_n - z_{n-1}), \tag{1.11}$$

$$x_{n+1} = J_{\mu \partial g}(z_n). \tag{1.12}$$

Thus, the second algorithm (1.9), which is closely related to the optimized gradient methods discussed by Kim-Fessler [28,29], uses a correction term other than the one involved in (1.7).

1.2.2 Monotone inclusions and acceleration processes.

Let us emphasize that, regarding the existing algorithmic solutions to (1.1)–(1.2) with an arbitrary monotone operator, there are no analogous theoretical convergence result to that obtained for (1.10). Many papers have been dedicated to accelerating PPA in its general form, by means of relaxation and inertial techniques. It seems (to the best of our knowledge) that only empirical accelerations have been obtained, except for the recent works by Attouch-Peypouquet [5], Attouch-László [2] and Kim [27]:

An accelerated variant of PPA has been proposed and investigated by Attouch-Peypouquet [5] through the following RIPA (Regularized Inertial Proximal Algorithm)

$$z_n = x_n + \left(1 - \frac{\alpha}{n}\right)(x_n - x_{n-1}), \tag{1.13}$$

$$x_{n+1} = \frac{\lambda_n}{\lambda_n + s}z_n + \frac{s}{\lambda_n + s}J_{(\lambda_n + s)A}(z_n), \tag{1.14}$$

where $\{s, \alpha, \lambda_n\}$ are positive parameters such that

$$\alpha > 2 \text{ and } \lambda_n = (1 + \epsilon)\frac{s}{\alpha^2}n^2 \text{ (for some } \epsilon > 0). \tag{1.15}$$

It was established (see [5, Theorem 3.6]) that (1.13)–(1.15) produces convergent sequences (x_n) with worst-case rates

$$\|x_{n+1} - x_n\| = \mathcal{O}(n^{-1}) \text{ and } \|A_{\lambda_n + s}x_n\| = o(n^{-1}). \tag{1.16}$$

This algorithm was applied by Attouch [1] to convex structured minimization problems with linear constraints and it gave rise to an inertial proximal ADMM algorithm.

Another numerical approach to (1.1)–(1.2) was addressed by Attouch-László [2] through the following PRINAM (Proximal Regularized Inertial Newton Algorithm)

$$z_{n-1} = \left(1 - \beta\left(\frac{1}{\lambda_n} - \frac{1}{\lambda_{n-1}}\right)\right)x_n + \left(\alpha_n - \frac{\beta}{\lambda_{n-1}}\right)\dot{x}_n$$
$$+ \beta\left(\frac{1}{\lambda_n}J_{\lambda_n A}(x_n) - \frac{1}{\lambda_{n-1}}J_{\lambda_{n-1}A}(x_{n-1})\right), \tag{1.17a}$$

$$x_{n+1} = \frac{\lambda_{n+1}}{\lambda_{n+1} + s} z_{n-1} + \frac{s}{\lambda_{n+1} + s} J_{(\lambda_{n+1}+s)A}(z_{n-1}), \qquad (1.17b)$$

where $\dot{x}_n = x_n - x_{n-1}$, while s, β, (λ_n) and (α_n) are positive parameters that satisfy

$$\alpha_n = \frac{rn + q - 1}{r(n+1) + q} \text{ with } r > 0 \text{ and } q \in (-\infty, \infty), \lambda_n = \lambda n^2 \text{ with } \lambda > \frac{(2\beta + s)^2 r^2}{s}.$$
$$(1.18)$$

It was established (see [2, Theorem 1.1]) that (1.17a)–(1.17b) generates convergent sequences (x_n) with worst-case rates

$$\|x_{n+1} - x_n\| = \mathcal{O}(n^{-1}) \text{ and } \|A_{\lambda_n}(x_n)\| = o(n^{-2}). \qquad (1.19)$$

This second method also use a correction term " $\beta\left(\frac{1}{\lambda_n} J_{\lambda_n A}(x_n) - \frac{1}{\lambda_{n-1}} J_{\lambda_{n-1} A}(x_{n-1})\right)$ " different from the one involved in (1.7).

Note that, for accelerating the proximal point algorithm, both the previous methods require proximal indexes (λ_n) that go to infinity as $n \to \infty$.

Remark 1 This last observation is fundamental in view of decomposition techniques which require to use only constant or bounded proximal indexes; for instance, in forward-backward algorithms ([33]), or in operator splitting methods ([18,19,21]).

This is not the case for the accelerated proximal point method proposed by Kim [27], based on the performance estimation problem (PEP) approach of Drori-Teboulle [20] and which writes, for initial iterates $\{x_0, z_0, z_{-1}\} \subset \mathcal{H}$ and for $n \geq 0$,

$$x_{n+1} = J_{\mu A}(z_n), \qquad (1.20a)$$
$$z_{n+1} = x_{n+1} + \frac{n}{n+2}(x_{n+1} - x_n) + \frac{n}{n+2}(z_{n-1} - x_n). \qquad (1.20b)$$

This yields the following worst-case convergence rate $\|x_n - z_{n-1}\| = \mathcal{O}(n^{-1})$ (see [27, Theorem 4.1]), or equivalently $\|A_\mu(z_n)\| = \mathcal{O}(n^{-1})$, which entails (since A_μ is Lipschtz continuous)

$$\|A_\mu(x_n)\| = \mathcal{O}(n^{-1}). \qquad (1.21)$$

However no convergence of the iterates was established for (1.20).

1.3 CRIPA and an overview of the related results.

1.3.1 Introducing CRIPA.

Our numerical approach to (1.1)–(1.2) is more precisely given by sequences $\{x_n, z_n\}$ generated by the following (corrected, relaxed and inertial) algorithm :

(CRIPA):
▷ **Step 1** (initialization): Let $\{z_{-1}, x_{-1}, x_0\} \subset \mathcal{H}, \lambda > 0$.
▷ **Step 2** (main step): Given $\{z_{n-1}, x_{n-1}, x_n\} \subset \mathcal{H}$ (with $n \geq 0$), we compute the updates by

$$z_n = x_n + \theta_n(x_n - x_{n-1}) + \gamma_n (z_{n-1} - x_n), \tag{1.22a}$$

$$x_{n+1} = \frac{1}{1 + k_n} z_n + \frac{k_n}{1 + k_n} J_{\lambda(1+k_n)A}(z_n), \tag{1.22b}$$

where k_n, θ_n and γ_n are non-negative parameters defined, for some constant $\{a, c, a_1, a_2\} \subset [0, \infty)$ and $\{b, \bar{c}\} \subset (0, \infty)$, as follows:

- (k_n) is given recursively, from $k_0 > 0$ and for $n \geq 1$, by

$$k_n = k_{n-1} \left(1 + \frac{a}{bn + c} \right). \tag{1.23}$$

- (θ_n) and (γ_n) are given for $n \geq 0$ by

$$\theta_n = 1 - \frac{a_1}{bn + \bar{c}}, \quad \gamma_n = 1 - \frac{a_2}{bn + \bar{c}}. \tag{1.24}$$

Let us stress that, for conventional reasons, we can assume that $\bar{c} > a_i$ (for $i = 1, 2$), so as to ensure that $\{\theta_n, \gamma_n\} \subset (0, 1)$ (even though this is not of great importance for convergence).

A particular attention will be paid to the special case of CRIPA when $a = 0$ (namely when using constant relaxation factors), which will be considered through the following formulation:
(CRIPA-S):
▷ **Step 1** (initialization): Let $\{z_{-1}, x_{-1}, x_0\} \subset \mathcal{H}, \lambda > 0$.
▷ **Step 2** (main step): Given $\{z_{n-1}, x_{n-1}, x_n\} \subset \mathcal{H}$ (with $n \geq 0$), we compute the updates by

$$z_n = x_n + \theta_n(x_n - x_{n-1}) + \gamma_n(z_{n-1} - x_n), \tag{1.25a}$$

$$x_{n+1} = \frac{1}{1 + k_0} z_n + \frac{k_0}{1 + k_0} J_{\lambda(1+k_0)A}(z_n). \tag{1.25b}$$

1.3.2 An overview of the main results.

Under convenient assumptions on the constants $\{a, b, c, a_1, a_2, \bar{c}\}$ we establish (among others) the weak convergence of $\{x_n\}$ (generated by CRIPA) towards equilibria, but we also set up the following estimates (see Theorems 1, 2 and 3):

$$\|x_{n+1} - x_n\| = o(n^{-1}), \sum_n n\|x_{n+1} - x_n\|^2 < \infty, \tag{1.26a}$$

$$\sum_n n^2 \|(x_{n+1} - x_n) - (x_n - x_{n-1})\|^2 < \infty. \tag{1.26b}$$

Moreover, given any arbitrary nonnegative integer p, we exhibit a set of conditions on the parameters that ensure the following properties (see Theorem 2)

$$\|A_\lambda(x_n)\| = o(n^{-(p+1)}), \sum_n n^{2p+1} \|A_\lambda(x_n)\|^2 < \infty, \qquad (1.27a)$$

$$\text{for any } q \in A^{-1}(0), \sum_n n^p \langle A_\lambda(x_n), x_n - q \rangle < \infty. \qquad (1.27b)$$

In particular, when A is the sub-differential of a proper and convex lower semi-continuous function $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$, we reach the following convergence rates of the values (see Theorem 4)

$$f_\lambda(x_n) - \min f = o(n^{-(p+1)}), \sum_n n^p (f_\lambda(x_n) - \min f) < \infty, \qquad (1.28)$$

where f_λ denotes the Moreau envelope of f.

Our process provides a significant acceleration of PPA, even when using a constant proximal index, besides generating convergent iterates. It also involves a correction term similar to that used by Kim [27]. It would be interesting to understand the role of the correction term used in the proposed method, through a continuous counterpart, but this is out of the scope of this work. Let us mention that there is a flourishing literature devoted to accelerated continuous counterpart of PPA with general operators (see, for instance, [2,3,5,8,11,35]).

1.4 Organization of the paper.

In Sect. 2, we give some preliminaries on CRIPA. The method is reformulated in terms of Yosida approximations and a crucial estimation is proposed for a Lyapunov analysis. Section 3 is devoted to the convergence analysis of CRIPA. A main result (Theorem 1) is established in a general setting of operators and parameters. As consequences of our main theorem, we claim two other results (Theorems 2 and 3) relative to the involved parameters. In Sect. 4, we specialize our main results in the setting of convex minimization. Two related results are proposed (Theorems 4 and 5) are set up. Next, numerical experiments are performed in Sect. 5 and several technical results are established in Appendix.

Remark 2 From now on, so as to simplify the presentation, we (often) use the following notation: given any sequence (u_n), we denote $\dot{u}_n = u_n - u_{n-1}$. We also denote the integer part of any real number a by $[a]$.

2 Preliminaries on CRIPA.

In this section, we provide a suitable reformulation of CRIPA in terms of Yosida approximations and we exhibit a Lyapunov sequence in connection with the algorithm.

These two arguments will allow us to obtain separately a series of preliminary estimates (in Sect. 3) that will be combined so as to reach our statements in Theorem 1.

2.1 Formulation of CRIPA by means of Yosida approximations.

Let sequences $\{x_n, z_n\}$ verify (1.22b) relative to some positive parameters $\{k_n, \lambda\}$ and some maximally monotone operator $A : \mathcal{H} \to 2^{\mathcal{H}}$.

By definition of the Yosida approximation we have $A_\lambda = \lambda^{-1}(I - J_{\lambda A})$ as well as

$$(I + \lambda k_n A_\lambda)^{-1} = I - \lambda k_n (A_\lambda)_{\lambda k_n}. \tag{2.1}$$

Hence, according to the resolvent equation (formulated as a semi-group property) $(A_\delta)_\kappa = A_{\delta+\kappa}$ (for any positive values δ and κ), we infer that

$$(I + \lambda k_n A_\lambda)^{-1} = I - \lambda k_n A_{\lambda(1+k_n)}. \tag{2.2}$$

So, by $x_{n+1} = z_n - \lambda k_n A_{\lambda(1+k_n)}(z_n)$ (from (1.22b)), we deduce that

$$x_{n+1} = (I + \lambda k_n A_\lambda)^{-1}(z_n). \tag{2.3}$$

This observation will be of crucial interest with regard to the methodology developed farther.

As a key result in our analysis, we give the following lemma.

Lemma 1 *The iterates $\{x_n, z_n\}$ generated by CRIPA satisfy (for $n \geq 0$)*

$$z_n - x_{n+1} = \lambda k_n A_\lambda(x_{n+1}), \tag{2.4a}$$

$$\dot{x}_{n+1} + (z_n - x_{n+1}) = \theta_n \dot{x}_n + \gamma_n(z_{n-1} - x_n). \tag{2.4b}$$

Proof It is no difficult to see for $n \geq 0$ that (2.3) is equivalent to $z_n = x_{n+1} + \lambda k_n A_\lambda(x_{n+1})$, which yields (2.4a). Furthermore, by $z_n = x_n + \theta_n \dot{x}_n + \gamma_n(z_{n-1} - x_n)$ (from (1.22a)), we simply obtain $z_n - x_{n+1} = -\dot{x}_{n+1} + \theta_n \dot{x}_n + \gamma_n(z_{n-1} - x_n)$, which entails (2.4b) □

As a consequence of the previous arguments we make the following observation.

Remark 3 The iterates $\{x_n, z_n\}$ given by CRIPA satisfy (for $n \geq 1$)

$$z_n = x_n + \theta_n(x_n - x_{n-1}) + \gamma_n \lambda k_{n-1} A_\lambda(x_n), \tag{2.5a}$$

$$x_{n+1} = J_{\lambda k_n A_\lambda}(z_n). \tag{2.5b}$$

This latter formulation enlightens the fact that the Moreau-Yosida regularization compatibilizes the operator's lack of co-coercivity with the acceleration scheme (see [5,6]).

2.2 A general inequality for a Lyapunov analysis

At once we provide a useful inequality dedicated to a Lyapunov analysis of the considered method. With the iterates $\{x_n\}$ produced by CRIPA, we associate the sequence $(\mathcal{E}_n(s, q))$ defined for $(s, q) \in (0, \infty) \times S$ and for $n \geq 1$ by

$$
\begin{aligned}
\mathcal{E}_n(s, q) = {} & \tfrac{1}{2}\|s(q - x_n) - (bn + \bar{c} - a_1)\dot{x}_n\|^2 \\
& + \tfrac{1}{2}s(a_1 - b - s)\|x_n - q\|^2 + s\lambda k_{n-1}(b(n - 1) + \bar{c})\langle A_\lambda(x_n), x_n - q\rangle.
\end{aligned} \tag{2.6}
$$

The next result will be helpful with regard to our forthcoming analysis.

Lemma 2 *Suppose that (1.2) holds and let $\{x_n\} \subset \mathcal{H}$ be generated by CRIPA with sequences (k_n), (θ_n) and (γ_n) (given by (1.23) and (1.24)), along with constants $\{\lambda, k_0\} \subset (0, \infty)$, $\{a, c, a_1, a_2\} \subset [0, \infty)$ and $\{b, \bar{c}\} \subset (0, \infty)$ verifying*

$$
a_1 > b. \tag{2.7}
$$

Then, for $(s, q) \in (0, a_1 - b] \times \mathcal{H}$ and for $n \geq N_0$ (with some N_0 large enough), we have

$$
\begin{aligned}
& \dot{\mathcal{E}}_{n+1}(s, q) + s\lambda k_{n-1}(a_2 - b)\langle A_\lambda(x_n), x_n - q\rangle \\
& \quad + \lambda k_{n-1}(bn + \bar{c})\left(a\tfrac{bn+\bar{c}-s}{bn+c} + a_2 - s\right)\langle A_\lambda(x_n), \dot{x}_{n+1}\rangle \\
& \quad + \lambda^2 k_n(bn + \bar{c})(bn + \bar{c} - s)\|A_\lambda(x_{n+1}) - A_\lambda(x_n)\|^2 \\
& \quad + \tfrac{1}{2}(bn + \bar{c})^2\|\dot{x}_{n+1} - \theta_n\dot{x}_n\|^2 \\
& \quad + \tfrac{1}{2}(a_1 - b - s)\left((2n + 1)b + 2\bar{c} - a_1\right)\|\dot{x}_{n+1}\|^2 \leq 0.
\end{aligned} \tag{2.8}
$$

The proof of Lemma 2 is divided into two parts through the next subsections.

2.2.1 Proof of Lemma 2 - part 1.

An important equality of independent interest is proposed here relative to our method through the wider framework of sequences $\{x_n, d_n\} \subset \mathcal{H}$ and parameters $\{e, v_n, \theta_n\} \subset (0, \infty)$ verifying (for $n \geq 0$)

$$
\dot{x}_{n+1} - \theta_n\dot{x}_n + d_n = 0, \tag{2.9a}
$$

$$
(e + v_{n+1})\theta_n = v_n. \tag{2.9b}
$$

For this purpose, we associate with (2.9) the quantity $G_n(s, q)$ given for $(s, q) \in [0, \infty) \times \mathcal{H}$ by

$$
G_n(s, q) = \frac{1}{2}\|s(q - x_n) - v_n\dot{x}_n\|^2 + \frac{1}{2}s(e - s)\|x_n - q\|^2, \tag{2.10}
$$

Basic properties regarding the sequence $(G_n(s, q))$ are established through the following proposition.

Proposition 1 *Let* $\{x_n, d_n\} \subset \mathcal{H}$ *and* $\{\theta_n, v_n, e\} \subset (0, \infty)$ *verify* (2.9). *Then for* $(s, q) \in (0, e] \times \mathcal{H}$ *and for* $n \geq 0$ *we have*

$$
\begin{aligned}
\dot{G}_{n+1}(s, q) &+ \tfrac{1}{2}(e + v_{n+1})^2 \|\dot{x}_{n+1} - \theta_n \dot{x}_n\|^2 \\
&+ s(e + v_{n+1})\langle d_n, x_{n+1} - q \rangle \\
&+ (e - s + v_{n+1})(e + v_{n+1})\langle d_n, \dot{x}_{n+1} \rangle = -\tfrac{1}{2}(e - s)(e + 2v_{n+1})\|\dot{x}_{n+1}\|^2.
\end{aligned}
\tag{2.11}
$$

The proof of Proposition 1 is given in Appendix 1.

2.2.2 Proof of Lemma 2 - part 2.

It can be checked (from (2.4)) that CRIPA enters the special case of algorithm (2.9) (for $n \geq 1$) when taking the particular parameters $e = a_1 - b$ (with $a_1 > b$), $v_n = bn + \bar{c} - a_1$, together with d_n as below

$$
d_n = \lambda k_n A_\lambda(x_{n+1}) - \gamma_n \lambda k_{n-1} A_\lambda(x_n).
\tag{2.12}
$$

In this specific situation, we get

$$
e + v_{n+1} = bn + \bar{c} \quad \text{and} \quad e + 2v_{n+1} = (2n + 1)b + 2\bar{c} - a_1.
\tag{2.13}
$$

Hence, for $n \geq N_0$, for some N_0 large enough (so as to ensure that v_n is positive), and for $s \in (0, a_1 - b]$, by Proposition 1 we obtain

$$
\begin{aligned}
\dot{G}_{n+1}(s, q) &+ \tfrac{1}{2}(bn + \bar{c})^2 \|\dot{x}_{n+1} - \theta_n \dot{x}_n\|^2 + \Gamma_n \\
&= -\tfrac{1}{2}(a_1 - b - s)((2n + 1)b + 2\bar{c} - a_1)\|\dot{x}_{n+1}\|^2,
\end{aligned}
\tag{2.14}
$$

where $G_n(s, q)$ and Γ_n are given by

$$
G_n(s, q) = \frac{1}{2}\|s(q - x_n) - (bn + \bar{c} - a_1)\dot{x}_n\|^2 + \frac{1}{2}s(a_1 - b - s)\|x_n - q\|^2,
\tag{2.15a}
$$

$$
\Gamma_n = s(bn + \bar{c})\langle d_n, x_{n+1} - q \rangle + (bn + \bar{c})(bn + \bar{c} - s)\langle d_n, \dot{x}_{n+1} \rangle.
\tag{2.15b}
$$

In order to evaluate Γ_n and for simplification, we set

$$
U_n := \langle \lambda k_{n-1} A_\lambda(x_n), x_n - q \rangle.
$$

According to the formulation of d_n (given in (2.12)) we have

$$
\begin{aligned}
&\langle d_n, x_{n+1} - q \rangle \\
&= \Big\langle \lambda k_n A_\lambda(x_{n+1}) - \gamma_n \lambda k_{n-1} A_\lambda(x_n), x_{n+1} - q \Big\rangle \\
&= \Big\langle \lambda k_n A_\lambda(x_{n+1}), x_{n+1} - q \Big\rangle - \gamma_n \Big\langle \lambda k_{n-1} A_\lambda(x_n), \dot{x}_{n+1} \rangle - \gamma_n \langle \lambda k_{n-1} A_\lambda(x_n), x_n - q \Big\rangle \\
&= U_{n+1} - \gamma_n U_n - \gamma_n \langle \lambda k_{n-1} A_\lambda(x_n), \dot{x}_{n+1} \rangle,
\end{aligned}
$$

as well as

$$
\begin{aligned}
&\langle d_n, \dot{x}_{n+1} \rangle \\
&= \left\langle \lambda k_n A_\lambda(x_{n+1}) - \gamma_n \lambda k_{n-1} A_\lambda(x_n), \dot{x}_{n+1} \right\rangle \\
&= \left\langle \lambda k_n A_\lambda(x_{n+1}) - \lambda k_{n-1} A_\lambda(x_n) + (1 - \gamma_n)\lambda k_{n-1} A_\lambda(x_n), \dot{x}_{n+1} \right\rangle \\
&= \left\langle \lambda k_n A_\lambda(x_{n+1}) - \lambda k_{n-1} A_\lambda(x_n), \dot{x}_{n+1} \right\rangle + (1 - \gamma_n)\langle \lambda k_{n-1} A_\lambda(x_n), \dot{x}_{n+1} \rangle \\
&= \lambda k_n \langle A_\lambda(x_{n+1}) - A_\lambda(x_n), \dot{x}_{n+1} \rangle \\
&\qquad + \lambda (k_n - k_{n-1}) \langle A_\lambda(x_n), \dot{x}_{n+1} \rangle + (1 - \gamma_n)\langle \lambda k_{n-1} A_\lambda(x_n), \dot{x}_{n+1} \rangle \\
&= (\lambda k_n)\langle A_\lambda(x_{n+1}) - A_\lambda(x_n), \dot{x}_{n+1} \rangle \\
&\qquad + \left(\frac{k_n - k_{n-1}}{k_{n-1}} + (1 - \gamma_n) \right) \langle \lambda k_{n-1} A_\lambda(x_n), \dot{x}_{n+1} \rangle.
\end{aligned}
$$

This in light of (2.15b) amounts to

$$
(bn + \bar{c})^{-1} \Gamma_n = s\langle d_n, x_{n+1} - q \rangle + (bn + \bar{c} - s)\langle d_n, \dot{x}_{n+1} \rangle
$$

$$
\begin{aligned}
&= s\left(U_{n+1} - \gamma_n U_n - \gamma_n \langle \lambda k_{n-1} A_\lambda(x_n), \dot{x}_{n+1} \rangle \right) \\
&\quad + (\lambda k_n)(bn + \bar{c} - s)\langle A_\lambda(x_{n+1}) - A_\lambda(x_n), \dot{x}_{n+1} \rangle \\
&\quad + (bn + \bar{c} - s)\left(\frac{k_n - k_{n-1}}{k_{n-1}} + 1 - \gamma_n \right) \langle \lambda k_{n-1} A_\lambda(x_n), \dot{x}_{n+1} \rangle
\end{aligned}
$$

$$
\begin{aligned}
&= s\,(U_{n+1} - \gamma_n U_n) \\
&\quad + (\lambda k_n)(bn + \bar{c} - s)\langle A_\lambda(x_{n+1}) - A_\lambda(x_n), \dot{x}_{n+1} \rangle \\
&\quad + \left((bn + \bar{c} - s)\frac{k_n - k_{n-1}}{k_{n-1}} \quad + (bn + \bar{c} - s)(1 - \gamma_n) - s\gamma_n \right) \langle \lambda k_{n-1} A_\lambda(x_n), \dot{x}_{n+1} \rangle.
\end{aligned}
$$

The latter equality can be rewritten as

$$
\begin{aligned}
(bn + \bar{c})^{-1} \Gamma_n &= s(U_{n+1} - \gamma_n U_n) \\
&\quad + \lambda k_n(bn + \bar{c} - s)\langle A_\lambda(x_{n+1}) - A_\lambda(x_n), \dot{x}_{n+1} \rangle \\
&\quad + \left((bn + \bar{c} - s)\frac{k_n - k_{n-1}}{k_{n-1}} + (bn + \bar{c})(1 - \gamma_n) - s \right) \langle \lambda k_{n-1} A_\lambda(x_n), \dot{x}_{n+1} \rangle.
\end{aligned} \tag{2.16}
$$

Moreover, we recall that $\gamma_n = 1 - \frac{a_2}{n + \bar{c}}$. Then, by an easy computation, we obtain the following two equalities

$$
\begin{aligned}
\gamma_n(bn + \bar{c}) &= (bn + \bar{c}) - a_2 = (b(n - 1) + \bar{c}) + b - a_2, \\
(bn + \bar{c})(1 - \gamma_n) - s &= a_2 - s.
\end{aligned}
$$

It is also readily checked from (1.24) that k_n satisfies

$$
\frac{k_n - k_{n-1}}{k_{n-1}} = \frac{a}{bn + c}. \tag{2.17}
$$

Therefore by the previous arguments we get

$$\Gamma_n = s(bn + \bar{c})U_{n+1} - s(b(n-1) + \bar{c})U_n + s(a_2 - b)U_n$$
$$+ (bn + \bar{c})\left(a\frac{bn + \bar{c} - s}{bn + c} + a_2 - s\right)\langle\lambda k_{n-1}A_\lambda(x_n), \dot{x}_{n+1}\rangle \quad (2.18)$$
$$+ \lambda k_n(n + \bar{c})(n + \bar{c} - s)\langle A_\lambda(x_{n+1}) - A_\lambda(x_n), \dot{x}_{n+1}\rangle.$$

Consequently, by noticing that $\mathcal{E}_n(s, q) = G_n(s, q) + s(b(n-1) + \bar{c})U_n$ (in light (2.6) and (2.15a)), and using (2.14) and (2.18), we deduce

$$\dot{\mathcal{E}}_{n+1}(s, q) + s(a_2 - b)U_n + (bn + \bar{c})\left(a\frac{bn + \bar{c} - s}{bn + c} + a_2 - s\right)\langle\lambda k_{n-1}A_\lambda(x_n), \dot{x}_{n+1}\rangle$$
$$+ \lambda k_n(bn + \bar{c})(bn + \bar{c} - s)\langle A_\lambda(x_{n+1}) - A_\lambda(x_n), \dot{x}_{n+1}\rangle$$
$$+ \frac{1}{2}(bn + \bar{c})^2\|\dot{x}_{n+1} - \theta_n\dot{x}_n\|^2$$
$$= -\frac{1}{2}(a_1 - b - s)\left((2n+1)b + 2\bar{c} - a_1\right)\|\dot{x}_{n+1}\|^2.$$

In addition, the well-known property of λ-co-coerciveness of A_λ implies that

$$\langle A_\lambda(x_{n+1}) - A_\lambda(x_n), \dot{x}_{n+1}\rangle \geq \lambda\|A_\lambda(x_{n+1}) - A_\lambda(x_n)\|^2. \quad (2.19)$$

Thus, for $n \geq N_0$, for some N_0 large enough (which also ensures that $(bn + \bar{c} - (a_1 - b))$ is positive), and for $s \in (0, a_1 - b]$, by the previous two inequalities we are led to

$$\dot{\mathcal{E}}_{n+1}(s, q) + s(a_2 - b)U_n + (bn + \bar{c})\left(a\frac{bn + \bar{c} - s}{bn + c} + a_2 - s\right)\langle\lambda k_{n-1}A_\lambda(x_n), \dot{x}_{n+1}\rangle$$
$$+ \lambda^2 k_n(bn + \bar{c})(bn + \bar{c} - s)\|A_\lambda(x_{n+1}) - A_\lambda(x_n)\|^2$$
$$+ \frac{1}{2}(bn + \bar{c})^2\|\dot{x}_{n+1} - \theta_n\dot{x}_n\|^2$$
$$+ \frac{1}{2}(a_1 - b - s)\left((2n+1)b + 2\bar{c} - a_1\right)\|\dot{x}_{n+1}\|^2 \leq 0.$$

This last inequality, recalling that $U_n := \langle\lambda k_{n-1}A_\lambda(x_n), x_n - q\rangle$, is nothing but (2.8)
□

3 CRIPA in the general case of monotone operators.

3.1 Main estimates.

A series of estimates are obtained here by means of a Lyapunov analysis (based upon Lemma 2) and using the reformulation of CRIPA (from Lemma 1). Our main results (in Theorem 1) will be derived as a combination of the previous series of estimates.

3.1.1 Estimates from the energy-like sequence.

The next result is obtained from Lyapunov properties of $(\mathcal{E}_n(s, q))$ for convenient choices of the involved parameters.

Lemma 3 *Suppose that (1.2) holds and that $\{x_n\} \subset \mathcal{H}$ is generated by CRIPA with sequences (k_n), (θ_n) and (γ_n) (given by (1.23) and (1.24)), along with constants $\{\lambda, k_0\} \subset (0, \infty)$, $\{a, c, a_1, a_2, \} \subset [0, \infty)$ and $\{b, \bar{c}\} \subset (0, \infty)$ verifying*

$$a_2 \geq b \quad \text{and} \quad a_1 > b + (a + a_2). \tag{3.1}$$

Assume, in addition, that c and $\bar{c}$ are chosen as follows:

$$\text{if } a > 0, \text{ then } c > a_1 - (a_2 + a), \bar{c} = c + (a_2 + a); \tag{3.2a}$$
$$\text{if } a = 0, \text{ then } \bar{c} > \max\{a_1, a_2\}. \tag{3.2b}$$

Then, for any $q \in S$, the sequence $(\mathcal{E}_n(a + a_2, q))_{n \geq N_1}$ (for some integer N_1 large enough) is non-increasing and convergent. Moreover, the following estimates are reached:

$$\sup_{n \geq N_1} \|x_n - q\|^2 \leq \frac{2\mathcal{E}_{N_1}(a + a_2, q)}{(a + a_2)(a_1 - b - (a + a_2))}, \tag{3.3a}$$

$$\sup_{n \geq N_1} (b(n - 1) + \bar{c})k_{n-1}\langle A_\lambda(x_n), x_n - q\rangle \leq \frac{\mathcal{E}_{N_1}(a + a_2, q)}{\lambda(a + a_2)}, \tag{3.3b}$$

$$\sup_{n} n\|\dot{x}_n\| < \infty, \tag{3.3c}$$

$$\sum_{n \geq N_1} (a_2 - b)k_{n-1}\langle A_\lambda(x_n), x_n - q\rangle \leq \frac{\mathcal{E}_{N_1}(a + a_2, q)}{\lambda(a + a_2)}, \tag{3.3d}$$

$$\sum_{n \geq N_1} k_n(bn + \bar{c})\big(bn + \bar{c} - (a + a_2)\big)\|A_\lambda(x_{n+1}) - A_\lambda(x_n)\|^2 \leq \frac{\mathcal{E}_{N_1}(a + a_2, q)}{\lambda^2},$$
$$\tag{3.3e}$$

$$\sum_{n \geq N_1} (bn + \bar{c})\|\dot{x}_{n+1}\|^2 \leq \frac{2\mathcal{E}_{N_1}(a + a_2, q)}{a_1 - b - (a + a_2)}, \tag{3.3f}$$

$$\sum_{n} n^2\|\dot{x}_{n+1} - \dot{x}_n\|^2 < \infty. \tag{3.3g}$$

Proof Clearly, we have $a + a_2 \in (0, a_1 - b)$ (by condition (3.1)). It can also be checked that condition (3.2) ensures that

$$a\frac{bn + \bar{c} - (a + a_2)}{bn + c} + a_2 = a + a_2. \tag{3.4}$$

Consequently, for $q \in S$ and using Lemma 2 with $s = a + a_2$, we know that, for $n \geq N_1$ (with N_1 large enough), we get

$$
\begin{aligned}
&\dot{\mathcal{E}}_{n+1}(a + a_2, q) + \lambda(a + a_2)k_{n-1}(a_2 - b)\langle A_\lambda(x_n), x_n - q\rangle \\
&\quad + \lambda^2 k_n(bn + \bar{c})(bn + \bar{c} - (a + a_2))\|A_\lambda(x_{n+1}) - A_\lambda(x_n)\|^2 \\
&\quad + \tfrac{1}{2}(bn + \bar{c})^2\|\dot{x}_{n+1} - \theta_n \dot{x}_n\|^2 \\
&\quad + \tfrac{1}{2}(a_1 - b - (a + a_2))\big((2n+1)b + 2\bar{c} - a_1\big)\|\dot{x}_{n+1}\|^2 \leq 0,
\end{aligned} \tag{3.5}
$$

together with

$$
bn + \bar{c} - (a + a_2) > 0, \tag{3.6a}
$$

$$
(2n+1)b + 2\bar{c} - a_1 > 0. \tag{3.6b}
$$

It follows immediately that the non-negative sequence $(\mathcal{E}_n(a + a_2, q))_{n \geq N_1}$ is non-increasing, since $a_2 - b$ is assumed to be non-negative and since $a_1 - b - (a + a_2)$ is assumed to be positive (in light of condition (3.1)). Hence, $(\mathcal{E}_n(a + a_2, q))_{n \geq N_1}$ is convergent and bounded. Moreover, from (2.6), we recall that

$$
\begin{aligned}
\mathcal{E}_n(a + a_2, q) &= \tfrac{1}{2}(a + a_2)\big(a_1 - b - (a + a_2)\big)\|x_n - q\|^2 \\
&\quad + \lambda(a + a_2)k_{n-1}(b(n-1) + \bar{c})\langle A_\lambda(x_n), x_n - q\rangle \\
&\quad + \tfrac{1}{2}\|(a + a_2)(q - x_n) - (bn + \bar{c} - a_1)\dot{x}_n\|^2.
\end{aligned} \tag{3.7}
$$

Then, by the inequality $\mathcal{E}_n(a + a_2, q) \leq \mathcal{E}_{N_1}(a + a_2, q)$ (for $n \geq N_1$), we get

$$
\tfrac{1}{2}(a + a_2)\big(a_1 - b - (a + a_2)\big)\|x_n - q\|^2 \leq \mathcal{E}_{N_1}(a + a_2, q), \tag{3.8}
$$

$$
\lambda(a + a_2)k_{n-1}(b(n-1) + \bar{c})\langle A_\lambda(x_n), x_n - q\rangle \leq \mathcal{E}_{N_1}(a + a_2, q), \tag{3.9}
$$

$$
(bn + \bar{c} - a_1)\|\dot{x}_n\| - (a + a_2)\|q - x_n\| \leq \sqrt{2\mathcal{E}_{N_1}(a + a_2, q)}. \tag{3.10}
$$

Estimates (3.3a), (3.3b) and (3.3c) are direct consequences of these last three inequalities. Furthermore, by adding (3.5) from $n = N_1$ to $n = N$ (for any given positive integer $N \geq N_1$) we obtain

$$
\begin{aligned}
&\mathcal{E}_{N+1}(a + a_2, q) \\
&\quad + \lambda(a + a_2)(a_2 - b)\sum_{n=N_1}^{N} k_{n-1}\langle A_\lambda(x_n), x_n - q\rangle \\
&\quad + \lambda^2 \sum_{n=N_1}^{N} k_n(bn + \bar{c})\big(bn + \bar{c} - (a + a_2)\big)\|A_\lambda(x_{n+1}) - A_\lambda(x_n)\|^2 \\
&\quad + \tfrac{1}{2}(a_1 - b - (a + a_2))\sum_{n=N_1}^{N}\big((2n+1)b + 2\bar{c} - a_1\big)\|\dot{x}_{n+1}\|^2 \\
&\quad + \tfrac{1}{2}\sum_{n=N_1}^{N}(n + \bar{c})^2\|\dot{x}_{n+1} - \theta_n \dot{x}_n\|^2 \leq \mathcal{E}_{N_1}(a + a_2, q),
\end{aligned} \tag{3.11}
$$

which, in light of (3.1) and (3.6), entails that

$$
\lambda(a_2 + a)(a_2 - b)\sum_{n=N_1}^{N} k_{n-1}\langle A_\lambda(x_n), x_n - q\rangle \leq \mathcal{E}_{N_1}(a_2 + a, q), \tag{3.12a}
$$

$$\lambda^2 \sum_{n=N_1}^{N} k_n (bn + \bar{c})\big(bn + \bar{c} - (a + a_2)\big) \|A_\lambda(x_{n+1}) - A_\lambda(x_n)\|^2 \le \mathcal{E}_{N_1}(a_2 + a, q),$$

(3.12b)

$$\frac{1}{2}(a_1 - b - (a + a_2)) \sum_{n=N_1}^{N} \big((2n+1)b + 2\bar{c} - a_1\big)\|\dot{x}_{n+1}\|^2 \le \mathcal{E}_{N_1}(a_2 + a, q),$$

(3.12c)

$$\frac{1}{2} \sum_{n=N_1}^{N} (n + \bar{c})^2 \|\dot{x}_{n+1} - \theta_n \dot{x}_n\|^2 \le \mathcal{E}_{N_1}(a_2 + a, q).$$

(3.12d)

This straightforwardly yields (3.3d), (3.3e) and (3.3f). The last estimate (3.3g) is simply deduced from (3.3f) and (3.12d) (in light of the definition of θ_n) $\square$

3.1.2 Estimates from the reformulation of the method.

In this section we establish additional estimates regarding CRIPA, especially on $(A_\lambda(x_n))$, by combining the results of Lemma 3 with the formulation of the method given in (2.4b).

Lemma 4 *Assume, in addition to the assumptions of Lemma 3, that a_2 and b satisfy*

$$a_2 > 2b.$$

(3.13)

Then we have the following results:

$$\sum_n n \|\dot{x}_n + \lambda k_{n-1} A_\lambda(x_n)\|^2 < \infty,$$

(3.14a)

$$\|\dot{x}_n + \lambda k_{n-1} A_\lambda(x_n)\| = o(n^{-1}),$$

(3.14b)

$$\sum_n n k_{n-1}^2 \|A_\lambda(x_n)\|^2 < \infty,$$

(3.14c)

$$\sum_n n k_{n-1} |\langle A_\lambda(x_n), \dot{x}_{n+1}\rangle| < \infty.$$

(3.14d)

Proof Let us prove (3.14a) and (3.14b). For $n \ge 1$, according to Lemma 1 we have

$$\dot{x}_{n+1} + \lambda k_n A_\lambda(x_{n+1}) = \theta_n \dot{x}_n + \gamma_n \lambda k_{n-1} A_\lambda(x_n),$$

(3.15)

where $\{\theta_n, \gamma_n\} \subset (0, 1)$ are given by (1.24) with $\bar{c} = c + a_2 + a$ (under the assumptions of Lemma 3), namely

$$\theta_n = 1 - a_1(bn + \bar{c})^{-1} \text{ and } \gamma_n = 1 - a_2(bn + \bar{c})^{-1}.$$

(3.16)

Observe that (3.15) can be rewritten as

$$\dot{x}_{n+1} + \lambda k_n A_\lambda(x_{n+1}) = \gamma_n(\dot{x}_n + \lambda k_{n-1} A_\lambda(x_n)) + (\theta_n - \gamma_n)\dot{x}_n, \qquad (3.17)$$

so, by setting

$$H_n = \dot{x}_n + \gamma_n \lambda k_{n-1} A_\lambda(x_n), \qquad (3.18)$$

we can equivalently formulate (3.17) as

$$H_{n+1} = \gamma_n H_n + (\theta_n - \gamma_n)\dot{x}_n = \gamma_n H_n + (1 - \gamma_n)\tfrac{\theta_n - \gamma_n}{1 - \gamma_n}\dot{x}_n. \qquad (3.19)$$

Then by convexity of the squared norm we infer that

$$\begin{aligned}
\|H_{n+1}\|^2 &\leq \gamma_n \|H_n\|^2 + (1 - \gamma_n)\left(\frac{\gamma_n - \theta_n}{1 - \gamma_n}\right)^2 \|\dot{x}_n\|^2 \\
&= \gamma_n \|H_n\|^2 + \frac{(\gamma_n - \theta_n)^2}{1 - \gamma_n}\|\dot{x}_n\|^2.
\end{aligned} \qquad (3.20)$$

Moreover, by (3.16) we readily have

$$\theta_n - \gamma_n = (a_2 - a_1)(bn + \bar{c})^{-1} \text{ and } 1 - \gamma_n = a_2(bn + \bar{c})^{-1}, \qquad (3.21)$$

which amounts to

$$\frac{(\gamma_n - \theta_n)^2}{1 - \gamma_n} = \frac{(a_2 - a_1)^2(n + \bar{c})^{-2}}{a_2(n + \bar{c})^{-1}} = \frac{(a_2 - a_1)^2}{a_2}(bn + \bar{c})^{-1}. \qquad (3.22)$$

Consequently, in light of (3.16), (3.20) and (3.22), we obtain

$$\|H_{n+1}\|^2 \leq \left(1 - a_2(bn + \bar{c})^{-1}\right)\|H_n\|^2 + a_2^{-1}(a_2 - a_1)^2(bn + \bar{c})^{-1}\|\dot{x}_n\|^2. \qquad (3.23)$$

Next, multiplying this last inequality by $(bn + \bar{c})^2$ gives us

$$\begin{aligned}
(bn + \bar{c})^2\|H_{n+1}\|^2 \leq\ &(bn + \bar{c})(bn + \bar{c} - a_2)\|H_n\|^2 \\
&+ a_2^{-1}(a_2 - a_1)^2(bn + \bar{c})\|\dot{x}_n\|^2,
\end{aligned} \qquad (3.24)$$

while we simply have

$$\begin{aligned}
&(bn + \bar{c})(bn + \bar{c} - a_2) - (b(n-1) + \bar{c})^2 \\
&= (bn + \bar{c})^2 - (b(n-1) + \bar{c})^2 - a_2(bn + \bar{c}) \\
&\leq 2b(bn + \bar{c}) - a_2(bn + \bar{c}) = -(a_2 - 2b)(bn + \bar{c}),
\end{aligned}$$

or equivalently

$$(bn + \bar{c})\,(bn + \bar{c} - a_2) \leq (b(n-1) + \bar{c})^2 - (a_2 - 2b)(bn + \bar{c}). \qquad (3.25)$$

It follows from the two estimates (3.24) and (3.25) that

$$\begin{aligned}
&(bn + \bar{c})^2 \|H_{n+1}\|^2 - (b(n-1) + \bar{c})^2 \|H_n\|^2 \\
&+ (bn + \bar{c})(a_2 - 2b)\|H_n\|^2 \leq a_2^{-1}(a_2 - a_1)^2 (bn + \bar{c})\|\dot{x}_n\|^2.
\end{aligned} \qquad (3.26)$$

Thus, assuming that $a_2 - 2b > 0$ and recalling that $\sum_n n\|\dot{x}_n\|^2 < \infty$ (according to Lemma 3), we classically deduce that $\sum_n n\|H_n\|^2 < \infty$ (that is (3.14a)) and that there exists $l_1 \geq 0$ such that

$$\lim_{n \to +\infty} (b(n-1) + \bar{c})^2 \|H_n\|^2 = l_1.$$

Notice that we clearly have $\lim_{n \to \infty} (bn)^2 \|H_n\|^2 = l_1$ (since $\frac{(b(n-1)+\bar{c})^2}{(bn)^2} \to 1$ as $n \to \infty$). So, by $\sum_n n\|H_n\|^2 < \infty$ (in light of (3.14a)), and noticing that $\sum_n n^{-1} = \infty$, we deduce that $l_1 = 0$, which leads to (3.14b).

Let us prove (3.14c) and (3.14d). Clearly, according to the definition of H_n, we simply have

$$n\|\lambda k_{n-1} A_\lambda(x_n)\|^2 \leq 2n\|\dot{x}_n + \lambda k_{n-1} A_\lambda(x_n)\|^2 + 2n\|\dot{x}_n\|^2 = 2n\|H_n\|^2 + 2n\|\dot{x}_n\|^2,$$

hence, by $\sum_n n\|\dot{x}_n\|^2 < \infty$ (from (3.3f)) and $\sum_n n\|H_n\|^2 < \infty$ (from (3.14a)), we immediately obtain (3.14c). In addition, Young's inequality readily gives us

$$n k_{n-1} |\langle A_\lambda(x_n), \dot{x}_{n+1} \rangle| \leq \tfrac{1}{2} n \|k_{n-1} A_\lambda(x_n)\|^2 + \tfrac{1}{2} n \|\dot{x}_{n+1}\|^2. \qquad (3.27)$$

The estimate (3.14d) is then obtained as an immediate consequence of (3.27), along with the results $\sum_n n\|k_{n-1} A_\lambda(x_n)\|^2 < \infty$ (from (3.14c)) and $\sum_n n\|\dot{x}_n\|^2 < \infty$ (from (3.3f)) $\qquad \square$

3.2 Asymptotic convergence and main results.

3.2.1 Convergence in the general case of parameters.

The following result establishes the convergence of CRIPA in a general setting of parameters.

Theorem 1 *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator such that $S := A^{-1}(0) \neq \emptyset$. Let $\{\lambda, k_0\}$ be positive constants and assume that $\{z_n, x_n\} \subset \mathcal{H}$ are generated by CRIPA with (k_n), (θ_n) and (γ_n) (given by (1.23) and (1.24)), along with constants $\{a, c, a_1, a_2\} \subset [0, \infty)$ and $\{b, \bar{c}\} \subset (0, \infty)$ verifying*

$$a_2 > 2b, \quad a_1 > b + a + a_2; \qquad (3.28a)$$

$$\text{if } a > 0, \text{ then } c > a_1 - (a_2 + a), \bar{c} = c + a_2 + a; \tag{3.28b}$$

$$\text{if } a = 0, \text{ then } \bar{c} > \max\{a_1, a_2\}. \tag{3.28c}$$

Then (x_n) and (z_n) converge weakly to some element of S and the following results are reached:

$$\|\dot{x}_{n+1}\| = o(n^{-1}), \sum_n n\|\dot{x}_{n+1}\|^2 < \infty, , \tag{3.29a}$$

$$\|A_\lambda(x_n)\| = o((nk_n)^{-1}), \sum_n nk_n^2\|A_\lambda(x_n)\|^2 < \infty, \tag{3.29b}$$

$$\sum_n n^2\|\dot{x}_{n+1} - \dot{x}_n\|^2 < \infty, \sum_n k_n n^2\|A_\lambda(x_{n+1}) - A_\lambda(x_n)\|^2 < \infty, \tag{3.29c}$$

$$\text{for any } q \in S, \sum_n k_n\langle A_\lambda(x_n), x_n - q\rangle < \infty. \tag{3.29d}$$

Theorem 1 will be proved in Appendix (Sect. 3).

3.2.2 Convergence results in particular cases of parameters.

The estimates given in Theorem 1 are depending on the parameter (k_n), while the next result enlightens some specific properties of (k_n) given by (1.23).

Proposition 2 *Let $(k_n) \subset (0, \infty)$ be given by (1.23) with $b > 0$ and $\{a, c\} \subset [0, \infty)$. Suppose for some nonnegative integer p that $[\frac{a}{b}] \geq p$ (where $[\frac{a}{b}]$ denotes the integer part of $\frac{a}{b}$). Then there exist some positive constant C (depending on k_0, $[\frac{a}{b}]$, $[\frac{c}{b}]$) and some positive integer n_p (depending on $[\frac{a}{b}]$) for which k_n satisfies*

$$k_n \geq Cn^p \text{ for } n \geq n_p. \tag{3.30}$$

The proof of Proposition 2 is given in Appendix 2.

The above proposition allows us to give more precise estimates with respect to the involved parameters.

In specific, the next two results are immediate consequences of Theorem 1 and Proposition 2.

The first theorem is related to the special case of CRIPA when $a = 0$.

Theorem 2 *(Convergence of CRIPA with constant relaxation factors) Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, with $S := A^{-1}(0) \neq \emptyset$. Let $\{\lambda, k_0\}$ be positive constants and assume that $\{z_n, x_n\} \subset \mathcal{H}$ are generated by CRIPA-S with parameters (θ_n) and (γ_n) (given (1.24)), along with constants $\{b, \bar{c}, a_1, a_2\} \subset (0, \infty)$ verifying*

$$a_2 > 2b, a_1 > b + a_2, \tag{3.31a}$$

$$\bar{c} > \max\{a_1, a_2\}. \tag{3.31b}$$

Then (x_n) and (z_n) converge weakly to some element of S and the following results are reached:

$$\|\dot{x}_{n+1}\| = o(n^{-1}), \sum_n n\|\dot{x}_{n+1}\|^2 < \infty, \sum_n n^2\|\dot{x}_{n+1} - \dot{x}_n\|^2 < \infty, \quad (3.32\text{a})$$

$$\|A_\lambda(x_n)\| = o(n^{-1}), \sum_n n\|A_\lambda(x_n)\|^2 < \infty, \quad (3.32\text{b})$$

$$\sum_n n^2\|\dot{x}_{n+1} - \dot{x}_n\|^2 < \infty, \sum_n n^2\|A_\lambda(x_{n+1}) - A_\lambda(x_n)\|^2 < \infty, \quad (3.32\text{c})$$

$$\text{for any } q \in S, \sum_n \langle A_\lambda(x_n), x_n - q \rangle < \infty. \quad (3.32\text{d})$$

The second theorem is related to the particular case of CRIPA when $a > 0$.

Theorem 3 (Convergence of CRIPA with varying relaxation factors) *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, with $S := A^{-1}(0) \neq \emptyset$, and let $\{x_n, z_n\} \subset \mathcal{H}$ be generated by CRIPA with parameters (θ_n) and (γ_n) (given by (1.24)). Suppose for some nonnegative integer p that $\{a, b, c, a_1, a_2, \bar{c}\}$ are positive constants verifying*

$$a_2 > 2b, \ [\frac{a}{b}] \geq p, a_1 > b + a + a_2, \quad (3.33\text{a})$$

$$c > a_1 - (a_2 + a), \ \bar{c} = c + a_2 + a. \quad (3.33\text{b})$$

Then (x_n) and (z_n) converge weakly to some element of S and the following results are reached:

$$\|\dot{x}_{n+1}\| = o(n^{-1}), \sum_n n\|\dot{x}_{n+1}\|^2 < \infty, \quad (3.34\text{a})$$

$$\|A_\lambda(x_n)\| = o(n^{-(p+1)}), \sum_n n^{2p+1}\|A_\lambda(x_n)\|^2 < \infty, \quad (3.34\text{b})$$

$$\sum_n n^2\|\dot{x}_{n+1} - \dot{x}_n\|^2 < \infty, \sum_n n^{p+2}\|A_\lambda(x_{n+1}) - A_\lambda(x_n)\|^2 < \infty, \quad (3.34\text{c})$$

$$\text{for any } q \in S, \sum_n n^p \langle A_\lambda(x_n), x_n - q \rangle < \infty. \quad (3.34\text{d})$$

4 CRIPA in the convex case.

In this section, by following the methodology used by Attouch-László [2], we expose our main results relative to the minimization problem

$$\inf_{x \in \mathcal{H}} f(x), \quad (4.1)$$

where $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a proper convex and lower semi-continuous function such that $\operatorname{argmin} f \neq \emptyset$.

Indeed, by Fermat's rule we know that (4.1) is equivalent to the monotone inclusion problem

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in \partial f(x). \tag{4.2}$$

Moreover, in the special case when $A = \partial f$, CRIPA reduces to the following algorithm.

(CRIPA-convex):
▷ **Step 1** (initialization): Let $\{z_{-1}, x_{-1}, x_0\} \subset \mathcal{H}$.
▷ **Step 2** (main step): Given $\{z_{n-1}, x_{n-1}, x_n\} \subset \mathcal{H}$ (with $n \geq 0$), we compute the updates by

$$z_n = x_n + \theta_n(x_n - x_{n-1}) + \gamma_n(z_{n-1} - x_n), \tag{4.3a}$$

$$x_{n+1} = \frac{1}{1 + k_n} z_n + \frac{k_n}{1 + k_n} \text{prox}_{\lambda(1+k_n)} f(z_n), \tag{4.3b}$$

where (k_n), (θ_n) and (γ_n) are given by (1.23) and (1.24).

As the specific case of the latter algorithm when $a = 0$ we also consider the following method:

(CRIPA-S-convex):
▷ **Step 1** (initialization): Let $\{z_{-1}, x_{-1}, x_0\} \subset \mathcal{H}$.
▷ **Step 2** (main step): Given $\{z_{n-1}, x_{n-1}, x_n\} \subset \mathcal{H}$ (with $n \geq 0$), we compute the updates by

$$z_n = x_n + \theta_n(x_n - x_{n-1}) + \gamma_n(z_{n-1} - x_n), \tag{4.4a}$$

$$x_{n+1} = \frac{1}{1 + k_0} z_n + \frac{k_0}{1 + k_0} \text{prox}_{\lambda(1+k_0)} f(z_n), \tag{4.4b}$$

where k_0 is a positive constant, and where (θ_n) and (γ_n) are given by (1.24).

Remark 4 As a fundamental tool, we also recall that the Yosida approximation of ∂f is equal to the gradient of the Moreau envelope of f. Namely, for any $\lambda > 0$, we have $(\partial f)_\lambda = \nabla f_\lambda$, where $f_\lambda : \mathcal{H} \to \mathbb{R}$ is a $C^{1,1}$ function, which is defined for any $x \in \mathcal{H}$ by:

$$f_\lambda(x) = \inf_{\xi \in \mathcal{H}} \left\{ f(\xi) + \tfrac{1}{2}\lambda^{-1} \|x - \xi\|^2 \right\}. \tag{4.5}$$

So, an alternative formulation of CRIPA-convex in terms of Moreau envelope is given by

$$z_n = x_n + \theta_n(x_n - x_{n-1}) + \gamma_n(z_{n-1} - x_n), \tag{4.6a}$$

$$x_{n+1} = z_n - \lambda k_n \nabla f_{\lambda(1+k_n)}(z_n). \tag{4.6b}$$

Before exposing our results regarding CRIPA-convex, we recall some properties of the Moreau envelope through the following lemma.

Lemma 5 *Let* $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ *be a lower semi-continuous convex and proper function such that* $\operatorname{argmin} f \neq \emptyset$, *and let* $q \in S := \operatorname{argmin} f$. *Then the following properties are obtained:*

$$0 \leq f_\lambda(x_n) - \min f \leq \langle \nabla f_\lambda(x_n), x_n - q \rangle, \tag{4.7a}$$

$$0 \leq f(\operatorname{prox}_{\lambda f}(x_n)) - \min f \leq f_\lambda(x_n) - \min f, \tag{4.7b}$$

$$2\lambda^{-1}\|x_n - \operatorname{prox}_{\lambda f} x_n\|^2 \leq f_\lambda(x_n) - \min f. \tag{4.7c}$$

Proof Item (4.7a) is immediate from the gradient inequality. In addition, by definition of f_λ and the proximal mapping, we have

$$f_\lambda(x_n) - \min f = f(\operatorname{prox}_{\lambda f}(x_n)) - \min f + 2\lambda^{-1}\|x_n - \operatorname{prox}_{\lambda f} x_n\|^2. \tag{4.8}$$

This obviously implies items (4.7b) and (4.7c)   □

Now, we are in position to claim the main result of the this section.

Theorem 4 (Convergence of CRIPA-convex) *Let* $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ *be a lower semi-continuous convex and proper function such that* $S := \operatorname{argmin} f \neq \emptyset$, *and let* $\{x_n, z_n\} \subset \mathcal{H}$ *be generated by CRIPA-convex with parameters* (θ_n) *and* (γ_n) *(given by (1.24)). Suppose for some nonnegative integer* p *that* $\{a, b, c, a_1, a_2, \bar{c}\}$ *are positive constants verifying*

$$a_2 > 2b, \quad [\frac{a}{b}] \geq p, a_1 > b + a + a_2, \tag{4.9a}$$

$$c > a_1 - (a_2 + a), \bar{c} = c + (a_2 + a). \tag{4.9b}$$

Then the following properties are obtained:

$$\|x_{n+1} - x_n\| = o(n^{-1}), \|\nabla f_\lambda(x_n)\| = o(n^{-(p+1)}), \tag{4.10a}$$

$$\sum_n n\|x_{n+1} - x_n\|^2 < \infty, \sum_n n^{2p+1}\|\nabla f_\lambda(x_n)\|^2 < \infty, \tag{4.10b}$$

$$for\ any\ \ q \in S, \sum_n n^p \langle \nabla f_\lambda(x_n), x_n - q \rangle < \infty, \tag{4.10c}$$

$$\exists \bar{x} \in S,\ s.t.\ (x_n, z_n) \to (\bar{x}, \bar{x})\ weakly\ in\ \mathcal{H}^2. \tag{4.10d}$$

We also have the convergence rates:

$$f_\lambda(x_n) - \min f = o(n^{-(p+1)}), \quad \sum_n n^p (f_\lambda(x_n) - \min f) < \infty, \tag{4.11a}$$

$$f(\operatorname{prox}_{\lambda f}(x_n)) - \min f = o(n^{-(p+1)}), \quad \sum_n n^p \left(f(\operatorname{prox}_{\lambda f}(x_n)) - \min f \right) < \infty, \tag{4.11b}$$

$$\|x_n - \operatorname{prox}_{\lambda f}(x_n)\| = o(n^{-\frac{p+1}{2}}), \sum_n n^p \|x_n - \operatorname{prox}_{\lambda f}(x_n)\|^2 < \infty. \tag{4.11c}$$

Proof The results given by item (4.10) are direct consequences of Theorem 3. So, (4.7a), in light of $\|\nabla f_\lambda(x_n)\| = o(n^{-(p+1)})$ (from (4.10a)) and by boundedness of (x_n) (from (4.10d)), yields the first result in item (4.11a). The second result in item (4.11a) follows immediately from (4.7a) and (4.10c). In addition, item (4.11b) is readily deduced from (4.7b) and (4.11a), while (4.11c) is obtained from (4.7c) and (4.11a). □

Theorem 5 (Convergence of CRIPA-S-convex) *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a lower semi-continuous convex and proper function such that $S := \operatorname{argmin} f \neq \emptyset$. Let $\{\lambda, k_0\}$ be positive constants and assume that $\{x_n, z_n\} \subset \mathcal{H}$ are generated by CRIPA-S with parameters (θ_n) and (γ_n) (given by (1.24)), along with constants $\{b, a_1, a_2, \bar{c}\} \subset (0, \infty)$ verifying*

$$a_2 > 2b, \ a_1 > b + a_2, \tag{4.12a}$$

$$\bar{c} > \max\{a_1, a_2\}. \tag{4.12b}$$

Then the following results are reached:

$$\|\dot{x}_{n+1}\| = o(n^{-1}), \ \sum_n n\|\dot{x}_{n+1}\|^2 < \infty, \tag{4.13a}$$

$$\|\nabla f_\lambda(x_n)\| = o(n^{-1}), \ \sum_n n\|\nabla f_\lambda(x_n)\|^2 < \infty, \tag{4.13b}$$

$$\sum_n n^2\|\dot{x}_{n+1} - \dot{x}_n\|^2 < \infty, \ \sum_n n^2\|\nabla f_\lambda(x_{n+1}) - \nabla f_\lambda(x_n)\|^2 < \infty, \tag{4.13c}$$

$$\text{for any } q \in S, \ \sum_n \langle \nabla f_\lambda(x_n), x_n - q \rangle < \infty, \tag{4.13d}$$

$$\exists \bar{x} \in S, \ s.t. \ (x_n, z_n) \rightharpoonup (\bar{x}, \bar{x}) \text{ weakly in } \mathcal{H}^2. \tag{4.13e}$$

Proof The items in (4.13) are clearly given by Theorem 2 □

5 Numerical experiments.

Some numerical experiments are performed in this section so as to illustrate the behavior of CRIPA relative to some benchmarks.

5.1 The maximally monotone case.

As done for illustrating the performance of PRINAM in [2], we consider a model example of the skew symmetric and maximally monotone operator $A : \mathbb{R}^2 \to \mathbb{R}^2$ defined for $(\xi, \eta) \in \mathbb{R}^2$ by $A(\xi, \eta) = (-\eta, \xi)$. It is well-known that A is not the sub-differential of a convex function. We also recall that A possesses a single zero $x^* = (0, 0)$, and that A and its Yosida regularization A_λ can be identified respectively

with the matrices

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, A_\lambda = \begin{pmatrix} \frac{\lambda}{\lambda^2+1} & -\frac{1}{\lambda^2+1} \\ \frac{1}{\lambda^2+1} & \frac{\lambda}{\lambda^2+1} \end{pmatrix}. \tag{5.1}$$

We approximate a zero of A by means of several algorithms: CRIPA, KAPPA (namely Kim's accelerated proximal point algorithm given in (1.20)) and PRINAM (given in (1.17)–(1.18)). On Figs. 1 and 2 are displayed the profiles of $\|x_n - x^*\|$ for the sequences (x_n) generated by these algorithms:

- Figure 1 illustrates the behavior of the iterates (x_n) generated by CRIPA (with constant relaxation factors and proximal index) and by KAPPA (which only uses constant indexes). The profile obtained for KAPPA with the proximal parameter $\mu = 0.01$ is compared with these of CRIPA for several values of k_0 and λ such that $\lambda(k_0 + 1) = \mu$. The starting points used are $x_0 = x_{-1} = z_{-1} = (1, -1)$ for CRIPA, and $x_0 = z_0 = z_{-1} = (1, -1)$ for KAPPA. We run each method until the stopping criteria $\|x_n - x^*\| \leq 10^{-7}$ holds. The performance of both algorithms are similar on this simple example. However, one can notice the so many oscillations that are exhibited by KAPPA and which do not happen for CRIPA.

- Figure 2 illustrates the behavior of the iterates (x_n) generated by CRIPA (using varying relaxation factors and unbounded proximal indexes) and by PRINAM (also using unbounded proximal indexes). The profile obtained for PRINAM (when using the same parameters as in the optimal simulation proposed in [2]) is compared with these of CRIPA (for several values of p). The starting points used are $x_0 = (-1, 1)$ and $x_1 = (1, -1)$ for PRINAM, and $x_0 = z_{-1} = (-1, 1)$ and $x_1 = (1, -1)$ for CRIPA. Here we use the stopping criteria $\|x_n - x^*\| \leq 10^{-5}$. A faster convergence can be noticed for PRINAM than for CRIPA with $p = 1$. However the convergence of the trajectories of CRIPA is considerably accelerated for $p \geq 2$.

5.2 The convex case.

Given a symmetric and positive definite matrix $A : \mathbb{R}^N \to \mathbb{R}^N$, we consider the convex quadratic programming problem

$$\min_{x \in \mathbb{R}^N} \left\{ f(x) := \frac{1}{2} \langle Ax, x \rangle \right\}. \tag{5.2}$$

It is clear that A possesses a single zero $x^* = 0$ and that (5.2) is equivalent to solve

$$0 \in (\partial f)(\bar{x}) = A\bar{x}. \tag{5.3}$$

We approximate the solution to (5.3) by means of the following algorithms: CRIPA, AFB (given in (1.10)) and IGAHD (see [4]).

Remark 5 For convenience of the reader, we recall that IGAHD was introduced in [4] for minimizing smooth convex function $f : \mathcal{H} \to \mathbb{R}$ with L-Lipschitz continuous
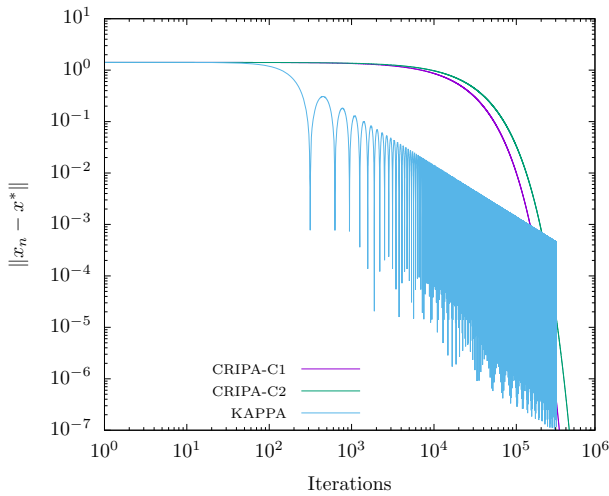
Fig. 1 CRIPA with $b = 1$, $a_2 = 2.5 * b$, $a_1 = 1.5 * (b + a_2)$, $\bar{c} = 1.5 * max\{a_1, a_2\}$, $\lambda = 0.001$ for (C1) and $\lambda = 0.005$ for (C2). For (C1) and (C2), k_0 depends on λ through $k_0 = 0.01 * \lambda^{-1} - 1$. KAPPA is considered with $\mu = 0.01$
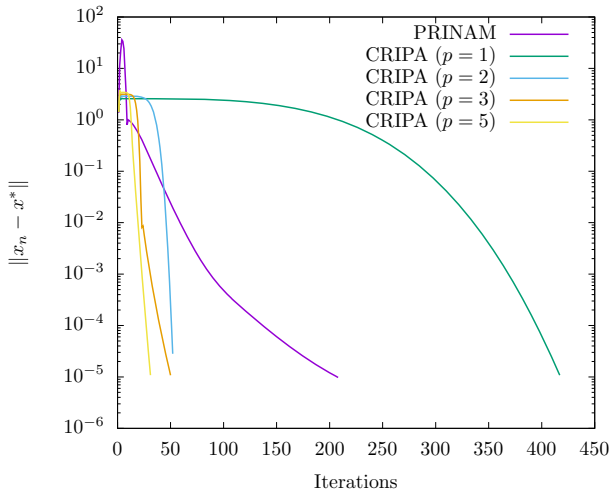

Fig. 2 CRIPA with $\lambda = 0.001$, $k_0 = 0.01$, $b = 1$, $a_2 = 3$, $a = 1.5 * (2^p - 1) * b$, $a_1 = 1.5 * (b + a + a_2)$, $c = 1.5 * (a_1 - a_2 - a)$, $\bar{c} = c + a_2 + a$. PRINAM is considered with $s = 0.1$, $\beta = 0.025$, $r = 0.1$, $q = -0.1$, $\lambda_1 = 1.01 * (2\beta + s)^2 r^2 s^{-1}$

gradient. This procedure is given, for some nonnegative values $\{s, \alpha, \beta\}$, by

$$y_n = x_n + \left(1 - \frac{\alpha}{n}\right)(x_n - x_{n-1}) - \beta\sqrt{s}\left(\nabla f(x_n) - \nabla f(x_{n-1})\right) - \frac{\beta\sqrt{s}}{n}\nabla f(x_{n-1}),$$
$$x_{n+1} = y_n - s\nabla f(y_n)).$$
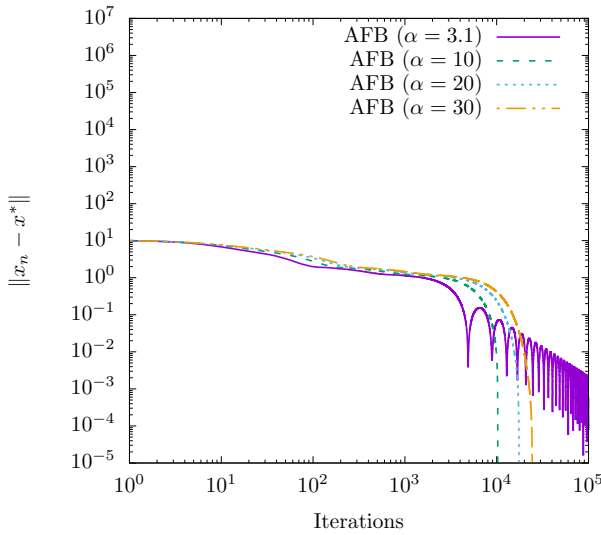
$$(5.4)$$

Fig. 3 AFB with $\lambda = 0.001$ and using several values for α

Convergence of the function values with the rates $o(n^{-2})$ as well as the property of fast convergence to zero of the gradient (that is, $\sum_n n^2 \|\nabla f(x_n)\|^2 < \infty$) were established under the conditions

$$\alpha \geq 3, 0 \leq \beta < 2\sqrt{s} \text{ and } s \leq L^{-1}. \tag{5.5}$$

Regarding our numerical simulation we take $N = 100$ and $A = B^T B$ where $B = (b_{i,j})_{1 \leq i, j \leq N}$ is a randomized invertible matrix such that $b_{i,j} \in [-1, 1]$.
On Figs. 3, 4 and 5 are displayed the profiles of $\|x_n - x^*\|$ for the iterates (x_n) generated by these algorithms. The starting points used are $x_1 = x_0 = (1, 1, ..., 1)$ for IGAHD and for AFB, while we similarly choose $x_1 = x_0 = z_{-1} = (1, 1, ..., 1)$ for CRIPA. Here we use the stopping criteria $\|x_n - x^*\| \leq 10^{-5}$:

- In order to compare CRIPA with AFB and IGAHD, we first give some insights into the influence of the parameter α on the trajectories generated by the latter two algorithms. Figures 3 and 4 feature the profiles obtained for AFB and IGAHD (for several values of α).

- On Fig. 5, some profiles obtained for IGAHD and AFB are compared with that of CRIPA (for several values of p).
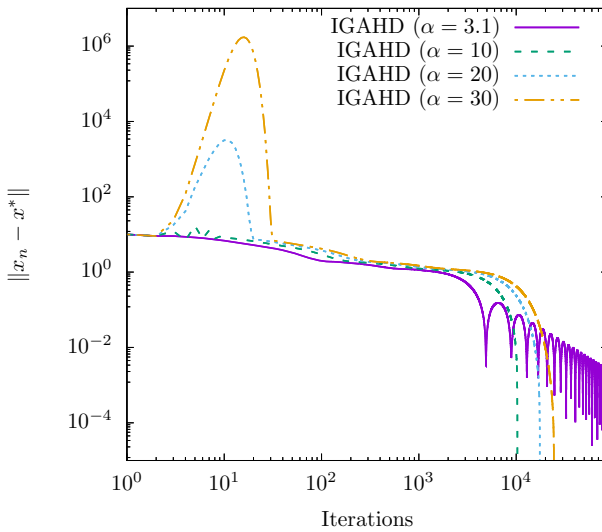
Fig. 4 IGAHD with $\lambda = 0.001$, $\beta = 0.9 * 2\sqrt{\lambda}$ and for several values of α
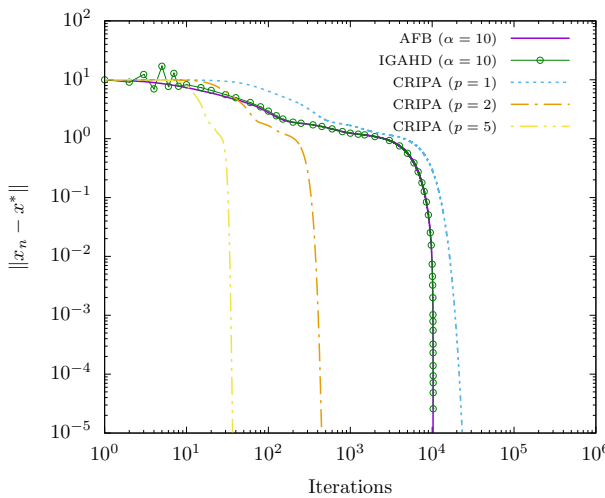

Fig. 5 CRIPA with the same parameters as in Figs. 2. IGAHD and AFB are considered with $\alpha = 10$ and the other parameters unchanged from Fig. 2

APPENDIX.

Proof of Proposition 1.

In order to get this result, we compute the discrete derivative $\dot{G}_{n+1}(s, q) := G_{n+1}(s, q) - G_n(s, q)$. For convenience of the reader we recall that $G_n(s, q)$ is given

from (2.10) by

$$G_n(s, q) = \frac{1}{2}\|s(q - x_n) - v_n \dot{x}_n\|^2 + \frac{1}{2}s(e - s)\|x_n - q\|^2. \qquad \text{(A.1)}$$

It is then readily noticed that $\dot{G}_{n+1}(s, q)$ can be formulated as

$$\dot{G}_{n+1}(s, q) = s(v_{n+1}a_{n+1} - v_n a_n) + se\dot{b}_{n+1} + v_{n+1}^2 c_{n+1} - v_n^2 c_n, \qquad \text{(A.2)}$$

where

$$a_n := \langle x_n - q, \dot{x}_n \rangle, \ b_n := \frac{1}{2}\|x_n - q\|^2 \text{ and } c_n := \frac{1}{2}\|\dot{x}_n\|^2.$$

Note also that for any bilinear symmetric form $\langle ., . \rangle_E$ on a real vector space E and for any sequences $\{\phi_n, \varphi_n\} \subset E$ we have the discrete derivative rules:

$$\langle \phi_{n+1}, \varphi_{n+1} \rangle_E - \langle \phi_n, \varphi_n \rangle_E = \langle \dot{\phi}_{n+1}, \varphi_{n+1} \rangle_E + \langle \phi_n, \dot{\varphi}_{n+1} \rangle_E, \qquad \text{(A.3a)}$$

$$\langle \phi_{n+1}, \varphi_{n+1} \rangle_E - \langle \phi_n, \varphi_n \rangle_E = \langle \dot{\phi}_{n+1}, \varphi_n \rangle_E + \langle \phi_{n+1}, \dot{\varphi}_{n+1} \rangle_E. \qquad \text{(A.3b)}$$

The sequel of the proof can be divided into the following parts **(1)–(4)**:

(1) Basic estimates. Setting $P_n = \langle q - x_{n+1}, \dot{x}_{n+1} \rangle$, we establish the elementary but useful facts below:

$$\dot{b}_{n+1} = -P_n - \frac{1}{2}\|\dot{x}_{n+1}\|^2, \qquad \text{(A.4a)}$$

$$v_{n+1}a_{n+1} - v_n a_n = v_n \langle \dot{x}_{n+1}, \dot{x}_n \rangle + e P_n - (e + v_{n+1})\langle d_n, x_{n+1} - q \rangle. \qquad \text{(A.4b)}$$

- Let us prove (A.4a). From $2b_{n+1} = \|x_{n+1} - q\|^2$, by the derivative rule (A.3a) we get

$$2\dot{b}_{n+1} = \langle \dot{x}_{n+1}, x_{n+1} - q \rangle + \langle x_n - q, \dot{x}_{n+1} \rangle$$
$$= \langle \dot{x}_{n+1}, x_{n+1} - q \rangle + \langle x_n - x_{n+1}, \dot{x}_{n+1} \rangle + \langle x_{n+1} - q, \dot{x}_{n+1} \rangle,$$

namely $2\dot{b}_{n+1} = -2P_n - \|\dot{x}_{n+1}\|^2$, which leads to (A.4a).
- Let us prove (A.4b). From $a_n = \langle q - x_n, -\dot{x}_n \rangle$, we simply get

$$a_n = \langle q - x_{n+1}, -\dot{x}_n \rangle + \langle \dot{x}_{n+1}, -\dot{x}_n \rangle = -\langle q - x_{n+1}, \dot{x}_n \rangle - \langle \dot{x}_{n+1}, \dot{x}_n \rangle. \qquad \text{(A.5)}$$

Moreover, from $a_n := \langle x_n - q, \dot{x}_n \rangle$, by the rule (A.3b) we readily get

$$\dot{a}_{n+1} = \langle \dot{x}_{n+1}, \dot{x}_n \rangle + \langle x_{n+1} - q, \dot{x}_{n+1} - \dot{x}_n \rangle. \qquad \text{(A.6)}$$

In addition, the derivative rule (A.3b) yields

$$v_{n+1}a_{n+1} - v_n a_n = \dot{v}_{n+1}a_n + v_{n+1}\dot{a}_{n+1}.$$

This latter result, in light of (A.5) and (A.6), amounts to

$$
\begin{aligned}
a_{n+1}v_{n+1} &- a_n v_n \\
&= \dot{v}_{n+1}\Big(-\langle q - x_{n+1}, \dot{x}_n \rangle - \langle \dot{x}_{n+1}, \dot{x}_n \rangle \Big) \\
&\quad + v_{n+1}\Big(\langle \dot{x}_{n+1}, \dot{x}_n \rangle - P_n + \langle q - x_{n+1}, \dot{x}_n \rangle \Big) \\
&= v_n \langle \dot{x}_{n+1}, \dot{x}_n \rangle - v_{n+1}P_n + v_n \langle q - x_{n+1}, \dot{x}_n \rangle.
\end{aligned}
\tag{A.7}
$$

Furthermore, (2.9) gives us $\dot{x}_{n+1} + d_n - \theta_n \dot{x}_n = 0$. Taking the scalar product of each side of this equality by $q - x_{n+1}$ yields

$$\theta_n \langle q - x_{n+1}, \dot{x}_n \rangle = P_n - \langle d_n, x_{n+1} - q \rangle.$$

So, recalling that $v_n = (e + v_{n+1})\theta_n$ (from (2.9)), we get

$$
\begin{aligned}
v_n \langle q - x_{n+1}, \dot{x}_n \rangle &= (e + v_{n+1})(\theta_n \langle q - x_{n+1}, \dot{x}_n \rangle) \\
&= (e + v_{n+1})\left(P_n - \langle d_n, x_{n+1} - q \rangle \right),
\end{aligned}
$$

which, in light of (A.7), entails

$$
\begin{aligned}
a_{n+1}v_{n+1} &- a_n v_n \\
&= v_n \langle \dot{x}_{n+1}, \dot{x}_n \rangle - v_{n+1}P_n + (e + v_{n+1})\left(P_n - \langle d_n, x_{n+1} - q \rangle \right) \\
&= v_n \langle \dot{x}_{n+1}, \dot{x}_n \rangle + eP_n - (e + v_{n+1})\langle d_n, x_{n+1} - q \rangle,
\end{aligned}
$$

that is (A.4b).

(2) An estimate from the inertial part. Now, given $(s, q) \in [0, \infty) \times \mathcal{H}$, we prove that the discrete derivative $\dot{G}_{n+1}(s, q)$ satisfies

$$
\begin{aligned}
\dot{G}_{n+1}(s, q) &+ s(e + v_{n+1})\langle d_n, x_{n+1} - q \rangle \\
&= s v_n \langle \dot{x}_{n+1}, \dot{x}_n \rangle - \tfrac{1}{2}\left(se - v_{n+1}^2 \right) \|\dot{x}_{n+1}\|^2 - \tfrac{1}{2}v_n^2 \|\dot{x}_n\|^2.
\end{aligned}
\tag{A.8}
$$

Indeed, in light of (A.2) and (A.4), we obtain

$$
\begin{aligned}
\dot{G}_{n+1} &= s\left(v_n \langle \dot{x}_{n+1}, \dot{x}_n \rangle + eP_n - (e + v_{n+1})\langle d_n, x_{n+1} - q \rangle \right) \\
&\quad + se\left(-P_n - \tfrac{1}{2}\|\dot{x}_{n+1}\|^2 \right) + \tfrac{1}{2}\left(v_{n+1}^2 \|\dot{x}_{n+1}\|^2 - v_n^2 \|\dot{x}_n\|^2 \right) \\
&= s v_n \langle \dot{x}_{n+1}, \dot{x}_n \rangle + \tfrac{1}{2}\left(v_{n+1}^2 - se \right) \|\dot{x}_{n+1}\|^2 - \tfrac{1}{2}v_n^2 \|\dot{x}_n\|^2 \\
&\quad - s(e + v_{n+1})\langle d_n, x_{n+1} - q \rangle,
\end{aligned}
$$

which leads obviously to the desired equality.

(3) An estimate from the proximal part. We prove that, for any $\xi_n \neq 1$, it holds that

$$
\begin{aligned}
&\xi_n \langle d_n, \dot{x}_{n+1} \rangle + \tfrac{1}{2} \| \dot{x}_{n+1} - \theta_n \dot{x}_n \|^2 \\
&= -\theta_n (1 - \xi_n) \langle \dot{x}_n, \dot{x}_{n+1} \rangle + \tfrac{1}{2} \theta_n^2 \| \dot{x}_n \|^2 - \left(\xi_n - \tfrac{1}{2} \right) \| \dot{x}_{n+1} \|^2.
\end{aligned}
\tag{A.9}
$$

Indeed, we have $\dot{x}_{n+1} = \theta_n \dot{x}_n - d_n$ (from (2.9)), hence, for any $\xi_n \neq 1$, and setting $H_n = \dot{x}_{n+1} - (1 - \xi_n)^{-1} \theta_n \dot{x}_n$, we have

$$
(1 - \xi_n) H_n = (1 - \xi_n) \dot{x}_{n+1} - \theta_n \dot{x}_n = -d_n - \xi_n \dot{x}_{n+1},
$$

or equivalently

$$
\xi_n \dot{x}_{n+1} = -(1 - \xi_n) H_n - d_n.
\tag{A.10}
$$

Furthermore, by $-d_n = \dot{x}_{n+1} - \theta_n \dot{x}_n$ (again using (2.9)) we simply obtain

$$
\begin{aligned}
\langle (-d_n), H_n \rangle &= \langle \dot{x}_{n+1} - \theta_n \dot{x}_n, \dot{x}_{n+1} - (1 - \xi_n)^{-1} \theta_n \dot{x}_n \rangle \\
&= \| \dot{x}_{n+1} \|^2 + (1 - \xi_n)^{-1} \theta_n^2 \| \dot{x}_n \|^2 - \tfrac{2 - \xi_n}{(1 - \xi_n)} \theta_n \langle \dot{x}_{n+1}, \dot{x}_n \rangle.
\end{aligned}
\tag{A.11}
$$

Therefore, taking the scalar product of each side of (A.10) with d_n, also adding $(1/2) \| d_n \|^2$ to the resulting equality, and next using (A.11) we get

$$
\begin{aligned}
\xi_n \langle d_n, \dot{x}_{n+1} \rangle + \tfrac{1}{2} \| d_n \|^2 &= (1 - \xi_n) \langle (-d_n), H_n \rangle - \tfrac{1}{2} \| d_n \|^2 \\
&= (1 - \xi_n) \left(\| \dot{x}_{n+1} \|^2 + \tfrac{\theta_n^2}{(1 - \xi_n)} \| \dot{x}_n \|^2 - \tfrac{2 - \xi_n}{(1 - \xi_n)} \theta_n \langle \dot{x}_{n+1}, \dot{x}_n \rangle \right) \\
&\quad - \tfrac{1}{2} \left(\| \dot{x}_{n+1} \|^2 + \theta_n^2 \| \dot{x}_n \|^2 - 2 \theta_n \langle \dot{x}_{n+1}, \dot{x}_n \rangle \right) \\
&= -(1 - \xi_n) \theta_n \langle \dot{x}_{n+1}, \dot{x}_n \rangle + \tfrac{1}{2} \theta_n^2 \| \dot{x}_n \|^2 + \left(\tfrac{1}{2} - \xi_n \right) \| \dot{x}_{n+1} \|^2.
\end{aligned}
$$

This, while noticing that $\| d_n \|^2 = \| \dot{x}_{n+1} - \theta_n \dot{x}_n \|^2$ (from (2.9)) , yields (A.9).

(4) Combining proximal and inertial effects. At once we show for $s \in (0, e]$ that the iterates verify (for $n \geq 0$)

$$
\begin{aligned}
&\dot{G}_{n+1}(s, q) + \tfrac{1}{2} (e + v_{n+1})^2 \| \dot{x}_{n+1} - \theta_n \dot{x}_n \|^2 \\
&+ s(e + v_{n+1}) \langle d_n, x_{n+1} - q \rangle + (e + v_{n+1})(e - s + v_{n+1}) \langle d_n, \dot{x}_{n+1} \rangle \\
&= -\tfrac{1}{2} (e - s)(e + 2v_{n+1}) \| \dot{x}_{n+1} \|^2.
\end{aligned}
\tag{A.12}
$$

Indeed, denoting $\tau_n = e + v_{n+1}$, from (A.8) we know that

$$
\begin{aligned}
&\dot{G}_{n+1}(s, q) + s \tau_n \langle d_n, x_{n+1} - q \rangle \\
&= s v_n \langle \dot{x}_{n+1}, \dot{x}_n \rangle - \tfrac{1}{2} \left(se - v_{n+1}^2 \right) \| \dot{x}_{n+1} \|^2 - \tfrac{1}{2} v_n^2 \| \dot{x}_n \|^2.
\end{aligned}
\tag{A.13}
$$

Furthermore, assuming that $s \in (0, e]$ and taking $\xi_n = 1 - s\tau_n^{-1}$ in (A.9), while noticing that $\tau_n \theta_n = v_n$ (from (2.9b)), we obtain

$$
\begin{aligned}
&(1 - s\tau_n^{-1})\langle d_n, \dot{x}_{n+1}\rangle + \tfrac{1}{2}\|\dot{x}_{n+1} - \theta_n \dot{x}_n\|^2 \\
&= -s\theta_n \tau_n^{-1}\langle \dot{x}_n, \dot{x}_{n+1}\rangle + \tfrac{1}{2}\theta_n^2\|\dot{x}_n\|^2 - \left(\tfrac{1}{2} - s\tau_n^{-1}\right)\|\dot{x}_{n+1}\|^2 \\
&= -s v_n \tau_n^{-2}\langle \dot{x}_n, \dot{x}_{n+1}\rangle + \tfrac{1}{2}v_n^2\tau_n^{-2}\|\dot{x}_n\|^2 - \left(\tfrac{1}{2} - s\tau_n^{-1}\right)\|\dot{x}_{n+1}\|^2.
\end{aligned}
\tag{A.14}
$$

Then multiplying equality (A.14) by τ_n^2 and adding the resulting equality to (A.13) yields

$$
\begin{aligned}
&(1 - s\tau_n^{-1})\tau_n^2\langle d_n, \dot{x}_{n+1}\rangle + \tfrac{1}{2}\tau_n^2\|\dot{x}_{n+1} - \theta_n \dot{x}_n\|^2 \\
&\quad + \dot{G}_{n+1}(s, q) + s\tau_n\langle d_n, x_{n+1} - q\rangle \\
&= \left(-\tfrac{1}{2}\left(se - v_{n+1}^2\right) - \tau_n^2\left(\tfrac{1}{2} - s\tau_n^{-1}\right)\right)\|\dot{x}_{n+1}\|^2 = -T_n,
\end{aligned}
\tag{A.15}
$$

where T_n is defined by

$$
T_n = \frac{1}{2}\left(se - v_{n+1}^2 + \tau_n^2\left(1 - 2s\tau_n^{-1}\right)\right)\|\dot{x}_{n+1}\|^2.
\tag{A.16}
$$

In addition, as $\tau_n := e + v_{n+1}$, a simple computation yields

$$
\begin{aligned}
\tau_n^2\left(1 - 2s\tau_n^{-1}\right) &= e^2 + 2ev_{n+1} + (v_{n+1})^2 - 2s\left(e + v_{n+1}\right) \\
&= e\left(e - s\right) - se + 2v_{n+1}\left(e - s\right) + (v_{n+1})^2 \\
&= \left(e + 2v_{n+1}\right)\left(e - s\right) - se + (v_{n+1})^2.
\end{aligned}
$$

As a consequence, by (A.16) we obtain

$$
T_n = \frac{(e - s)}{2}\left(e + 2v_{n+1}\right)\|\dot{x}_{n+1}\|^2.
$$

This ends the proof $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of Proposition 2.

Let us first prove the result for $b = 1$. Consider the positive sequence $(k_n)_{n\geq 0}$ defined, for $n \geq 1$ and for some constants $\{a, c\} \subset [0, \infty)$, by the recursive formula $\frac{k_n}{k_{n-1}} = 1 + \frac{a}{n+c}$, . As a consequence, from the basic inequalities $a \geq [a]$ and $c \leq [c] + 1$, we get

$$
\frac{k_n}{k_{n-1}} \geq 1 + \frac{[a]}{n+[c]+1} = \frac{n+[c]+[a]+1}{n+[c]+1}.
$$

Hence, for $n \geq 1$, we obtain

$$
\frac{k_n}{k_0} \geq \prod_{j=1}^{n} \frac{j+[c]+[a]+1}{j+[c]+1} = \frac{\prod_{j=1}^{n}(j+[c]+[a]+1)}{\prod_{j=1}^{n}(j+[c]+1)} = \frac{\prod_{j=2+[c]+[a]}^{n+[c]+1+[a]} j}{\prod_{j=2+[c]}^{n+[c]+1} j}.
\tag{A.17}
$$

Moreover, for $n \geq 3 + [a]$, we simply get

$$\prod_{j=2+[c]+[a]}^{n+[c]+1+[a]} j = (2 + [c] + [a]) \left(\prod_{j=3+[c]+[a]}^{n+[c]} j \right) \left(\prod_{j=n+[c]+1}^{n+[c]+1+[a]} j \right), \quad \text{(A.18a)}$$

$$\prod_{j=2+[c]}^{n+[c]+1} j = \left(\prod_{j=2+[c]}^{2+[c]+[a]} j \right) \left(\prod_{j=3+[c]+[a]}^{n+[c]} j \right) (n + [c] + 1). \quad \text{(A.18b)}$$

As a consequence, for $n \geq 3 + [a]$, by (A.17) in light of (A.18) we get

$$k_n \geq k_0 \frac{(2+[c]+[a]) \left(\prod_{j=n+[c]+1}^{n+[c]+1+[a]} j \right)}{\left(\prod_{j=2+[c]}^{2+[c]+[a]} j \right)(n+[c]+1)} = C_0 \frac{\prod_{j=n+[c]+1}^{n+[c]+1+[a]} j}{n+[c]+1} = C_0 \frac{\prod_{j=[c]+1}^{[c]+1+[a]} (n+j)}{n+[c]+1},$$

where $C_0 := k_0 \left(\frac{2+[c]+[a]}{\prod_{j=2+[c]}^{2+[c]+[a]} j} \right)$, which implies that

$$k_n \geq C_0 (n + [c] + 1)^{[a]} \geq C_0 n^{[a]}.$$

The desired result follows immediately from the previous arguments when replacing a and c by $\frac{a}{b}$ and $\frac{c}{b}$, respectively. This completes the proof $\qquad \square$

Proof of Theorem 1

The proof will be divided into the following steps **(a)**, **(b)** and **(c)**:

(a) In order to establish (3.29a) and (3.29b), we first prove the following estimates

$$\|\dot{x}_n\| = o(n^{-1}), \quad \text{(A.19a)}$$

$$k_{n-1}\|A_\lambda(x_n)\| = o(n^{-1}). \quad \text{(A.19b)}$$

Indeed, passing to the limit as $s \to 0^+$ in (2.8) amounts to

$$\begin{aligned}
&(b(n+1) + \bar{c} - a_1)^2 \|\dot{x}_{n+1}\|^2 - (bn + \bar{c} - a_1)^2 \|\dot{x}_n\|^2 \\
&+ \lambda k_{n-1}(bn + \bar{c}) \left(a \frac{bn + \bar{c}}{bn + c} + a_2 \right) \langle A_\lambda(x_n), \dot{x}_{n+1} \rangle \\
&+ \frac{1}{2}(a_1 - b)\big((2n+1)b + 2\bar{c} - a_1\big) \|\dot{x}_{n+1}\|^2 \leq 0.
\end{aligned} \quad \text{(A.20)}$$

Concerning the second quantity in the right side of the above inequality, by $b > 0$, $a \geq 0$ and $a_2 > 0$ (hence $a + a_2 > 0$) we readily have

$$(bn + \bar{c}) \left(a \frac{bn + \bar{c}}{bn + c} + a_2 \right) \sim b(a + a_2)n \text{ as } n \to \infty.$$

Then, in light of $\sum_n n k_{n-1} |\langle A_\lambda(x_n), \dot{x}_{n+1}\rangle| < \infty$ (from (3.14d)), we infer that

$$\sum_n (bn + \bar{c}) \left(a \frac{bn + \bar{c}}{bn + c} + a_2 \right) k_{n-1} |\langle A_\lambda(x_n), \dot{x}_{n+1}\rangle| < \infty.$$

It is then a classical matter to derive from (A.20) that there exists some $l_2 \geq 0$ such that $\lim_{n\to\infty} n^2 \|\dot{x}_n\|^2 = l_2$. Moreover, recalling that $\sum_n n \|\dot{x}_n\|^2 < \infty$ (from (3.3f)), and noticing that $\sum_n n^{-1} = \infty$, we infer that $\liminf_{n\to\infty} n^2 \|\dot{x}_n\|^2 = 0$. It follows that $l_2 = 0$, that is (A.19a).

Next combining the latter result with $\|\dot{x}_n + \lambda k_{n-1} A_\lambda(x_n)\| = o(n^{-1})$ (from (3.14b)) gives us immediately that $\lim_{n\to\infty} n k_{n-1} \|A_\lambda(x_n)\| = 0$, that is (A.19b).

Then (3.29a) is given by (3.3f) and (A.19a), while (3.29b) follows from (3.14c) and (A.19b).

(b) The estimates in (3.29c) are derived from (3.3e) and (3.3g), while the estimate (3.29d) is given by (3.3d).

(c) Finally, we prove the weak convergence of the iterates $\{x_n, z_n\}$ given by CRIPA by means of the Opial lemma. This latter result guarantees that (x_n) converges weakly to some element of S, provided that the following results (h1)–(h2) hold: (h1) for any $q \in S$, the sequence $(\|x_n - q\|)$ is convergent; (h2) any weak-cluster point of (x_n) belongs to S.

- Let us prove (h1). Take $q \in S$. Clearly, as a straightforward consequence of the bounded-ness of (x_n) (given by (3.3a)) along with (A.19b) we have

$$k_{n-1}\langle A_\lambda(x_n), x_n - q\rangle = o(n^{-1}). \tag{1.21}$$

Moreover, we know that $(\mathcal{E}_n(a + a_2, q))$ is convergent (from Lemma 3) and writes

$$\begin{aligned}
\mathcal{E}_n(a + a_2, q) &= \frac{1}{2}\|(a + a_2)(q - x_n) - (bn + \bar{c} - a_1)\dot{x}_n\|^2 \\
&\quad + \frac{1}{2}(a + a_2)(a_1 - b - (a + a_2))\|x_n - q\|^2 \\
&\quad + (a + a_2)(b(n - 1) + \bar{c})(\lambda k_{n-1})\langle A_\lambda(x_n), x_n - q\rangle.
\end{aligned} \tag{1.22}$$

Then, by $n\|\dot{x}_n\| \to 0$ (from (A.19a)) and $n k_{n-1} |\langle A_\lambda(x_n), x_n - q\rangle| \to 0$ (according to (1.21)) as $n \to \infty$, we deduce that

$$\lim_{n\to\infty} \mathcal{E}_n(a + a_2, q) = \lim_{n\to\infty} \frac{1}{2}(a + a_2)(a_1 - b)\|x_n - q\|^2. \tag{1.23}$$

This entails (h1).

- Now, we prove (h2). Let u be a weak cluster point of (x_n), namely there exists a subsequence (x_{n_k}) that converges to u as $k \to \infty$. So, in view of (A.19b), we simply have $\lim_{k\to+\infty} \|A_\lambda(x_{n_k})\| = 0$, while a classical result gives us $A_\lambda(x_{n_k}) \in A(x_{n_k} - \lambda A_\lambda(x_{n_k}))$. Then passing to the limit as $k \to \infty$ in this latter result and recalling that the graph of a maximally monotone operator is demi-closed (see, for instance, [13]), we deduce that $0 \in A(u)$, namely $u \in S$. This proves (h2).

Then we infer by Opial's lemma that (x_n) converges weakly to some element $\bar{x} \in S$.

Finally, by $\sum_n n \|k_{n-1} A_\lambda (x_n)\|^2 < \infty$ (from (3.14c)), we get $k_{n-1} A_\lambda (x_n) \to 0$ (as $n \to \infty$), hence, by $z_{n-1} - x_n = \lambda k_{n-1} A_\lambda (x_n)$ (from (2.4a)), we are led to $z_{n-1} - x_n \to 0$ (as $n \to \infty$). It follows that (z_n) converges weakly to the element $\bar{x}$. This completes the proof   □

References

1. Attouch, H.: Fast inertial proximal ADMM algorithms for convex structured optimization with linear constraint. Minimax Theory Its Appl. **06**(1), 1–24 (2021)
2. Attouch, H., László, S.C.: Newton-like inertial dynamics and proximal algorithms governed by maximally monotone operators. SIAM J. Optim. **30**(4), 3252–3283 (2020)
3. Attouch, H., László, S.C.: Continuous Newton-like inertial dynamics for monotone inclusions. Set-Valued Var. Anal. (2020). https://doi.org/10.1007/s11228-020-00564-y
4. Attouch, H., Chbani, Z., Fadili, J., Riahi, H.: First-order optimization algorithms via inertial systems with Hessian driven damping, arXiv preprint, arXiv:1907.10536 (2019)
5. Attouch, H., Peypouquet, J.: Convergence of inertial dynamics and proximal algorithms governed by maximal monotone operators. Math. Programm. **174**, 391–432 (2019)
6. Attouch, H., Peypouquet, J.: Convergence Rate of Proximal Inertial Algorithms Associated with Moreau Envelopes of Convex Functions. In: Bauschke, H., Burachik, R., Luke, D. (eds.) Splitting Algorithms, Modern Operator Theory, and Applications. Springer, Cham (2019)
7. 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)
8. Attouch, H., Svaiter, B.F.: A continuous dynamical Newton-like approach to solving monotone inclusions. SIAM J. Control Optim. **49**(2), 574–598 (2011)
9. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2017)
10. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)
11. Boţ, R.I., Csetnek, E.R.: Second order forward-backward dynamical systems for monotone inclusion problems. SIAM J. Control Optim. **54**, 1423–1443 (2016)
12. Boţ, R.I., Csetnek, E.R.: ADMM for monotone operators: convergence analysis and rates. Adv. Comput. Math. **45**(1), 327–359 (2019)
13. Brezis, H.: Opérateurs maximaux monotones, Mathematical Studies, vol. 5. North-Holland, Amsterdam (1973)
14. Brezis, H.: Function Analysis, Sobolev Spaces and Partial Differential Equations. Springer, New York (2010). https://doi.org/10.1007/978-0-387-70914-7
15. Brezis, H., Lions, P.L.: Produits infinis de résolvantes. Isr. J. Math. **29**, 329–345 (1978)
16. Chambolle, A., Dossal, C.: On the convergence of the iterates of Fista. J. Optim. Theory Appl. **166**(3), 968–982 (2015)
17. Combettes, P.L.: Monotone operator theory in convex optimization. Math. Programm. Vol. B **170**(1), 177–206 (2018)
18. Corman, E., Yuan, X.: A generalized proximal point algorithm and its convergence rate. SIAM J. Optim. **24**(4), 1614–1638 (2014)
19. Douglas, J., Rachford, H.H.: On the numerical solution of heat conduction problems in two and three space variables. Trans. Am. Math. Soc. **82**, 421–439 (1956)
20. Drori, Y., Teboulle, M.: Performance of first-order methods for smooth convex minimization: a novel approach. Math. Programm. **145**(1–2), 451–482 (2014)
21. Eckstein, J., Bertsekas, D.P.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Programm. **55**(1–3), 293–318 (1992)
22. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite-element approximations. Comput. Math. Appl. **2**(1), 17–40 (1976)

23. Gu, G., Yang, J.: *Optimal nonergodic sublinear convergence rate of proximal point algorithm for maximal monotone inclusion problems*, (2019), arXiv:1904.05495
24. Güler, O.: On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. **29**, 403–419 (1991)
25. Güler, O.: New proximal point algorithms for convex minimization. SIAM J. Optim. **2**(4), 649–664 (1992)
26. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**, 302–320 (1969)
27. Kim, D.: Accelerated proximal point method for maximally monotone operators, (2019), arXiv:1905.05149
28. Kim, D., Fessler, J.A.: Another look at the Fast Iterative Shrinkage/Thresholding Algorithm (FISTA). SIAM J. Optim. **28**(1), 223–250 (2018)
29. Kim, D., Fessler, J.A.: Generalizing the optimized gradient method for smooth convex minimization. SIAM J. Optim. **28**(2), 1920–1950 (2018)
30. Lemaire, B.: The proximal algorithm, in: New methods in optimization and their industrial uses. In: J.P. Penot (ed), Internat. Ser. Numer. Math. 87, Birkhauser, Basel, pp. 73-87 (1989)
31. Lin, H., Mairal, J., Harchaoui, Z.: Catalyst acceleration for first-order convex optimization: from theory to practice. J. Mach. Learning Res. **18**(212), 1–54 (2018)
32. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**(6), 964–979 (1979)
33. Lorenz, D.A., Pock, T.: An inertial forward-backward algorithm for monotone inclusions. J. Math. Imaging Vis. **51**, 311–325 (2015)
34. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**, 964–979 (1979)
35. Maingé, P.E.: First-order continuous newton-like systems for monotone inclusions. SIAM J. Control Optim. **51**(2), 1615–1638 (2013)
36. Martinet, B.: Régularisation d'in' equations variationnelles par approximations successives. Rev. Fr. Infor. Rech. Opération. **4**, 154–158 (1970)
37. Nesterov, Y.: A method of solving a convex programming problem with convergence rate O(1/k2). Soviet Math. Doklady **27**, 372–376 (1983)
38. Nesterov, Y.: A method for unconstrained convex minimization problem with the rate of convergence O(1/k2). Dokl. Akad. Nauk. USSR **269**(3), 543–7 (1983)
39. Nesterov, Y.: Gradient methods for minimizing composite objective function. Math. Programm. Ser. B **140**, 125–161 (2013)
40. Peaceman, D.W., Rachford, H.H.: The numerical solution of parabolic and elliptic differential equations. J. Soc. Ind. Appl. Math. **3**(1), 28–41 (1955)
41. Powell, M.J.D.: Optimization. In: Fletcher, R. (ed.) A Method for Nonlinear Constraints in Minimization Problems, pp. 283–98. Academic Press, New York (1969)
42. Rockafellar, R.T.: Monotone operators associated with saddle functions and minimax problems. In: F. E. Browder, (ed), Nonlinear Functional Analysis, Part 1. Symposia in Pure Math., vol. 18, American Mathematical Society, Providence, RI., pp. 397–407 (1970)
43. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Oper. Res. **1**(2), 97–116 (1976)
44. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control. Opt. **14**(5), 877–898 (1976)
45. Rockafellar, R.T., Wets, J.B.: Variational Analysis. Springer, Berlin (1998)
46. Takahashi, W.: Nonlinear Functional Analysis. Yokohama Publishers, Yokohama (2000)
47. Zhang, X.Q., Burger, M., Bresson, X., Osher, S.: Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. SIAM J. Imaging Sci. **3**(3), 253–276 (2010)