



HAL
open science

Convergence rate of a relaxed inertial proximal algorithm for convex minimization

Hedy Attouch, Alexandre Cabot

► **To cite this version:**

Hedy Attouch, Alexandre Cabot. Convergence rate of a relaxed inertial proximal algorithm for convex minimization. 2018. hal-01807041

HAL Id: hal-01807041

<https://hal.science/hal-01807041>

Preprint submitted on 4 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONVERGENCE RATE OF A RELAXED INERTIAL PROXIMAL ALGORITHM FOR CONVEX MINIMIZATION

HEDY ATTOUCH AND ALEXANDRE CABOT

ABSTRACT. In a Hilbert space setting, the authors recently introduced a general class of relaxed inertial proximal algorithms that aim at solving monotone inclusions. In this paper, we specialize this study in the case of non-smooth convex minimization problems. We obtain convergence rates for the values which have similarities with the results based on the Nesterov accelerated gradient method. The joint adjustment of inertia, relaxation and proximal terms plays a central role. In doing so, we put to the fore inertial proximal algorithms that converge for general monotone inclusions, and which, in the case of convex minimization, give fast convergence rates of the values in the worst case.

1. INTRODUCTION

Throughout the paper, \mathcal{H} is a real Hilbert space equipped with the scalar product $\langle \cdot, \cdot \rangle$ and the corresponding norm $\|\cdot\|$. The Relaxed Inertial Proximal Algorithm (RIPA) was introduced by the authors in [6]. It aims at solving by fast methods the inclusion $A(x) \ni 0$, where $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ is a general maximally monotone operator. (RIPA) is a proximal-based algorithm involving both inertia and relaxation aspects. In [6], it is shown the weak convergence of the iterates to equilibria under general conditions on the inertia, relaxation, and proximal coefficients of the algorithm. It extends the work of Attouch-Peypouquet [9] which concerns the particular case where the inertial coefficient is of the form $1 - \frac{\alpha}{k}$, an important issue because of its connection with the accelerated method of Nesterov, see [4], [7], [17], [31].

In this paper, we specialize the study of (RIPA) in the case of (nonsmooth) convex minimization problems. This corresponds to taking $A = \partial\Phi$ where $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex lower semicontinuous proper function. The algorithm is then written

$$(1) \quad \text{(RIPA)} \quad \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = (1 - \rho_k)y_k + \rho_k \text{prox}_{\mu_k \Phi}(y_k), \end{cases}$$

where the proximal mapping $\text{prox}_{\mu\Phi} : \mathcal{H} \rightarrow \mathcal{H}$ is defined, for every $x \in \mathcal{H}$, and every $\mu > 0$ by the formula

$$\text{prox}_{\mu\Phi}(x) = \operatorname{argmin}_{\xi \in \mathcal{H}} \left\{ \mu\Phi(\xi) + \frac{1}{2}\|x - \xi\|^2 \right\}.$$

The sequence (α_k) of nonnegative numbers reflects the inertial aspects, while (ρ_k) is the sequence of positive relaxation parameters. The sequence (μ_k) of positive numbers makes it possible to consider the proximal operator with a varying parameter, which is an important issue in this context, see [9]. By using variational methods and well-chosen Lyapunov functions, we obtain convergence rates for the values. In doing so, we put to the fore inertial algorithms whose iterates converge in the case of general monotone inclusions, and which, in the case of convex minimization, give fast convergence

Date: May 24, 2018.

1991 Mathematics Subject Classification. 49M37, 65K05, 65K10, 90C25.

Key words and phrases. Inertial proximal method; Lyapunov analysis; maximally monotone operators; nonsmooth convex minimization; relaxation; subdifferential of convex functions;

rates of the values (in the worst case). Let's first recall the main results concerning (RIPA) in the general maximally monotone framework.

1.1. (RIPA) for general monotone inclusions. Given $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ a maximally monotone operator, the Relaxed Inertial Proximal Algorithm, (RIPA) for short, is defined by, for $k \geq 1$

$$(RIPA) \quad \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = (1 - \rho_k)y_k + \rho_k J_{\mu_k A}(y_k). \end{cases}$$

In the above formula, $J_{\mu A} = (I + \mu A)^{-1}$ is the *resolvent* of A with index $\mu > 0$. It plays a central role in the analysis of (RIPA), along with the *Yosida regularization* of A with parameter $\mu > 0$, which is defined by $A_{\mu} = \frac{1}{\mu}(I - J_{\mu A})$. We denote by $\text{zer } A$ the set of zeros of A , and we assume that $\text{zer } A \neq \emptyset$. The weak convergence of the sequences (x_k) generated by the algorithm (RIPA) is based on a joint tuning of the sequences (α_k) , (μ_k) , (ρ_k) . As a crucial ingredient, the analysis of the convergence properties of (RIPA) involves the sequence (t_k) defined by

$$(2) \quad t_k := 1 + \sum_{l=k}^{+\infty} \left(\prod_{j=k}^l \alpha_j \right).$$

The sequence (t_k) is well defined under the following standing assumption

$$(K_0) \quad \sum_{l=k}^{+\infty} \left(\prod_{j=k}^l \alpha_j \right) < +\infty \quad \text{for every } k \geq 1.$$

The formula (2) that allows to pass from (α_k) to (t_k) may seem complicated. By contrast, one can simply retrieve (α_k) from (t_k) thanks to the following formula that will prove very useful.

Lemma 1.1. *Assume that the nonnegative sequence (α_k) satisfies (K_0) . Then the sequence (t_k) satisfies*

$$(3) \quad 1 + \alpha_k t_{k+1} = t_k \quad \text{for every } k \geq 1.$$

Let us first formulate the convergence properties of (RIPA) in the case (ρ_k) bounded away from zero.

Theorem 1.2. ([6, Theorem 2.6]) *Let $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ be a maximally monotone operator such that $\text{zer } A \neq \emptyset$. Suppose that $\alpha_k \in [0, 1]$, $\rho_k \in]0, 2]$ and $\mu_k > 0$ for every $k \geq 1$. Under (K_0) , let (t_k) be the sequence defined by (2). Assume that there exists $\varepsilon \in]0, 1[$ such that for k large enough,*

$$(L) \quad (1 - \varepsilon) \frac{2 - \rho_{k-1}}{\rho_{k-1}} (1 - \alpha_{k-1}) \geq \alpha_k t_{k+1} \left(1 + \alpha_k + \left[\frac{2 - \rho_k}{\rho_k} (1 - \alpha_k) - \frac{2 - \rho_{k-1}}{\rho_{k-1}} (1 - \alpha_{k-1}) \right]_+ \right).$$

Then for any sequence (x_k) generated by (RIPA), we have

$$(i) \quad \sum_{k=1}^{+\infty} \frac{2 - \rho_{k-1}}{\rho_{k-1}} (1 - \alpha_{k-1}) \|x_k - x_{k-1}\|^2 < +\infty, \text{ and hence } \sum_{k=1}^{+\infty} \alpha_k t_{k+1} \|x_k - x_{k-1}\|^2 < +\infty.$$

$$(ii) \quad \sum_{k=1}^{+\infty} \rho_k (2 - \rho_k) t_{k+1} \|\mu_k A_{\mu_k}(x_k)\|^2 < +\infty.$$

(iii) *For any $z \in \text{zer } A$, $\lim_{k \rightarrow +\infty} \|x_k - z\|$ exists, and hence (x_k) is bounded.*

Assume moreover that $\limsup_{k \rightarrow +\infty} \rho_k < 2$, and $\liminf_{k \rightarrow +\infty} \rho_k > 0$. Then the following holds

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

(v) *If $\liminf_{k \rightarrow +\infty} \mu_k > 0$, then there exists $x_{\infty} \in \text{zer } A$ such that $x_k \rightharpoonup x_{\infty}$ weakly in \mathcal{H} as $k \rightarrow +\infty$.*

Let us now state the convergence results in the case of a possibly vanishing sequence (ρ_k) .

Theorem 1.3. ([6, Theorem 2.14]) *Let $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ be a maximally monotone operator such that $\text{zer}A \neq \emptyset$. Suppose that $\alpha_k \in [0, 1]$, $\rho_k \in]0, 2]$ and $\mu_k > 0$ for every $k \geq 1$. Assume moreover that conditions (K_0) and (L) hold true. Then for any sequence (x_k) generated by (RIPA),*

(i) *There exists $C \geq 0$ such that for every $k \geq 1$, $\|x_{k+1} - x_k\| \leq C \sum_{i=1}^k \left[\left(\prod_{j=i+1}^k \alpha_j \right) \rho_i \right]$.*

Assume moreover that $\limsup_{k \rightarrow +\infty} \rho_k < 2$, together with

- $\sum_{i=1}^k \left[\left(\prod_{j=i+1}^k \alpha_j \right) \rho_i \right] = \mathcal{O}(\rho_k t_{k+1})$, $\frac{|\mu_{k+1} - \mu_k|}{\mu_{k+1}} = \mathcal{O}(\rho_k t_{k+1})$, $\rho_{k-1} t_k = \mathcal{O}(\rho_k t_{k+1})$ as $k \rightarrow +\infty$;
- $\sum_{k=1}^{+\infty} \rho_k t_{k+1} = +\infty$.

Then the following holds

(ii) $\lim_{k \rightarrow +\infty} \mu_k A_{\mu_k}(x_k) = 0$. *If $\liminf_{k \rightarrow +\infty} \mu_k > 0$, then there exists $x_\infty \in \text{zer}A$ such that $x_k \rightarrow x_\infty$ weakly in \mathcal{H} as $k \rightarrow +\infty$.*

In particular, the above results provide estimates concerning the rate of convergence to zero of the discrete velocities. When A is the subdifferential of a convex lower semicontinuous proper function, we will complete these results by estimating the rate of convergence of the values.

1.2. (RIPA) for convex minimization. Let us now suppose that A is the subdifferential of a convex lower semicontinuous proper function $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$, let $A = \partial\Phi$. Then, as a classical result, $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ is a maximally monotone operator, and the minimizers of Φ are exactly the zeros of A . Let us specialize (RIPA) in this framework. We have $J_{\mu\partial\Phi} = \text{prox}_{\mu\Phi}$, hence the formulation (1) of (RIPA). Let's give another useful formulation of (RIPA). By definition of the Yosida approximation

$$(1 - \rho_k)y_k + \rho_k J_{\mu_k A}(y_k) = y_k - \rho_k \mu_k A_{\mu_k}(y_k).$$

We have $A_\mu = \nabla\Phi_\mu$, where Φ_μ is the Moreau envelope of Φ with index $\mu > 0$. Recall that, for all $x \in \mathcal{H}$

$$\Phi_\mu(x) = \inf_{\xi \in \mathcal{H}} \left\{ \Phi(\xi) + \frac{1}{2\mu} \|x - \xi\|^2 \right\},$$

see Appendix A for its classical properties. It follows that (RIPA) can be rewritten in an equivalent way

$$(4) \quad (\text{RIPA}) \quad \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = y_k - \rho_k \mu_k \nabla\Phi_{\mu_k}(y_k). \end{cases}$$

As a distinctive property, $\Phi_\mu : \mathcal{H} \rightarrow \mathbb{R}$ is a convex differentiable function, whose gradient is Lipschitz continuous. Thus, when A is the subdifferential of a convex lower semicontinuous proper function, this property naturally links (RIPA) to the inertial gradient methods, see [1], [4], [8], [10] [13], [17], [18], [22], [23], [26], [31], [32] for some of the rich literature that has been devoted to this class of algorithms in recent years. The new aspects of the algorithm (RIPA) are the general inertial coefficients (α_k) , the fact that the potential Φ_{μ_k} varies at each iteration, as well as the step size $\rho_k \mu_k$. A helpful study is [4], where the authors considered inertial forward-backward algorithms for structured convex minimization problems, and with general inertial coefficients.

1.3. Links with continuous dynamics. Linking (RIPA) to continuous inertial dynamical systems provides a physical intuition, and enlightens the role of parameters entering the algorithm. Following [5] and [9], let us consider the continuous dynamics

$$(5) \quad \ddot{x}(t) + \gamma(t)\dot{x}(t) + \nabla\Phi_{\mu(t)}(x(t)) = 0,$$

where $\gamma(\cdot)$ is a positive viscous damping parameter, which is time-dependent. It is governed by the Yosida regularization of the maximally monotone operator $\partial\Phi$, with a time-dependent regularization

parameter $\mu(t)$. Thanks to the Lipschitz continuity property of the Yosida approximation, this dynamics is relevant owing to the Cauchy-Lipschitz theorem, and the corresponding Cauchy problem is well-posed. Let us describe successively explicit and implicit discretization of (5).

Take a time step $h_k > 0$, and set $t_k = \sum_{i=1}^k h_i$, $x_k = x(t_k)$, $\mu_k = \mu(t_k)$, $\gamma_k = \gamma(t_k)$.

1.3.1. *Explicit discretization.* A finite-difference scheme for (5) with centered second-order variation gives

$$(6) \quad \frac{1}{h_k^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\gamma_k}{h_k}(x_k - x_{k-1}) + \nabla\Phi_{\mu_k}(y_k) = 0,$$

where y_k is an extrapolated point from x_k and x_{k-1} that will be chosen later (because $\nabla\Phi_{\mu_k}$ is Lipschitz continuous, there is some flexibility in this choice). After developing (6), we obtain

$$x_{k+1} = x_k + (1 - \gamma_k h_k)(x_k - x_{k-1}) - h_k^2 \nabla\Phi_{\mu_k}(y_k).$$

Take $y_k = x_k + (1 - \gamma_k h_k)(x_k - x_{k-1})$, which is the Nesterov choice for the extrapolation term. We obtain

$$\begin{cases} y_k &= x_k + (1 - \gamma_k h_k)(x_k - x_{k-1}) \\ x_{k+1} &= y_k - h_k^2 \nabla\Phi_{\mu_k}(y_k). \end{cases}$$

That's algorithm (RIPA), as formulated in (4) with $\alpha_k = 1 - \gamma_k h_k$ and $\rho_k = \frac{h_k^2}{\mu_k}$. Thus $\rho_k \mu_k$ has the dimension of the square of a length, and the vanishing viscosity property $\gamma_k \rightarrow 0$ is related to $\alpha_k \rightarrow 1$.

1.3.2. *Implicit discretization.* Implicit discretization of (5) gives

$$\frac{1}{h_k^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\gamma_k}{h_k}(x_k - x_{k-1}) + \nabla\Phi_{\mu_k}(x_{k+1}) = 0.$$

Equivalently, $x_{k+1} + h_k^2 \nabla\Phi_{\mu_k}(x_{k+1}) = x_k + (1 - \gamma_k h_k)(x_k - x_{k-1})$, which gives

$$x_{k+1} = (I + h_k^2 \nabla\Phi_{\mu_k})^{-1}(x_k + (1 - \gamma_k h_k)(x_k - x_{k-1})).$$

Using the resolvent equation (see [6] for more details) we obtain

$$\begin{cases} y_k &= x_k + (1 - \gamma_k h_k)(x_k - x_{k-1}) \\ x_{k+1} &= \frac{\mu_k}{\mu_k + h_k^2} y_k + \frac{h_k^2}{\mu_k + h_k^2} \text{PROX}_{(\mu_k + h_k^2)\Phi}(y_k). \end{cases}$$

That's algorithm (RIPA) with $\alpha_k = 1 - \gamma_k h_k$, $\rho_k = \frac{h_k^2}{\mu_k + h_k^2}$, and proximal parameter $\mu_k + h_k^2$.

1.4. **Inertia and relaxation.** In his pioneering work [30], Polyak introduced the Heavy Ball method in order to speed up the classical gradient algorithm. Later Alvarez [1] studied the Inertial Proximal algorithm, which can be viewed as an implicit version of the Heavy Ball method. The extension from the subgradient case to the maximally monotone case was first considered by Alvarez-Attouch [3]. Recently, an extensive literature has been devoted to inertial methods, in relation with their acceleration properties.

Relaxation has proven to be an essential ingredient in the resolution of monotone inclusions, see Bauschke-Combettes [12], Eckstein-Bertsekas [19]. After having reformulated the problem in the form of a fixed point, it makes it possible to use Krasnoselskii-Mann theorem. Without using inertia, over-relaxation provides a natural way to speed up the algorithm. By contrast, for the resolution of monotone inclusions by inertial methods, we will see that under-relaxation allows to balance the inertial extrapolation effect. This makes it play a crucial role in the convergence of inertial proximal methods for general monotone inclusions. Let us mention some related contributions: Alvarez [2], Attouch-Cabot [6], Attouch-Peypouquet [9], Bot-Csetnek [15], Maingé [24]. In [20], Iutzeler-Hendrickx compare the numerical performances of inertia and relaxation techniques.

2. CONVERGENCE RATE OF THE VALUES FOR (RIPA)

Throughout this section, we assume that $A = \partial\Phi$, and make the following assumptions on the function Φ and the sequences (α_k) , (ρ_k) , (μ_k) :

$$(H) \left\{ \begin{array}{l} \bullet \Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\} \text{ is a convex lower semicontinuous proper function;} \\ \bullet S := \operatorname{argmin}\Phi \neq \emptyset; \\ \bullet (\alpha_k) \text{ is a nonnegative sequence;} \\ \bullet 0 < \rho_k \leq 1 \text{ for all } k \geq 1; \\ \bullet (\mu_k) \text{ and } (\rho_k \mu_k) \text{ are nondecreasing sequences of positive numbers.} \end{array} \right.$$

For each $k \geq 1$, we write briefly Φ_k for Φ_{μ_k} , and set $s_k = \rho_k \mu_k$. Therefore, (RIPA) is written in an equivalent way

$$(7) \quad \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = y_k - s_k \nabla \Phi_k(y_k). \end{cases}$$

A key point of our study will be to control the variations of Φ_k with respect to k . To do this, we will use the fact that the application $\mu \mapsto \Phi_\mu$ is nonincreasing (a direct consequence of its definition), and assume that $k \mapsto \mu_k$ is nondecreasing (assumption (H)).

2.1. Preliminary results. Let us establish some lemmas concerning the energy sequence (W_k) and the sequence of anchor terms (h_k) . They will play a central role in the Lyapunov analysis of (RIPA).

2.1.1. Descent rule. As a key classical ingredient, we use the descent rule for the gradient method, that we recall below, see [13, Lemma 2.3], [17, Lemma 1].

Lemma 2.1. *(descent rule) Suppose that $\Theta : \mathcal{H} \rightarrow \mathbb{R}$ is a convex, differentiable function, whose gradient is L -Lipschitz continuous. Then for any $0 < s \leq \frac{1}{L}$, for all $x, y \in \mathcal{H}$, we have*

$$\Theta(y - s\nabla\Theta(y)) \leq \Theta(x) + \langle \nabla\Theta(y), y - x \rangle - \frac{s}{2} \|\nabla\Theta(y)\|^2.$$

Note that $\nabla\Phi_k$ is Lipschitz continuous with the Lipschitz constant $\frac{1}{\mu_k}$. We have $s_k = \rho_k \mu_k \leq \mu_k$, therefore the descent rule applies for Φ_k with the step size s_k . So, the following inequality is true for all $x, y \in \mathcal{H}$, and all $k \geq 1$

$$(8) \quad \Phi_k(y - s_k \nabla \Phi_k(y)) \leq \Phi_k(x) + \langle \nabla \Phi_k(y), y - x \rangle - \frac{s_k}{2} \|\nabla \Phi_k(y)\|^2.$$

2.1.2. Energy sequence (W_k) . Let us introduce the sequence (W_k)

$$W_k := \Phi_k(x_k) - \min \Phi + \frac{1}{2s_k} \|x_k - x_{k-1}\|^2.$$

The term W_k is naturally interpreted as the global mechanical energy (potential + kinetic) at the stage k . Since $\inf \Phi_k = \inf \Phi$, we have $\Phi_k(x_k) - \min \Phi \geq 0$. Therefore, W_k is nonnegative, as the sum of two nonnegative terms. Let us evaluate the energy decay. The following result is consistent with the fact that the friction effect, and hence the dissipation of mechanical energy, is related to $\alpha_k \leq 1$.

Proposition 2.2. *Let us make assumption (H). Let (x_k) be a sequence generated by the algorithm (RIPA). Then, the energy sequence (W_k) satisfies for every $k \geq 1$,*

$$(9) \quad W_{k+1} - W_k \leq -\frac{1 - \alpha_k^2}{2s_k} \|x_k - x_{k-1}\|^2.$$

As a consequence, the sequence (W_k) is nonincreasing if $\alpha_k \in [0, 1]$ for every $k \geq 1$.

Proof. By applying formula (8) with $y = y_k$ and $x = x_k$, we obtain

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

Since $x_{k+1} = y_k - s_k \nabla \Phi_k(y_k)$ and $y_k - x_k = \alpha_k(x_k - x_{k-1})$, this implies that

$$\Phi_k(x_{k+1}) \leq \Phi_k(x_k) - \frac{1}{2s_k} \|x_{k+1} - x_k\|^2 + \frac{\alpha_k^2}{2s_k} \|x_k - x_{k-1}\|^2.$$

Equivalently

$$\Phi_k(x_{k+1}) + \frac{1}{2s_k} \|x_{k+1} - x_k\|^2 \leq \Phi_k(x_k) + \frac{1}{2s_k} \|x_k - x_{k-1}\|^2 + \left(\frac{\alpha_k^2}{2s_k} - \frac{1}{2s_k} \right) \|x_k - x_{k-1}\|^2.$$

By assumption (H), the sequences $(s_k) = (\rho_k \mu_k)$ and (μ_k) are nondecreasing. These properties respectively imply that $\frac{1}{2s_{k+1}} \leq \frac{1}{2s_k}$ and $\Phi_{k+1} \leq \Phi_k$. It follows that

$$\Phi_{k+1}(x_{k+1}) + \frac{1}{2s_{k+1}} \|x_{k+1} - x_k\|^2 \leq \Phi_k(x_k) + \frac{1}{2s_k} \|x_k - x_{k-1}\|^2 + \left(\frac{\alpha_k^2}{2s_k} - \frac{1}{2s_k} \right) \|x_k - x_{k-1}\|^2.$$

This can be equivalently rewritten as

$$W_{k+1} \leq W_k - \frac{1 - \alpha_k^2}{2s_k} \|x_k - x_{k-1}\|^2.$$

The last assertion is immediate. \square

2.1.3. *Anchor sequence (h_k) .* Let us now fix $x^* \in \mathcal{H}$, and consider the anchor sequence (h_k) which is defined by $h_k = \frac{1}{2} \|x_k - x^*\|^2$.

Proposition 2.3. *Let us make assumption (H). Let (x_k) be a sequence generated by the algorithm (RIPA). We have for every $k \geq 1$*

$$(10) \quad h_{k+1} - h_k - \alpha_k(h_k - h_{k-1}) = \frac{1}{2}(\alpha_k^2 + \alpha_k) \|x_k - x_{k-1}\|^2 - s_k \langle \nabla \Phi_{\mu_k}(y_k), y_k - x^* \rangle + \frac{s_k^2}{2} \|\nabla \Phi_{\mu_k}(y_k)\|^2.$$

If moreover $x^* \in \operatorname{argmin} \Phi$, then

$$h_{k+1} - h_k - \alpha_k(h_k - h_{k-1}) \leq \frac{1}{2}(\alpha_k^2 + \alpha_k) \|x_k - x_{k-1}\|^2 - s_k(\Phi_{\mu_k}(x_{k+1}) - \min \Phi).$$

Proof. Observe that

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

Setting briefly $A_k = h_{k+1} - h_k - \alpha_k(h_k - h_{k-1}) = \frac{1}{2}\|x_{k+1} - x^*\|^2 - [h_k + \alpha_k(h_k - h_{k-1})]$, we deduce that

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

Using the equality $x_{k+1} = y_k - s_k \nabla \Phi_{\mu_k}(y_k)$, we obtain (10).

Let us now assume that $x^* \in \operatorname{argmin} \Phi$. From inequality (8) applied with $y = y_k$ and $x = x^*$ (recall the simplified notation $\Phi_k = \Phi_{\mu_k}$)

$$\Phi_k(x_{k+1}) = \Phi_k(y_k - s_k \nabla \Phi_k(y_k)) \leq \Phi_k(x^*) + \langle \nabla \Phi_k(y_k), y_k - x^* \rangle - \frac{s_k}{2} \|\nabla \Phi_k(y_k)\|^2.$$

Since $\Phi_k(x^*) = \min \Phi$, we infer that

$$-s_k \langle \nabla \Phi_k(y_k), y_k - x^* \rangle + \frac{s_k^2}{2} \|\nabla \Phi_k(y_k)\|^2 \leq -s_k(\Phi_k(x_{k+1}) - \min \Phi),$$

which completes the proof of Proposition 2.3. \square

2.2. Convergence rates for the values. Given $x^* \in \operatorname{argmin} \Phi$, let us define (\mathcal{E}_k) by, for each $k \geq 1$

$$(11) \quad \mathcal{E}_k = t_k^2(\Phi_{\mu_k}(x_k) - \min \Phi) + \frac{1}{2\rho_k\mu_k}\|x_{k-1} + t_k(x_k - x_{k-1}) - x^*\|^2.$$

With our simplified notations $\Phi_k = \Phi_{\mu_k}$, and $s_k = \rho_k\mu_k$, the energy \mathcal{E}_k writes

$$\mathcal{E}_k = t_k^2(\Phi_k(x_k) - \min \Phi) + \frac{1}{2s_k}\|x_{k-1} + t_k(x_k - x_{k-1}) - x^*\|^2.$$

The next result shows that the sequence (\mathcal{E}_k) is nonincreasing, under some suitable condition (K_1) . The following statement is an adaptation of [13, Lemma 4.1] to our setting.

Proposition 2.4. *Let us make assumptions (H) and (K_0) . Let (x_k) be a sequence generated by the algorithm (RIPA), and let (\mathcal{E}_k) be the sequence defined by (11). Then we have*

$$(12) \quad \mathcal{E}_{k+1} - \mathcal{E}_k \leq (t_{k+1}^2 - t_k^2 - t_{k+1})(\Phi_{\mu_k}(x_k) - \min \Phi).$$

Under the assumption

$$(K_1) \quad t_{k+1}^2 - t_k^2 \leq t_{k+1} \quad \text{for every } k \geq 1,$$

then the sequence (\mathcal{E}_k) is nonincreasing.

Proof. Let us write successively formula (8) at $y = y_k$ and $x = x_k$, then at $y = y_k$ and $x = x^*$. Recalling that $x_{k+1} = y_k - s_k \nabla \Phi_k(y_k)$, we obtain the two inequalities

$$(13) \quad \Phi_k(x_{k+1}) \leq \Phi_k(x_k) + \langle \nabla \Phi_k(y_k), y_k - x_k \rangle - \frac{s_k}{2} \|\nabla \Phi_k(y_k)\|^2,$$

$$(14) \quad \Phi_k(x_{k+1}) \leq \min \Phi + \langle \nabla \Phi_k(y_k), y_k - x^* \rangle - \frac{s_k}{2} \|\nabla \Phi_k(y_k)\|^2.$$

Multiplying (13) by $t_{k+1} - 1 \geq 0$, then adding (14), we derive that

$$(15) \quad \begin{aligned} t_{k+1}\Phi_k(x_{k+1}) &\leq (t_{k+1} - 1)\Phi_k(x_k) + \min \Phi \\ &\quad + \langle \nabla \Phi_k(y_k), (t_{k+1} - 1)(y_k - x_k) + y_k - x^* \rangle - \frac{s_k}{2} t_{k+1} \|\nabla \Phi_k(y_k)\|^2. \end{aligned}$$

Observe that

$$\begin{aligned}
(t_{k+1} - 1)(y_k - x_k) + y_k &= t_{k+1} y_k - (t_{k+1} - 1)x_k \\
&= x_k + t_{k+1} \alpha_k (x_k - x_{k-1}) \\
&= x_{k-1} + (1 + t_{k+1} \alpha_k)(x_k - x_{k-1}) \\
&= x_{k-1} + t_k (x_k - x_{k-1}) \quad \text{in view of (3)}.
\end{aligned}$$

Setting $z_k = x_{k-1} + t_k(x_k - x_{k-1})$, we then deduce from (15) that

$$(16) \quad t_{k+1}(\Phi_k(x_{k+1}) - \min \Phi) \leq (t_{k+1} - 1)(\Phi_k(x_k) - \min \Phi) + \langle \nabla \Phi_k(y_k), z_k - x^* \rangle - \frac{s_k}{2} t_{k+1} \|\nabla \Phi_k(y_k)\|^2.$$

On the other hand, observe that

$$\begin{aligned}
z_{k+1} - z_k &= x_k + t_{k+1}(x_{k+1} - x_k) - x_{k-1} - t_k(x_k - x_{k-1}) \\
&= t_{k+1}(x_{k+1} - x_k) - (t_k - 1)(x_k - x_{k-1}) \\
&= t_{k+1}(x_{k+1} - x_k - \alpha_k(x_k - x_{k-1})) \quad \text{in view of (3)} \\
&= t_{k+1}(x_{k+1} - y_k) = -s_k t_{k+1} \nabla \Phi_k(y_k).
\end{aligned}$$

It ensues that $z_{k+1} - x^* = z_k - x^* - s_k t_{k+1} \nabla \Phi_k(y_k)$, which gives

$$\|z_{k+1} - x^*\|^2 = \|z_k - x^*\|^2 - 2s_k t_{k+1} \langle \nabla \Phi_k(y_k), z_k - x^* \rangle + s_k^2 t_{k+1}^2 \|\nabla \Phi_k(y_k)\|^2.$$

By using this equality in (16), we find

$$t_{k+1}(\Phi_k(x_{k+1}) - \min \Phi) \leq (t_{k+1} - 1)(\Phi_k(x_k) - \min \Phi) + \frac{1}{2s_k t_{k+1}} (\|z_k - x^*\|^2 - \|z_{k+1} - x^*\|^2),$$

which is equivalent to

$$(17) \quad t_{k+1}^2(\Phi_k(x_{k+1}) - \min \Phi) + \frac{1}{2s_k} \|z_{k+1} - x^*\|^2 \leq (t_{k+1}^2 - t_{k+1})(\Phi_k(x_k) - \min \Phi) + \frac{1}{2s_k} \|z_k - x^*\|^2.$$

By assumption (H), the sequences $(s_k) = (\rho_k \mu_k)$ and (μ_k) are nondecreasing. These properties respectively imply that $\frac{1}{s_{k+1}} \leq \frac{1}{s_k}$ and $\Phi_{k+1} \leq \Phi_k$. Therefore, we deduce from (17) that

$$t_{k+1}^2(\Phi_{k+1}(x_{k+1}) - \min \Phi) + \frac{1}{2s_{k+1}} \|z_{k+1} - x^*\|^2 \leq (t_{k+1}^2 - t_{k+1})(\Phi_k(x_k) - \min \Phi) + \frac{1}{2s_k} \|z_k - x^*\|^2.$$

Using the expression of the sequence (\mathcal{E}_k) , we obtain

$$\mathcal{E}_{k+1} \leq \mathcal{E}_k + (t_{k+1}^2 - t_k^2 - t_{k+1})(\Phi_k(x_k) - \min \Phi).$$

The last assertion is immediate. \square

We can now state our main result concerning the rate of convergence of the values for (RIPA).

Theorem 2.5. *Under (H), assume that the nonnegative sequence (α_k) satisfies (K_0) - (K_1) . Let (x_k) be a sequence generated by the algorithm (RIPA). Then we have*

(i) *For every $k \geq 1$,*

$$\Phi_{\mu_k}(x_k) - \min \Phi \leq \frac{C}{t_k^2},$$

with $C = t_1^2(\Phi_{\mu_1}(x_1) - \min \Phi) + \frac{1}{\rho_1 \mu_1} (d(x_0, S)^2 + t_1^2 \|x_1 - x_0\|^2)$.

As a consequence, setting $p_k = \text{prox}_{\mu_k \Phi}(x_k)$, we have

$$\Phi(p_k) - \min \Phi = \mathcal{O}\left(\frac{1}{t_k^2}\right) \quad \text{and} \quad \|x_k - p_k\|^2 = \mathcal{O}\left(\frac{\mu_k}{t_k^2}\right) \quad \text{as } k \rightarrow +\infty.$$

(ii) Assume moreover that there exists $0 \leq m < 1$ such that

$$(K_1^+) \quad t_{k+1}^2 - t_k^2 \leq m t_{k+1} \quad \text{for every } k \geq 1.$$

Then we have

$$\sum_{k=1}^{+\infty} t_{k+1}(\Phi_{\mu_k}(x_k) - \min \Phi) < +\infty.$$

Proof. (i) By Proposition 2.4, the sequence (\mathcal{E}_k) is nonincreasing. It ensues that $\mathcal{E}_k \leq \mathcal{E}_1$ for every $k \geq 1$. Recalling the expression of \mathcal{E}_k , we deduce that

$$\begin{aligned} t_k^2(\Phi_{\mu_k}(x_k) - \min \Phi) &\leq \mathcal{E}_1 = t_1^2(\Phi_{\mu_1}(x_1) - \min \Phi) + \frac{1}{2\rho_1\mu_1} \|x_0 + t_1(x_1 - x_0) - x^*\|^2 \\ &\leq t_1^2(\Phi_{\mu_1}(x_1) - \min \Phi) + \frac{1}{\rho_1\mu_1} (\|x_0 - x^*\|^2 + t_1^2\|x_1 - x_0\|^2). \end{aligned}$$

Since x^* can be taken arbitrarily in $S = \operatorname{argmin} \Phi$, we obtain $t_k^2(\Phi_{\mu_k}(x_k) - \min \Phi) \leq C$, with

$$C = t_1^2(\Phi_{\mu_1}(x_1) - \min \Phi) + \frac{1}{\rho_1\mu_1} (d(x_0, S)^2 + t_1^2\|x_1 - x_0\|^2).$$

(ii) By summing inequality (12) from $k = 1$ to n , we find

$$\mathcal{E}_{n+1} + \sum_{k=1}^n (t_{k+1} - t_{k+1}^2 + t_k^2)(\Phi_{\mu_k}(x_k) - \min \Phi) \leq \mathcal{E}_1.$$

Since $\mathcal{E}_{n+1} \geq 0$ and since $t_{k+1}^2 - t_k^2 \leq m t_{k+1}$, this implies that

$$(1 - m) \sum_{k=1}^n t_{k+1}(\Phi_{\mu_k}(x_k) - \min \Phi) \leq \mathcal{E}_1.$$

The expected estimate is obtained by letting n tend to infinity. \square

Remark 2.6. By contrast with the classical inertial gradient methods, the rate of convergence of the values for (RIPA) is expressed in terms of the proximal sequence (p_k) , with $p_k = \operatorname{prox}_{\mu_k \Phi}(x_k)$, and not directly in terms of (x_k) . This has some analogy with the convergence properties of the shadow sequence in the Douglas-Rachford algorithm.

Remark 2.7. Let us assume that x_1 belongs to the domain of Φ , that is $\Phi(x_1) < +\infty$. Then $\Phi_{\mu_1}(x_1) \leq \Phi(x_1) < +\infty$, and Theorem 2.5 gives, for every $k \geq 1$

$$\Phi_{\mu_k}(x_k) - \min \Phi \leq \frac{\mathcal{E}_1(x_0, x_1)}{t_k^2}$$

where

$$\mathcal{E}_1(x_0, x_1) = t_1^2[\Phi(x_1) - \min \Phi] + \frac{1}{\rho_1\mu_1} \|x_0 - x^* + t_1(x_1 - x_0)\|^2.$$

As a function of (x_0, x_1) , \mathcal{E}_1 achieves its minimum when $x_1 \in \operatorname{argmin} \Phi$ and $x_1 - x_0 = \frac{1}{t_1}(x^* - x_0)$. Of course, taking $x_1 \in \operatorname{argmin} \Phi$ is not realistic, since this would mean that the problem is already solved. But this suggests taking the initial direction $x_1 - x_0$ as a multiple of an approximation of $x^* - x_0$, such as the opposite of the gradient $\nabla \Phi_{\mu_0}(x_0)$. This amounts to take a gradient step at the first stage.

2.3. Convergence rate of the velocities to zero. Let us first show estimates in a summation form.

Proposition 2.8. *Under (H), assume that the sequence (α_k) satisfies (K_0) - (K_1^+) . Let (x_k) be a sequence generated by the algorithm (RIPA). Then we have*

$$(18) \quad \sum_{k=1}^{+\infty} \frac{t_k}{s_k} \|x_k - x_{k-1}\|^2 < +\infty.$$

Moreover

$$\sum_{k=1}^{+\infty} \frac{t_{k+1}}{s_k} \|x_k - x_{k-1}\|^2 < +\infty,$$

and hence

$$\sum_{k=1}^{+\infty} t_{k+1} W_k < +\infty.$$

Proof. Let us recall the inequality (9) from Proposition 2.2

$$W_{k+1} - W_k \leq -\frac{1 - \alpha_k^2}{2s_k} \|x_k - x_{k-1}\|^2.$$

Multiplying this inequality by t_{k+1}^2 and summing from $k = 1$ to n , we obtain

$$\sum_{k=1}^n t_{k+1}^2 (W_{k+1} - W_k) + \sum_{k=1}^n \frac{1}{2s_k} t_{k+1}^2 (1 - \alpha_k^2) \|x_k - x_{k-1}\|^2 \leq 0.$$

By rearranging the terms (we perform a discrete form of the integration by parts formula), we find

$$t_{n+1}^2 W_{n+1} + \sum_{k=1}^n (t_k^2 - t_{k+1}^2) W_k + \sum_{k=1}^n \frac{1}{2s_k} t_{k+1}^2 (1 - \alpha_k^2) \|x_k - x_{k-1}\|^2 \leq t_1^2 W_1.$$

Recalling the expression of W_k , we deduce that

$$\sum_{k=1}^n \frac{1}{2s_k} [t_k^2 - t_{k+1}^2 + t_{k+1}^2 (1 - \alpha_k^2)] \|x_k - x_{k-1}\|^2 \leq t_1^2 W_1 + \sum_{k=1}^n (t_{k+1}^2 - t_k^2) (\Phi_k(x_k) - \min \Phi).$$

Since $t_{k+1} \alpha_k = t_k - 1$ and $t_{k+1}^2 - t_k^2 \leq t_{k+1}$ by assumption (K_1) , this implies

$$\sum_{k=1}^n \frac{1}{2s_k} [t_k^2 - (t_k - 1)^2] \|x_k - x_{k-1}\|^2 \leq t_1^2 W_1 + \sum_{k=1}^n t_{k+1} (\Phi_k(x_k) - \min \Phi).$$

Observing that $t_k \geq 1$, we have

$$t_k^2 - (t_k - 1)^2 = 2t_k - 1 \geq t_k.$$

Hence,

$$\sum_{k=1}^n \frac{1}{2s_k} t_k \|x_k - x_{k-1}\|^2 \leq t_1^2 W_1 + \sum_{k=1}^n t_{k+1} (\Phi_k(x_k) - \min \Phi).$$

By letting n tend to infinity in the above inequality we obtain

$$\sum_{k=1}^{+\infty} \frac{t_k}{s_k} \|x_k - x_{k-1}\|^2 \leq 2t_1^2 W_1 + 2 \sum_{k=1}^{+\infty} t_{k+1} (\Phi_k(x_k) - \min \Phi) < +\infty,$$

which gives the conclusion (18). Let us complete this result by observing that assumption (K_1) implies

$$t_{k+1} - t_k \leq \frac{t_{k+1}}{t_{k+1} + t_k} \leq 1.$$

Since $t_k \geq 1$, we deduce that $t_{k+1} \leq 2t_k$ for every $k \geq 1$. The estimate (18) then yields

$$\sum_{k=1}^{+\infty} \frac{t_{k+1}}{s_k} \|x_k - x_{k-1}\|^2 < +\infty.$$

Recalling from Theorem 2.5 that $\sum_{k=1}^{+\infty} t_{k+1}(\Phi_k(x_k) - \min \Phi) < +\infty$, we obtain $\sum_{k=1}^{+\infty} t_{k+1} W_k < +\infty$. \square

Let us now establish pointwise estimates for the rate of convergence of the velocities.

Theorem 2.9. *Under (H), assume that the sequence (α_k) satisfies (K_0) - (K_1^+) , and $\alpha_k \in [0, 1]$ for every $k \geq 1$. Then for any sequence (x_k) generated by the algorithm (RIPA), the following holds true*

$$(19) \quad \Phi_{\mu_k}(x_k) - \min \Phi = o\left(\frac{1}{\sum_{i=1}^k t_i}\right) \quad \text{and} \quad \|x_k - x_{k-1}\| = o\left(\frac{s_k}{\sum_{i=1}^k t_i}\right)^{\frac{1}{2}} \quad \text{as } k \rightarrow +\infty,$$

thus implying

$$(20) \quad \Phi(p_k) - \min \Phi = o\left(\frac{1}{\sum_{i=1}^k t_i}\right) \quad \text{and} \quad \|x_k - p_k\| = o\left(\frac{\mu_k}{\sum_{i=1}^k t_i}\right)^{\frac{1}{2}} \quad \text{as } k \rightarrow +\infty.$$

As a consequence, we have

$$\Phi(p_k) - \min \Phi = o\left(\frac{1}{t_k^2}\right), \quad \|x_k - p_k\| = o\left(\frac{\sqrt{\mu_k}}{t_k}\right) \quad \text{and} \quad \|x_k - x_{k-1}\| = o\left(\frac{\sqrt{s_k}}{t_k}\right) \quad \text{as } k \rightarrow +\infty.$$

Proof. According to Proposition 2.2 and $\alpha_k \in [0, 1]$ for all $k \geq 1$, we deduce that the energy sequence (W_k) is nonincreasing. By Proposition 2.8, we have that $\sum_{k=1}^{+\infty} t_{k+1} W_k < +\infty$. Since (W_k) is nonincreasing, this implies that $\sum_{k=1}^{+\infty} t_{k+1} W_{k+1} < +\infty$, hence $\sum_{k=1}^{+\infty} t_k W_k < +\infty$. Let us apply Lemma B.2 in the Appendix, with the sequences (t_k) and (W_k) , respectively in place of (τ_k) and (ε_k) . We obtain that

$$W_k = o\left(\frac{1}{\sum_{i=1}^k t_i}\right) \quad \text{as } k \rightarrow +\infty.$$

The estimates (19) follow immediately. From the definition of Φ_{μ_k} , we clearly deduce (20). In view of assumption (K_1) , we have $t_{i+1}^2 - t_i^2 \leq t_{i+1}$ for every $i \geq 1$, hence by summing from $i = 1$ to $k - 1$

$$t_k^2 \leq t_1^2 + \sum_{i=1}^{k-1} t_{i+1} = t_1^2 - t_1 + \sum_{i=1}^k t_i.$$

We easily deduce the last estimates. \square

2.4. Convergence of the sequence of iterates (x_k) .

Theorem 2.10. *Under (H), assume that the sequence (α_k) satisfies (K_0) - (K_1^+) , together with $\alpha_k \in [0, 1]$ for every $k \geq 1$. Assume moreover that*

$$\sup_k \frac{\mu_k}{\sum_{i=1}^k t_i} < +\infty \quad \text{and} \quad \sup_k \rho_k \mu_k < +\infty.$$

Then, for any sequence (x_k) generated by algorithm (RIPA), the following holds

$$(i) \quad \lim_{k \rightarrow +\infty} \|x_k - p_k\| = 0, \quad \text{where } p_k = \text{prox}_{\mu_k \Phi}(x_k);$$

(ii) (x_k) converges weakly, as $k \rightarrow +\infty$, to some $x^* \in \operatorname{argmin} \Phi$.

Proof. (i) By the second estimate of (20) we have

$$\|x_k - p_k\| = o\left(\frac{\mu_k}{\sum_{i=1}^k t_i}\right)^{\frac{1}{2}} \rightarrow 0 \quad \text{as } k \rightarrow +\infty,$$

because $\sup_k \frac{\mu_k}{\sum_{i=1}^k t_i} < +\infty$ by assumption.

(ii) Let us verify successively the items (i) and (ii) of the Opial lemma, see Lemma B.1 in the Appendix. Take $S = \operatorname{argmin} \Phi$. From the first estimate of (20), the sequence (p_k) is minimizing, which implies that any weak cluster point of (p_k) belongs to S . Since $\lim \|x_k - p_k\| = 0$, the same property holds true for the sequence (x_k) , which completes item (i) of Opial's lemma. Let us fix $x^* \in S$, and show that $\lim_{k \rightarrow +\infty} \|x_k - x^*\|$ exists. For that purpose, let us set $h_k = \frac{1}{2} \|x_k - x^*\|^2$. By Proposition 2.3, the sequence (h_k) satisfies the following inequalities

$$\begin{aligned} h_{k+1} - h_k - \alpha_k(h_k - h_{k-1}) &\leq \frac{1}{2}(\alpha_k^2 + \alpha_k)\|x_k - x_{k-1}\|^2 - s_k(\Phi_{\mu_k}(x_{k+1}) - \min \Phi) \\ &\leq \frac{1}{2}(\alpha_k^2 + \alpha_k)\|x_k - x_{k-1}\|^2 \\ &\leq \|x_k - x_{k-1}\|^2 \quad \text{since } \alpha_k \in [0, 1]. \end{aligned}$$

Taking the positive part, we find

$$(h_{k+1} - h_k)_+ \leq \alpha_k(h_k - h_{k-1})_+ + \|x_k - x_{k-1}\|^2.$$

By Proposition 2.8 we have $\sum_{k=1}^{+\infty} \frac{t_{k+1}}{s_k} \|x_k - x_{k-1}\|^2 < +\infty$. Since (s_k) has been supposed to be bounded from above, this implies $\sum_{k=1}^{+\infty} t_{k+1} \|x_k - x_{k-1}\|^2 < +\infty$. By applying Lemma B.3 (given in the Appendix) with $a_k = (h_k - h_{k-1})_+$ and $\omega_k = \|x_k - x_{k-1}\|^2$, we obtain

$$\sum_{k=1}^{+\infty} (h_k - h_{k-1})_+ < +\infty.$$

Since (h_k) is nonnegative, this classically implies that $\lim_{k \rightarrow +\infty} h_k$ exists. The second point of the Opial lemma is shown, which ends the proof. \square

2.5. Geometrical aspects. The results obtained in the previous section deal with general convex lower semicontinuous proper functions $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$. They give convergence rates for values in the worst case. It is a well-documented fact that under additional geometric properties on Φ , one can expect better convergence rates. Let us mention [4], [7], [31] for results based on strong convexity, [23] for results based on partial smoothness, and [14, 21] for results based on the Kurdyka-Lojasiewicz property. For the Nesterov acceleration (vanishing damping coefficient of the form $\frac{\alpha}{t}$), [11] gives a recent account on these issues with optimal convergence rates.

Since (RIPA) is governed by $\nabla \Phi_{\mu_k}$ at iteration k , a natural question is to ask whether such geometric assumptions about Φ , are preserved by taking the Moreau envelopes. We will examine several cases for which we will answer this question affirmatively. This suggests that in these cases, it is likely that one can improve the rate of convergence that we obtained for (RIPA) in the worst case. This is an interesting question for future studies.

2.5.1. Strong convexity. We recall that $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ is strongly convex with constant $\beta > 0$ if $\Phi - \frac{\beta}{2} \|\cdot\|^2$ is convex. One can consult [12, Section 10.2] for classical results concerning strong convexity. First, as a direct consequence of Theorems 2.9 and 2.10, we obtain the strong convergence of the iterates in this case.

Proposition 2.11. *Let's make the hypotheses of Theorem 2.10, and suppose that $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex lower semicontinuous proper function that is strongly convex. Then, for any sequence (x_k) generated by (RIPA), (x_k) and (p_k) converge strongly, as $k \rightarrow +\infty$, to the unique minimizer x^* of Φ .*

Proof. By Theorem 2.9, we have $\Phi(p_k) - \min \Phi = o\left(\frac{1}{t_k^2}\right)$. Since $t_k \geq 1$, this implies $\Phi(p_k) \rightarrow \min \Phi$. Hence, (p_k) is a minimizing sequence of the strongly convex function Φ . This classically implies that (p_k) converges strongly to the unique minimizer x^* of Φ . By Theorem 2.10, we have $\lim_{k \rightarrow +\infty} \|x_k - p_k\| = 0$, which implies that, likewise, (x_k) converges strongly, as $k \rightarrow +\infty$, to the unique minimizer x^* of Φ . \square

Proposition 2.12. *Suppose that $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex lower semicontinuous proper function. Then, the following implications hold: for any $\beta, \gamma > 0$ and $\mu > 0$*

- (i) Φ strongly convex with constant $\beta \implies \Phi_\mu$ strongly convex with constant $\frac{\beta}{1 + \beta\mu}$.
- (ii) Φ_μ strongly convex with constant $\gamma < \frac{1}{\mu} \implies \Phi$ strongly convex with constant $\frac{\gamma}{1 - \gamma\mu}$.

Proof. (i) If Φ is strongly convex with constant $\beta > 0$, we have $\Phi = \Psi + \frac{\beta}{2}\|\cdot\|^2$ for some convex function Ψ . Elementary calculus (see [12, Exercise 12.6] for example) gives, with $\theta = \frac{\mu}{1 + \beta\mu}$,

$$\Phi_\mu(x) = \Psi_\theta\left(\frac{1}{1 + \mu\beta}x\right) + \frac{\beta}{2(1 + \beta\mu)}\|x\|^2.$$

Since $x \mapsto \Psi_\theta\left(\frac{1}{1 + \mu\beta}x\right)$ is a convex function, the above formula shows that Φ_μ is strongly convex with constant $\gamma = \frac{\beta}{1 + \beta\mu}$. Note that $\gamma\mu = \frac{\beta\mu}{1 + \beta\mu} < 1$, which gives $\gamma < \frac{1}{\mu}$.

(ii) Conversely, suppose that $\Phi_\mu = G + \frac{\gamma}{2}\|\cdot\|^2$ for some convex function G and some $\gamma > 0$ such that $\gamma\mu < 1$. Since the Moreau envelope Φ_μ is of class \mathcal{C}^1 , the function G is also of class \mathcal{C}^1 . Taking the Legendre-Fenchel conjugate, we obtain $(\Phi_\mu)^* = (G + \frac{\gamma}{2}\|\cdot\|^2)^*$. Using classical calculus rules, we have

$$(\Phi_\mu)^* = \left(\Phi \# \frac{1}{2\mu}\|\cdot\|^2\right)^* = \Phi^* + \frac{\mu}{2}\|\cdot\|^2, \text{ and } \left(G + \frac{\gamma}{2}\|\cdot\|^2\right)^* = G^* \# \frac{1}{2\gamma}\|\cdot\|^2 = (G^*)_\gamma,$$

where $\#$ denotes the infimal convolution (also called epi-addition). It ensues that $\Phi^* = (G^*)_\gamma - \frac{\mu}{2}\|\cdot\|^2$. Since Φ is convex lower semicontinuous and proper, we have $\Phi = \Phi^{**} = ((G^*)_\gamma - \frac{\mu}{2}\|\cdot\|^2)^*$.

By definition of the Legendre-Fenchel conjugate and the infimal convolution, this gives

$$\Phi(x) = \sup_{\xi} \sup_u \left\{ \langle x, \xi \rangle - G^*(u) - \frac{1}{2\gamma}\|\xi - u\|^2 + \frac{\mu}{2}\|\xi\|^2 \right\}.$$

Elementary calculus (switch order of supremum in the above expression) gives $\Phi = J + \frac{\gamma}{2(1 - \gamma\mu)}\|\cdot\|^2$, where J is convex lower semicontinuous and proper. Hence Φ is strongly convex with constant $\frac{\gamma}{1 - \gamma\mu}$. \square

2.5.2. Growth condition. Let us consider $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ a convex lower semicontinuous proper function that satisfies the growth condition: there exists some positive constant β such that for all $x \in \mathcal{H}$

$$\Phi(x) \geq \min \Phi + \frac{\beta}{2}d(x, S)^2,$$

where $S = \operatorname{argmin} \Phi$. When Φ is convex, which is our case, it is equivalent to a Kurdyka-Lojasiewicz inequality. This makes it play an important role in the mathematical analysis of continuous and discrete dynamical systems. It is also called conditioning or Hölderian error bounds. Note that

$\frac{1}{2}d(x, S)^2 = (\delta_S)_1$ where δ_S is the indicator function of the set S . The resolvent equality (semi-group property) gives

$$\left(\frac{1}{2}d(\cdot, S)^2\right)_\mu = ((\delta_S)_1)_\mu = (\delta_S)_{1+\mu} = \frac{1}{2(1+\mu)}d(\cdot, S)^2.$$

From this, it is an easy exercise to verify that for such a function Φ , the Moreau envelope Φ_μ satisfies

$$\Phi_\mu(x) \geq \min \Phi + \frac{\beta}{2(1+\beta\mu)}d(x, S)^2.$$

Since $\min \Phi = \min \Phi_\mu$, we deduce that Φ_μ satisfies a similar growth condition (with another constant).

2.6. Related algorithms. The analysis of (RIPA) performed in the previous sections is based on its formulation as an inertial gradient algorithm

$$(21) \quad \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = y_k - s_k \nabla \Phi_k(y_k), \end{cases}$$

where, as a distinctive feature, the potential function Φ_k varies at each iteration. A careful inspection at the proofs shows that the results listed in section 2 up to the first part of Theorem 2.9 are valid in a more general context than that of Moreau-Yosida regularization. As part of the algorithm (21), these results remain true for sequences (Φ_k) and (s_k) that meet the following conditions:

- (Φ_k) is a nonincreasing sequence of differentiable convex functions;
- $\nabla \Phi_k$ is L_k -Lipschitzian;
- $\operatorname{argmin} \Phi_k$ is not empty and $\min \Phi_k \equiv m$ for all $k \geq 1$;
- s_k is a nondecreasing sequence such that $s_k \leq \frac{1}{L_k}$ for all $k \geq 1$.

This suggests a natural extension, which consists of replacing the quadratic regularization term of the Moreau envelopes with a general Bregman distance. This is another direction for future research.

3. APPLICATION TO PARTICULAR CLASSES OF PARAMETERS α_k , μ_k AND ρ_k

We will now specialize the parameters α_k , μ_k and ρ_k in order to obtain inertial proximal-based algorithms whose iterations converge in the case of general monotone inclusions, and which, in the case of convex minimization, give fast convergence rates of values (in the worst case).

Theorem 1.3 provides convergence of the iterates for general maximally monotone operators, and Theorem 2.5 gives fast convergence of the values for convex minimization. So, each of them meets one of the two above objectives, and we just need to synthesize them. This amounts to finding assumptions about the data (α_k) , (ρ_k) , (μ_k) for which both results are valid. Note that apart from the (K_0) assumption common to both, and which is basic to define the sequence (t_k) , each of these results is based on a specific set of assumptions. For general monotone inclusions, we use Theorem 1.3, which concerns the case $\rho_k \rightarrow 0$. Indeed, when $\alpha_k \rightarrow 1$, which is the situation for which one can expect rapid convergence results of the values for convex minimization, a direct inspection of the condition (L) shows that we must have $\rho_k \rightarrow 0$. The assumptions of Theorem 1.3 and Theorem 2.5 are expressed in terms of the sequences (α_k) , (ρ_k) , (μ_k) , and (t_k) . Only the sequence (t_k) is not given explicitly from the data. The following proposition provides a criterion for simply obtaining an asymptotic equivalent of (t_k) , see [4, Propositions 14 and 15]. It also guarantees that the property (K_0) is satisfied.

Proposition 3.1. *Let (α_k) be a sequence such that $\alpha_k \in [0, 1[$ for every $k \geq 1$. Assume that*

$$(22) \quad \lim_{k \rightarrow +\infty} \left(\frac{1}{1 - \alpha_{k+1}} - \frac{1}{1 - \alpha_k} \right) = c,$$

for some $c \in [0, 1[$. Then the property (K_0) is satisfied, and $t_{k+1} \sim \frac{1}{(1-c)(1-\alpha_k)}$ as $k \rightarrow +\infty$.

The following result gives a practical form of Theorem 1.3 that relies on Proposition 3.1. The conditions are now expressed simply and directly in terms of the sequences (α_k) , (ρ_k) , (μ_k) .

Theorem 3.2. [6, Theorem 3.6] *Let $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ be a maximally monotone operator such that $\text{zer}A \neq \emptyset$. Suppose that the sequences (α_k) and (ρ_k) satisfy $\alpha_k \in [0, 1[$ and $\rho_k \in]0, 2[$ for every $k \geq 1$. Let us assume that there exist $\bar{\alpha} \in [0, 1]$, $\bar{\rho} \in [0, 2[$, $c \in [0, 1[$ and $c'' \in \mathbb{R}$, with $-(1 - \bar{\rho}/2) < c'' \leq -(1 - \bar{\rho}/2)c$ such that*

$$(23) \quad \lim_{k \rightarrow +\infty} \alpha_k = \bar{\alpha} \quad \text{and} \quad \lim_{k \rightarrow +\infty} \rho_k = \bar{\rho};$$

$$(24) \quad \lim_{k \rightarrow +\infty} \left(\frac{1}{1 - \alpha_{k+1}} - \frac{1}{1 - \alpha_k} \right) = c;$$

$$(25) \quad \lim_{k \rightarrow +\infty} \frac{\rho_{k+1} - \rho_k}{\rho_{k+1}(1 - \alpha_k)} = c'';$$

$$(26) \quad \liminf_{k \rightarrow +\infty} \frac{(1 - \alpha_k)^2}{\rho_k} > \frac{\bar{\alpha}(1 + \bar{\alpha})}{2 - \bar{\rho} + 2c''}.$$

Then for any sequence (x_k) generated by (RIPA), we have

$$(i) \quad \|x_{k+1} - x_k\| = \mathcal{O}\left(\frac{\rho_k}{1 - \alpha_k}\right) \quad \text{as } k \rightarrow +\infty.$$

Assume additionally that $\frac{|\mu_{k+1} - \mu_k|}{\mu_{k+1}} = \mathcal{O}\left(\frac{\rho_k}{1 - \alpha_k}\right)$ as $k \rightarrow +\infty$, together with $\sum_{k=1}^{+\infty} \frac{\rho_k}{1 - \alpha_k} = +\infty$.

Then the following holds

$$(ii) \quad \lim_{k \rightarrow +\infty} \mu_k A_{\mu_k}(x_k) = 0. \quad \text{If } \liminf_{k \rightarrow +\infty} \mu_k > 0, \text{ then there exists } x_\infty \in \text{zer}A \text{ such that } x_k \rightharpoonup x_\infty \text{ weakly in } \mathcal{H} \text{ as } k \rightarrow +\infty.$$

3.1. The case $\alpha_k = 1 - \frac{\alpha}{k}$. In a seminal paper [25], Nesterov considered the accelerated gradient method

$$\begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = y_k - \lambda_k \nabla \Phi(y_k), \end{cases}$$

which aims at minimizing a differentiable convex function $\Phi : \mathcal{H} \rightarrow \mathbb{R}$. It is assumed that Φ has a L -Lipschitz gradient, and that the parameters satisfy $\lambda_k \equiv 1/L$, and $\alpha_k = \frac{t_k - 1}{t_{k+1}}$ with $t_1 = 1$ and $t_{k+1} = \frac{\sqrt{4t_k^2 + 1} + 1}{2}$ (this corresponds to the case of equality in (K_1)). This choice leads to an increasing sequence of extrapolated coefficients (α_k) , that behaves like $1 - \frac{3}{k}$ as $k \rightarrow +\infty$. This scheme exhibits the convergence rate for the values $\Phi(x_k) - \min \Phi = \mathcal{O}\left(\frac{1}{k^2}\right)$ as $k \rightarrow +\infty$, which, for gradient methods, is known to be optimal among all first-order methods having only information about the gradients and functions at previous iterates [26]. Then, Beck-Teboulle [13] extended Nesterov's method to inertial proximal-gradient methods for solving structured minimization problems, that's (FISTA). Other choices of the extrapolation factors α_k have been recently considered in [7, 8, 10, 17, 31]. Notably, the case $\alpha_k = 1 - \frac{\alpha}{k}$, with $\alpha > 3$ offers many advantages. First, it ensures the convergence of the sequences (x_k) , as proved by Chambolle and Dossal [17], see also [7]. Let us recall that the convergence of the sequences generated by (FISTA) has not been established so far. Second, as proved by Attouch and Peypouquet in [8], it provides the better rate of convergence $\Phi(x_k) - \min \Phi = o\left(\frac{1}{k^2}\right)$. A unified point of view on these methods has been lately brought to light by Attouch-Cabot [4], where a general extrapolation coefficient α_k is considered.

The following result is a direct application of Theorem 3.2 to the case $\alpha_k = 1 - \frac{\alpha}{k}$.

Corollary 3.3. [6, Corollary 3.8] *Let $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ be a maximally monotone operator such that $\text{zer}A \neq \emptyset$. Suppose that $\alpha_k = 1 - \frac{\alpha}{k}$ and $\rho_k = \frac{\beta}{k^r}$ for $k \geq 1$, where the parameters α, β, r are supposed to satisfy:*

$r \geq 2, \alpha > r$ and $\beta > 0$ is such that $\beta < \alpha(\alpha - 2)$ if $r = 2$ (no condition on β if $r > 2$).

Then for any sequence (x_k) generated by (RIPA), we have

$$(i) \ \|x_{k+1} - x_k\| = \mathcal{O}\left(\frac{1}{k^{r-1}}\right) \quad \text{as } k \rightarrow +\infty.$$

Assume additionally that $r = 2$ and that $\frac{|\mu_{k+1} - \mu_k|}{\mu_{k+1}} = \mathcal{O}\left(\frac{1}{k}\right)$ as $k \rightarrow +\infty$. Then the following holds

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

(iii) If $\liminf_{k \rightarrow +\infty} \mu_k > 0$, then there exists $x_\infty \in \text{zer}A$ such that $x_k \rightarrow x_\infty$ weakly in \mathcal{H} as $k \rightarrow +\infty$.

Proof. Let's verify that, with this choice of parameters, the assumptions of Theorem 3.2 are satisfied.

- Assumption (23) is clearly satisfied, with $\bar{\alpha} = 1$ and $\bar{\rho} = 0$ respectively.
- Observe that $\frac{1}{1-\alpha_{k+1}} - \frac{1}{1-\alpha_k} = \frac{1}{\alpha}(k+1) - \frac{1}{\alpha}k = \frac{1}{\alpha}$, hence assumption (24) is verified with $c = \frac{1}{\alpha}$.
- We have $\frac{\rho_{k+1} - \rho_k}{\rho_{k+1}(1-\alpha_k)} = \left(\frac{1}{(k+1)^r} - \frac{1}{k^r}\right)(k+1)^r \frac{k}{\alpha} \rightarrow -\frac{r}{\alpha}$ as $k \rightarrow +\infty$. This shows that assumption (25) is fulfilled with $c'' = -\frac{r}{\alpha}$. The hypothesis $-(1 - \bar{\rho}/2) < c'' \leq -(1 - \bar{\rho}/2)c$ amounts to $-1 < -\frac{r}{\alpha} \leq -\frac{1}{\alpha}$, which is in turn equivalent to $1 \leq r < \alpha$.
- Assumption (26) can be rewritten as $\liminf_{k \rightarrow +\infty} \frac{(1-\alpha_k)^2}{\rho_k} > \frac{1}{1-r/\alpha} = \frac{\alpha}{\alpha-r}$.

We have $(1-\alpha_k)^2/\rho_k = (\alpha^2/k^2)(k^r/\beta) = \frac{\alpha^2}{\beta}k^{r-2}$. Therefore

$$\lim_{k \rightarrow +\infty} \frac{(1-\alpha_k)^2}{\rho_k} = \begin{cases} +\infty & \text{if } r > 2 \\ \alpha^2/\beta & \text{if } r = 2. \end{cases}$$

Hence, assumption (26) is satisfied if $r > 2$, while it is equivalent to $\alpha(\alpha - 2) > \beta$ if $r = 2$.

- Condition $\sum_{k=1}^{+\infty} \frac{\rho_k}{1-\alpha_k} = +\infty$ amounts to $r \leq 2$, which boils down to $r = 2$.
- $\frac{|\mu_{k+1} - \mu_k|}{\mu_{k+1}} = \mathcal{O}\left(\frac{\rho_k}{1-\alpha_k}\right)$ is equivalent to $\frac{|\mu_{k+1} - \mu_k|}{\mu_{k+1}} = \mathcal{O}\left(\frac{1}{k}\right)$.

□

Let's now check the assumptions of Theorem 2.5. Since condition (22) is satisfied with $c = \frac{1}{\alpha}$, Proposition 3.1 shows that the property (K_0) holds true when $\alpha > 1$, and $t_{k+1} \sim \frac{k}{\alpha-1}$. Indeed one can easily verify that equality holds, that is, $t_{k+1} = \frac{k}{\alpha-1}$. We then obtain that condition (K_1) , namely, $t_{k+1}^2 - t_k^2 \leq t_{k+1}$ for every $k \geq 1$, is equivalent to $\alpha \geq 3$. A similar calculation gives that condition (K_1^+) is equivalent to $\alpha > 3$. Thus when $\alpha \geq 3$, and under assumption (H) , we obtain the following convergence rate of the values for (RIPA) $\Phi(p_k) - \min \Phi = \mathcal{O}\left(\frac{1}{k^2}\right)$, where $p_k = \text{prox}_{\mu_k \Phi}(x_k)$.

The previous results are summarized in Figure 1. In the general maximally monotone case, weak convergence of the iterates is ensured by the assumptions of the first column. When $A = \partial\Phi$, the convergence rate of the values $\mathcal{O}\left(\frac{1}{k^2}\right)$ is guaranteed by the hypotheses of the second column. The last column provides assumptions that ensure the weak convergence of the iterates in the case $A = \partial\Phi$. In this case, it is not necessary to assume $\rho_k \rightarrow 0$ to get convergence of iterates.

(RIPA)			
A	maximally monotone	$A = \partial\Phi, \quad \Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ convex lsc.	
α_k	$\alpha_k = 1 - \frac{\alpha}{k}$		
	$\alpha > 2$	$\alpha \geq 3$	$\alpha > 3$
ρ_k	$\rho_k = \frac{\beta}{k^2}, \quad \beta < \alpha(\alpha - 2)$	$0 < \rho_k \leq 1$	
μ_k	$\frac{ \mu_{k+1} - \mu_k }{\mu_{k+1}} = \mathcal{O}\left(\frac{1}{k}\right)$ $\liminf \mu_k > 0$	(μ_k) and $(\rho_k \mu_k)$ nondecreasing	
		$\sup \rho_k \mu_k < +\infty$ $\sup \frac{\mu_k}{k^2} < +\infty$	
	weak convergence of iterates $\ x_{k+1} - x_k\ = \mathcal{O}\left(\frac{1}{k}\right)$	$\Phi(p_k) - \min \Phi = \mathcal{O}\left(\frac{1}{k^2}\right)$ with $p_k = \text{prox}_{\mu_k \Phi}(x_k)$	weak convergence of iterates

Figure 1: Summary of the convergence results for (RIPA) with $\alpha_k = 1 - \frac{\alpha}{k}$.

We can now easily deduce a set of assumptions for which both types of results are satisfied.

Corollary 3.4. *Take the parameters α_k, ρ_k, μ_k of (RIPA) as follows: for each $k \geq 1$*

- $\alpha_k = 1 - \frac{\alpha}{k}, \quad \alpha \geq 3; \quad \rho_k = \frac{\beta}{k^2}, \quad \beta < \alpha(\alpha - 2); \quad \mu_k = ck^{r'}$ for some $r' \geq 2$ and $c > 0$.

Then, we have both

a) the weak convergence of the iterates (x_k) generated by (RIPA) when A is a general maximally monotone operator. Moreover, $\|x_{k+1} - x_k\| = \mathcal{O}\left(\frac{1}{k}\right)$.

b) the rate of convergence of the values $\mathcal{O}\left(\frac{1}{k^2}\right)$ for (RIPA) when $A = \partial\Phi$ and $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex, lower semicontinuous, proper. Precisely, $\Phi(p_k) - \min \Phi = \mathcal{O}\left(\frac{1}{k^2}\right)$ where $p_k = \text{prox}_{\mu_k \Phi}(x_k)$.

Proof. It is an immediate consequence of the results contained in Figure 1. Concerning the choice of μ_k just notice that, when taking $\mu_k = ck^{r'}$ for some $r' \geq 2$, the sequences (μ_k) and $(\rho_k \mu_k) = c\beta k^{r'-2}$ are nondecreasing. Moreover $\frac{|\mu_{k+1} - \mu_k|}{\mu_{k+1}} \sim \frac{r'}{k} = \mathcal{O}\left(\frac{1}{k}\right)$. \square

The case $r = r' = 2$ has been studied by Attouch and Peypouquet [9] with $\alpha_k = 1 - \frac{\alpha}{k}, \quad \rho_k = \frac{s}{\lambda_k + s}$ and $\mu_k = \lambda_k + s$, where $\alpha, s > 0$ and $\lambda_k = (1 + \varepsilon) \frac{s}{\alpha^2} k^2$, for some $\varepsilon > 0$. With this choice,

the sequence (s_k) with $s_k = \rho_k \mu_k$ is constant, equal to s . This particular situation naturally occurs when deriving the algorithm from a dynamical system with constant step size, see section 1.3.2.

3.2. The case $\alpha_k = 1 - \frac{\alpha}{k^q}$. The following result is a direct application of Theorem 3.2 to the case $\alpha_k = 1 - \alpha/k^q$ and $\rho_k = \beta/k^r$, for some $\alpha, \beta > 0$, $q \in]0, 1[$ and $r > 0$.

Corollary 3.5. [6, Corollary 3.7] *Let $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ be a maximally monotone operator such that $\text{zer} A \neq \emptyset$. Suppose that $\alpha_k = 1 - \frac{\alpha}{k^q}$ and $\rho_k = \frac{\beta}{k^r}$ for $k \geq 1$, where the parameters α, β, q and r are supposed to satisfy:*

$(q, r) \in]0, 1[\times \mathbb{R}_+^$, $r \geq 2q$, and $(\alpha, \beta) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$ satisfies $\alpha^2/\beta > 1$ if $r = 2q$ (no condition if $r > 2q$).*

Then for any sequence (x_k) generated by (RIPA), we have

$$(i) \quad \|x_{k+1} - x_k\| = \mathcal{O}\left(\frac{1}{k^{r-q}}\right) \quad \text{as } k \rightarrow +\infty.$$

Assume additionally that $r \leq q + 1$ and that $\frac{|\mu_{k+1} - \mu_k|}{\mu_{k+1}} = \mathcal{O}\left(\frac{1}{k^{r-q}}\right)$ as $k \rightarrow +\infty$. Then the following holds

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

(iii) If $\liminf_{k \rightarrow +\infty} \mu_k > 0$, then there exists $x_\infty \in \text{zer} A$ such that $x_k \rightharpoonup x_\infty$ weakly in \mathcal{H} as $k \rightarrow +\infty$.

Proof. Let's verify that, with this choice of parameters, the assumptions of Theorem 3.2 are satisfied.

- Assumption (23) is clearly satisfied, with $\bar{\alpha} = 1$ and $\bar{\rho} = 0$.
- Observe that $\frac{1}{1-\alpha_{k+1}} - \frac{1}{1-\alpha_k} = \frac{1}{\alpha}((k+1)^q - k^q) \sim \frac{q}{\alpha} k^{q-1} \rightarrow 0$ as $k \rightarrow +\infty$, where we have used $q \in]0, 1[$. Hence assumption (24) is verified with $c = 0$.
- We have $\frac{\rho_{k+1} - \rho_k}{\rho_{k+1}(1-\alpha_k)} = \left(\frac{1}{(k+1)^r} - \frac{1}{k^r}\right) (k+1)^r \frac{k^q}{\alpha} \sim -\frac{r}{\alpha} k^{q-1} \rightarrow 0$ as $k \rightarrow +\infty$.

Therefore, assumption (25) is fulfilled with $c'' = 0$.

- Assumption (26) amounts to $\liminf_{k \rightarrow +\infty} \frac{(1-\alpha_k)^2}{\rho_k} > 1$. We have

$$(1-\alpha_k)^2/\rho_k = (\alpha^2/k^{2q})(k^r/\beta) = \frac{\alpha^2}{\beta} k^{r-2q}, \text{ hence } \lim_{k \rightarrow +\infty} \frac{(1-\alpha_k)^2}{\rho_k} = \begin{cases} +\infty & \text{if } r > 2q \\ \alpha^2/\beta & \text{if } r = 2q. \end{cases}$$

It ensues that assumption (26) is satisfied if $r > 2q$, while it is equivalent to $\alpha^2 > \beta$ if $r = 2q$.

- Condition $\sum_{k=1}^{+\infty} \frac{\rho_k}{1-\alpha_k} = +\infty$ amounts to $\sum_{k=1}^{+\infty} \frac{1}{k^{r-q}} = +\infty$, that is, $r \leq q + 1$.
- $\frac{|\mu_{k+1} - \mu_k|}{\mu_{k+1}} = \mathcal{O}\left(\frac{\rho_k}{1-\alpha_k}\right)$ is equivalent to $\frac{|\mu_{k+1} - \mu_k|}{\mu_{k+1}} = \mathcal{O}\left(\frac{1}{k^{r-q}}\right)$.

□

Let's now check the assumptions of Theorem 2.9. We have $\frac{1}{1-\alpha_{k+1}} - \frac{1}{1-\alpha_k} \rightarrow 0$ as $k \rightarrow +\infty$. Hence, $c = 0$ in Proposition 3.1, which implies that (K_0) is satisfied, and $t_{k+1} \sim \frac{1}{1-\alpha_k} = \frac{1}{\alpha} k^q$. Since $c < \frac{1}{3}$, it follows from [4, Proposition 14] that (K_1^+) is satisfied. Note that $\sum_{i=1}^k t_i \sim \frac{1}{\alpha(1+q)} k^{1+q}$. According to Theorem 2.9, under (H) , the following holds true $\Phi_{\mu_k}(x_k) - \min \Phi = o\left(\frac{1}{k^{1+q}}\right)$ and $\|x_k - x_{k-1}\| = o\left(\frac{s_k}{k^{1+q}}\right)^{\frac{1}{2}}$. Therefore, $\Phi(p_k) - \min \Phi = o\left(\frac{1}{k^{1+q}}\right)$, where $p_k = \text{prox}_{\mu_k \Phi}(x_k)$. Under the additional assumptions $\sup \frac{\mu_k}{k^{1+q}} < +\infty$ and $\sup \rho_k \mu_k < +\infty$, Theorem 2.10 shows the weak convergence of the iterates (x_k) .

The previous results are summarized in the table of Figure 2.

(RIPA)		
A	maximally monotone	$A = \partial\Phi, \quad \Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ convex lsc.
α_k	$\alpha_k = 1 - \frac{\alpha}{k^q}, \quad q \in]0, 1[, \alpha > 0$	
ρ_k	$\rho_k = \frac{\beta}{k^r}, \quad \beta > 0$	$0 < \rho_k \leq 1$
	$r = 2q, \beta < \alpha^2$	$r \in]2q, q + 1]$
μ_k	$\frac{ \mu_{k+1} - \mu_k }{\mu_{k+1}} = \mathcal{O}\left(\frac{1}{k^{r-q}}\right)$ $\liminf \mu_k > 0$	(μ_k) and $(\rho_k \mu_k)$ nondecreasing $\sup \rho_k \mu_k < +\infty$ $\sup \frac{\mu_k}{k^{1+q}} < +\infty$
	weak convergence of iterates $\ x_k - x_{k-1}\ = \mathcal{O}\left(\frac{1}{k^{r-q}}\right)$	$\Phi(p_k) - \min \Phi = o\left(\frac{1}{k^{1+q}}\right)$ $\ x_k - x_{k-1}\ = o\left(\frac{\rho_k \mu_k}{k^{1+q}}\right)^{\frac{1}{2}}$ weak convergence of iterates

Figure 2: Summary of the convergence results for (RIPA) with $\alpha_k = 1 - \frac{\alpha}{k^q}$, $q \in]0, 1[$.

We can now easily deduce a set of assumptions for which both types of results are satisfied.

Corollary 3.6. *Take the parameters α_k, ρ_k, μ_k of (RIPA) as follows: for each $k \geq 1$*

- $\alpha_k = 1 - \frac{\alpha}{k^q}, \quad \alpha > 0, q \in]0, 1[;$
- $\rho_k = \frac{\beta}{k^r}, \quad \beta > 0, r \in [2q, q + 1],$ and $\beta < \alpha^2$ if $r = 2q$ (no condition if $r > 2q$);
- $\mu_k = ck^{r'}$ for some $r' \geq r$ and some positive constant c .

Then, we have both

a) the weak convergence of the iterates (x_k) generated by (RIPA) when A is a general maximally monotone operator. Moreover, $\|x_k - x_{k-1}\| = \mathcal{O}\left(\frac{1}{k^{r-q}}\right)$.

b) the rate of convergence of the values $o\left(\frac{1}{k^{1+q}}\right)$ for the algorithm (RIPA) when $A = \partial\Phi$ and $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex, lower semicontinuous, proper. Precisely, $\Phi(p_k) - \min \Phi = o\left(\frac{1}{k^{1+q}}\right)$

where $p_k = \text{prox}_{\mu_k \Phi}(x_k)$. Moreover, $\|x_k - x_{k-1}\| = o\left(\frac{k^{r'-r}}{k^{1+q}}\right)^{\frac{1}{2}}$.

Proof. It is an immediate consequence of the results contained in Figure 2. Concerning the choice of μ_k just notice that, when taking $\mu_k = c k^{r'}$ for some $r' \geq r$, the sequences (μ_k) and $(\rho_k \mu_k) = c\beta k^{r'-r}$ are nondecreasing. Moreover $\frac{|\mu_{k+1} - \mu_k|}{\mu_{k+1}} \sim \frac{r'}{k} = \mathcal{O}\left(\frac{1}{k^{r-q}}\right)$, since $r \leq q+1$. \square

Remark 3.7. As a general rule, we can observe a balance between the inertial effect and the relaxation effect. As α_k gets closer to one, the relaxation parameter ρ_k gets closer to zero. As well, there is a balance between the relaxation and the regularization parameters. As ρ_k tends to zero, the regularization parameter μ_k tends to infinity. A typical situation is $\mu_k = \frac{C}{\rho_k}$, for some $C > 0$.

APPENDIX A. MOREAU ENVELOPES

Recall some basic facts concerning the Moreau envelope. Let $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex lower semicontinuous proper function. For any $\mu > 0$, and any $x \in \mathcal{H}$,

$$(27) \quad \Phi_\mu(x) = \inf_{\xi \in \mathcal{H}} \left\{ \Phi(\xi) + \frac{1}{2\mu} \|x - \xi\|^2 \right\}.$$

The infimum in (27) is achieved at a unique point $\text{prox}_{\mu\Phi}(x)$, which gives

$$\Phi_\mu(x) = \Phi(\text{prox}_{\mu\Phi}(x)) + \frac{1}{2\mu} \|x - \text{prox}_{\mu\Phi}(x)\|^2.$$

Writing the optimality condition for (27), we get

$$\text{prox}_{\mu\Phi}(x) + \mu \partial\Phi(\text{prox}_{\mu\Phi}(x)) \ni x,$$

that is $\text{prox}_{\mu\Phi}(x) = (I + \mu\partial\Phi)^{-1}(x)$. Thus, $\text{prox}_{\mu\Phi}$ is the resolvent of index $\mu > 0$ of the maximal monotone operator $\partial\Phi$. Consequently, the mapping $\text{prox}_{\mu\Phi} : \mathcal{H} \rightarrow \mathcal{H}$ is firmly nonexpansive. For any $\mu > 0$, the function $x \mapsto \Phi_\mu(x)$ is continuously differentiable, with

$$\nabla\Phi_\mu(x) = \frac{1}{\mu} (x - \text{prox}_{\mu\Phi}(x)).$$

Thus $\nabla\Phi_\mu = \frac{1}{\mu} (I - (I + \mu\partial\Phi)^{-1}) = (\partial\Phi)_\mu$ is the Yosida approximation with index μ of the maximally monotone operator $\partial\Phi$. As such, $\nabla\Phi_\mu$ is Lipschitz continuous, with Lipschitz constant $\frac{1}{\mu}$, and $\Phi_\mu \in \mathcal{C}^{1,1}(\mathcal{H})$. See [12, 16, 28, 29], for further results.

APPENDIX B. AUXILIARY RESULTS

Let us present some auxiliary lemmas that are used throughout the paper. To establish the weak convergence of the iterates of (RIPA), we apply Opial's Lemma [27], that we recall in its discrete form.

Lemma B.1. *Let S be a nonempty subset of \mathcal{H} , and (x_k) a sequence of elements of \mathcal{H} . Assume that*

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

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

The next lemma provides an estimate of the convergence rate of a sequence that is summable with respect to weights.

Lemma B.2 ([4, Lemma 22]). *Let (τ_k) be a nonnegative sequence such that $\sum_{k=1}^{+\infty} \tau_k = +\infty$. Assume that (ε_k) is a nonnegative and nonincreasing sequence satisfying $\sum_{k=1}^{+\infty} \tau_k \varepsilon_k < +\infty$. Then we have*

$$\varepsilon_k = o\left(\frac{1}{\sum_{i=1}^k \tau_i}\right) \quad \text{as } k \rightarrow +\infty.$$

The following result allows us to establish the summability of a sequence (a_k) satisfying some suitable inequality.

Lemma B.3 ([4, Lemma 23]). *Given a nonnegative sequence (α_k) satisfying (K_0) , let (t_k) be the sequence defined by $t_k = 1 + \sum_{i=k}^{+\infty} \prod_{j=k}^i \alpha_j$. Let (a_k) and (ω_k) be two sequences of nonnegative numbers such that $a_{k+1} \leq \alpha_k a_k + \omega_k$, for all $k \geq 0$. If $\sum_{k=0}^{+\infty} t_{k+1} \omega_k < +\infty$, then $\sum_{k=0}^{+\infty} a_k < +\infty$.*

REFERENCES

- [1] F. ÁLVAREZ, *On the minimizing property of a second-order dissipative system in Hilbert spaces*, SIAM J. Control Optim., 38, No. 4, (2000), pp. 1102-1119.
- [2] F. ÁLVAREZ, *Weak Convergence of a Relaxed and Inertial Hybrid Projection-Proximal Point Algorithm for Maximal Monotone Operators in Hilbert Space*, SIAM J. Optim., 14 (2004), pp. 773-782.
- [3] F. ÁLVAREZ, H. ATTOUCH, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Set-Valued Analysis, 9 (2001), No. 1-2, pp. 3-11.
- [4] H. ATTOUCH, A. CABOT, *Convergence rates of inertial forward-backward algorithms*, SIAM J. Optim., 28 (1) (2018), pp. 849-874.
- [5] H. ATTOUCH, A. CABOT, *Convergence of damped inertial dynamics governed by regularized maximally monotone operators*, J. Differential Equations, 264 (2018), pp. 7138-7182.
- [6] H. ATTOUCH, A. CABOT, *Convergence of a relaxed inertial proximal algorithm for maximally monotone operators*, (2018), HAL-01708905.
- [7] H. ATTOUCH, Z. CHBANI, J. PEYPOUQUET, P. REDONT, *Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity*, Math. Program. Ser. B 168 (2018), pp. 123-175.
- [8] H. ATTOUCH, J. PEYPOUQUET, *The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $\frac{1}{k^2}$* , SIAM J. Optim., 26 (2016), No. 3, pp. 1824-1834.
- [9] H. ATTOUCH, J. PEYPOUQUET, *Convergence of inertial dynamics and proximal algorithms governed by maximal monotone operators*, Math. Program., Ser. B, <https://doi.org/10.1007/s10107-018-1252-x>, published online March 2018.
- [10] J.-F. AUJOL, CH. DOSSAL, *Stability of over-relaxations for the Forward-Backward algorithm, application to FISTA*, SIAM J. Optimization, 25 (2015), pp. 2408-2433.
- [11] J.-F. AUJOL, CH. DOSSAL, A. RONDEPIERRE, *Optimal convergence rates for Nesterov acceleration*, arXiv:1805.05719v1 [math.OC], 2018.
- [12] H. BAUSCHKE, P. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert spaces*, CMS Books in Mathematics, Springer, (2011).
- [13] A. BECK, M. TEOULLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), No. 1, pp. 183-202.
- [14] P. BÉGOUT, J. BOLTE, M.A. JENDOUBI, *On damped second-order gradient systems*, Journal of Differential Equations, 259 (7-8) (2015), pp. 3115-3143.
- [15] R. I. BOT, E. R. CSETNEK, *Second order forward-backward dynamical systems for monotone inclusion problems*, SIAM Journal on Control and Optimization 54(3) (2016), pp. 1423-1443.
- [16] H. BRÉZIS, *Opérateurs maximaux monotones dans les espaces de Hilbert et équations d'évolution*, Lecture Notes 5, North Holland, (1972).
- [17] A. CHAMBOLLE, CH. DOSSAL, *On the convergence of the iterates of the Fast Iterative Shrinkage/Thresholding Algorithm*, Journal of Optimization Theory and Applications, 166 (2015), pp. 968-982.
- [18] Y. DRORI, M. TEOULLE, *Performance of first-order methods for smooth convex minimization: a novel approach*, Mathematical Programming, Series A, 145 (2014), pp. 451-482.
- [19] J. ECKSTEIN, D. P. BERTSEKAS, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55 (1992), pp. 293-318.
- [20] F. IUTZELER, J.M. HENDRICKX, *Generic online acceleration scheme for optimization algorithms via relaxation and inertia*, (2017), arXiv:1603.05398v3 [math.OC].
- [21] F. IUTZELER, J. MALICK, *On the Proximal Gradient Algorithm with Alternated Inertia*, Journal of Optimization Theory and Applications, 176 (2018), pp. 688-710.
- [22] D. KIM, J.A. FESSLER, *Optimized first-order methods for smooth convex minimization*, Math. Program., 159 (2016), No. 1, pp. 81-107.
- [23] J. LIANG, J. FADILI, G. PEYRÉ, *Local linear convergence of forward-backward under partial smoothness*, Advances in Neural Information Processing Systems, 2014, pp. 1970-1978.
- [24] P.-E. MAINGÉ, *Convergence theorems for inertial KM-type algorithms*, Journal of Computational and Applied Mathematics, 219 (2008), pp. 223-236.

- [25] Y. NESTEROV, *A method of solving a convex programming problem with convergence rate $O(1/k^2)$* , Soviet Mathematics Doklady, 27 (1983), pp. 372–376.
- [26] Y. NESTEROV, *Introductory lectures on convex optimization: A basic course*, volume 87 of Applied Optimization. Kluwer Academic Publishers, Boston, MA, 2004.
- [27] Z. OPIAL, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. Amer. Math. Soc., 73 (1967), pp. 591–597.
- [28] N. PARIKH, S. BOYD, *Proximal algorithms*, Foundations and trends in optimization, volume 1, (2013), pp. 123–231.
- [29] J. PEYPOUQUET, *Convex optimization in normed spaces: theory, methods and examples*. Springer, 2015.
- [30] B.T. POLYAK, *Some methods of speeding up the convergence of iteration methods*, U.S.S.R. Comput. Math. Math. Phys., 4 (1964), pp. 1–17.
- [31] W. SU, S. BOYD, E. J. CANDÈS, *A Differential Equation for Modeling Nesterov’s Accelerated Gradient Method: Theory and Insights*, Journal of Machine Learning Research, 17 (2016), pp. 1–43.
- [32] S. VILLA, S. SALZO, L. BALDASSARRES, A. VERRI, *Accelerated and inexact forward-backward*, SIAM J. Optim., 23 (2013), No. 3, pp. 1607–1633.

INSTITUT MONTPELLIÉRAIN ALEXANDER GROTHENDIECK, UMR 5149 CNRS, UNIVERSITÉ MONTPELLIER, PLACE EUGÈNE BATAILLON, 34095 MONTPELLIER CEDEX 5, FRANCE

E-mail address: `hedy.attouch@univ-montp2.fr`

INSTITUT DE MATHÉMATIQUES DE BOURGOGNE, UMR 5584 CNRS, UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ, 21000 DIJON, FRANCE.

E-mail address: `alexandre.cabot@u-bourgogne.fr`