CONVERGENCE RATES OF INERTIAL FORWARD-BACKWARD ALGORITHMS

Hedy Attouch, Alexandre Cabot

CONVERGENCE RATES OF INERTIAL FORWARD-BACKWARD ALGORITHMS*

HEDY ATTOUCH† AND ALEXANDRE CABOT ‡

Abstract. In a Hilbert space $\mathcal{H}$, given (α_k) a general sequence of nonnegative numbers, we analyze the convergence properties of the inertial forward-backward algorithm

$$(IFB) \quad \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = \operatorname{prox}_{s\Psi}(y_k - s\nabla\Phi(y_k)), \end{cases}$$

where $\Psi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a proper lower-semicontinuous convex function, and $\Phi : \mathcal{H} \to \mathbb{R}$ is a differentiable convex function, whose gradient is Lipschitz continuous. Various options for the sequence (α_k) are considered in the literature. Among them, the Nesterov choice leads to the FISTA algorithm [10], and accelerates convergence from $\mathcal{O}(1/k)$ to $\mathcal{O}(1/k^2)$ for the values. Several variants are used to guarantee the convergence of the iterates, or also to improve the rate of convergence for the values, see [4, 5, 7, 19]. For the design of fast optimization methods, the tuning of the sequence (α_k) is a subtle issue, which we deal with in this paper in general. We show that the convergence rate of the algorithm can be obtained simply by analyzing the sequence of positive real numbers (α_k). In addition to the case $\alpha_k = 1 - \frac{\alpha}{k}$ with $\alpha \geq 3$, our results apply equally well to $\alpha_k = 1 - \frac{\alpha}{k^r}$, with an exponent $0 < r < 1$, and to Polyak's heavy ball method. Thus, we unify most of the existing results based on the accelerated gradient method of Nesterov. In the process, we improve some of them, and discover new ones.

Key words. Accelerated gradient method; FISTA; inertial forward-backward algorithms; Nesterov method; proximal-based methods; structured convex optimization; vanishing damping.

AMS subject classification. 49M37, 65K05, 90C25

1. Introduction. In this paper, $\mathcal{H}$ is a real Hilbert space endowed with the scalar product $\langle \cdot, \cdot \rangle$ and the corresponding norm $\| \cdot \|$.

1.1. Problem statement. Let us consider the structured convex minimization problem

$$\min \{\Psi(x) + \Phi(x) : \ x \in \mathcal{H}\} \tag{1}$$

where $\Psi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a proper lower-semicontinuous convex function, and $\Phi : \mathcal{H} \to \mathbb{R}$ is a continuously differentiable convex function, whose gradient is Lipschitz continuous. Given initial data x_0, $x_1 \in \mathcal{H}$ and a general sequence (α_k) of nonnegative numbers, we will analyze the convergence properties of the inertial forward-backward algorithm

$$(IFB) \quad \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = \operatorname{prox}_{s\Psi}(y_k - s\nabla\Phi(y_k)). \end{cases}$$

*Effort sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number F49550-1 5-1-0500. Also supported by ECOS-Conicyt Project C13E03.

†Institut Montpelliérain Alexander Grothendieck, UMR 5149 CNRS, Université Montpellier, place Eugène Bataillon, 34095 Montpellier cedex 5, France (hedy.attouch@univ-montp2.fr).

‡Institut de Mathématiques de Bourgogne, UMR 5584 CNRS, Université Bourgogne Franche-Comté, 21000 Dijon, France. (alexandre.cabot@u-bourgogne.fr).

The main structure of algorithm (IFB) is that of FISTA [10], with potential for studying variants through the choice of (α_k). We will systematically assume the following set of hypotheses

$$(H)\left\{\begin{array}{l}\bullet\ \Psi:\mathcal{H}\to\mathbb{R}\cup\{+\infty\}\text{ is convex, proper and lower semicontinuous;}\\ \bullet\ \Phi:\mathcal{H}\to\mathbb{R}\text{ is convex, differentiable with } L\text{-Lipschitz continuous gradient;}\\ \bullet\ \Theta:=\Psi+\Phi\text{ has a nonempty set of minimizers: } S=\operatorname{argmin}\Theta\neq\emptyset;\\ \bullet\ \text{The sequence } (\alpha_k) \text{ is nonnegative;}\\ \bullet\ \text{The parameter } s \text{ satisfies } s\in\,]0,1/L].\end{array}\right.$$

We will obtain rates of convergence for the sequences generated by algorithm (IFB) under conditions involving only the sequence (α_k). In particular, we will obtain fast convergence results that unify most of the existing results based on the Nesterov accelerated gradient method. Let us first recall some classical facts concerning the inertial forward-backward algorithms. Algorithm (IFB) combines the two basic blocks: a gradient step with respect to the smooth function Φ, and a proximal step with respect to the nonsmooth function Ψ, hence the equivalent terminology, inertial proximal-gradient algorithm. We recall the definition of the proximal mapping $\operatorname{prox}_{s\Psi}:\mathcal{H}\to\mathcal{H}$, which is defined, for every $x\in\mathcal{H}$, by the formula

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

One can consult [9, 32, 34], for a recent account on the proximal-based algorithms, that play a central role in nonsmooth optimization. The classical forward-backward algorithm corresponds to $\alpha_k=0$, and writes as follows $x_{k+1}=\operatorname{prox}_{s\Psi}(x_k-s\nabla\Phi(x_k))$. By contrast, for $\alpha_k>0$, algorithm (IFB) involves an additional extrapolation operation, given by $y_k=x_k+\alpha_k(x_k-x_{k-1})$. For α_k judiciously chosen, this inertial term improves the speed of convergence of the algorithm.

1.2. Historical perspective. In a seminal paper [28], Nesterov proposed an accelerated gradient method that corresponds to the particular choice

$$\alpha_k=\frac{t_k-1}{t_{k+1}}\quad\text{with } t_1=1 \text{ and } t_{k+1}=\frac{\sqrt{4t_k^2+1}+1}{2}.$$

This choice leads to an increasing sequence (α_k), that behaves like $1-\frac{3}{k}$ as $k\to+\infty$. It has been further extended to the structured minimization problem (1) by Beck and Teboulle in [10], that's FISTA. This scheme exhibits the convergence rate $\mathcal{O}(\frac{1}{k^2})$, which, for gradient methods, is known to be optimal among all methods having only information about the gradient at consecutive iterates [29]. FISTA has been successfully applied to large size problems, especially in the domain of the signal/imaging processing. Accelerating first-order methods is a subject of active research. Since the introduction of Nesterov's scheme, much has been done on the development of first-order accelerated methods, to name but a few [9, 10, 19, 20, 24, 25, 26, 32, 34, 35, 36]. Recently, a special attention has been devoted to the case $\alpha_k=\frac{k-1}{k+\alpha-1}$, where $\alpha>0$, see [4, 5, 19, 35]. Given $x_0,\ x_1\in\mathcal{H}$, for $k\geq 1$ the algorithm writes

$$\left\{\begin{array}{rcl} y_k &=& x_k+\frac{k-1}{k+\alpha-1}(x_k-x_{k-1})\\ x_{k+1} &=& \operatorname{prox}_{s\Psi}\left(y_k-s\nabla\Phi(y_k)\right).\end{array}\right.\tag{2}$$

The asymptotic properties of the iterates of (2) remain unchanged if the coefficient $\alpha_k=\frac{k-1}{k+\alpha-1}$ is replaced[1] with the equivalent expression $\alpha_k=1-\frac{\alpha}{k}$. For $\alpha=3$ we recover a first-order approximation

[1] This will be established from Proposition 15 and Theorem 16.

of the original FISTA algorithm. The great novelty with the algorithm (2) is that, while keeping the same computational complexity as in the case $\alpha = 3$, taking $\alpha > 3$ offers many advantages. First, it ensures the convergence of the sequences (x_k), as proved by Chambolle and Dossal [19], see also [4]. Let us recall that the convergence of the sequences generated by FISTA has not been established so far. This is a puzzling question in the study of numerical optimization methods. Second, as proved by Attouch and Peypouquet in [5], it provides the better rate of convergence

$$(\Psi + \Phi)(x_k) - \min(\Psi + \Phi) = o\left(\frac{1}{k^2}\right).$$

Thus, when taking $\alpha_k = 1 - \frac{\alpha}{k}$, $\alpha = 3$ appears as a critical value. Precisely, we shall see that the fast convergence properties of algorithm (IFB) are based on a subtle tuning of α_k. We will highlight the critical values that are attached to these questions, and show that the rate of convergence of the algorithm can be obtained simply by analyzing the sequence of positive real numbers (α_k). In the case $\alpha_k = 1 - \frac{\alpha}{k}$, we will recover $\alpha = 3$ as a critical value. So doing, we will unify and somehow improve some of the recent papers concerning these questions, [4, 5, 19, 35].

Let us end this paragraph by evoking some of the recent trends. For optimization, there is a strong numerical evidence that inertia is a good thing at the beginning of trajectories, but that ultimately it can induce undesirable oscillations with negative effect. Recent studies by Liang, Fadili, and Peyré [25] have analyzed this fact. To overcome these difficulties, various strategies are developed, such as restarting (O'Donoghue and Candès [30]), geometrical damping (Attouch, Peypouquet and Redont [6]), periodic damping (Ghisi, Gobbino and Haraux [21]), introduction of an additional damping term (Kim and Fessler [24]). In this perspective, our study with a general sequence (α_k) provides a valuable tool to analyze these various situations. It paves the way to further studies concerning the analysis of algorithm (IFB) from the optimal control point of view of dynamical systems, where (α_k) is the control variable, and (x_k) is the corresponding state variable. Passing from the open-loop control (which is the case in the paper) to a closed-loop control is also a promising direction of research. It should be noted that several research studies aim to extend the above results beyond the convex minimization framework, considering for example the equations governed by maximal monotone operators (Bot and Csetnek [12]), or considering non-convex minimization problems based on the Kurdyka-Lojasiewicz property (Bégout, Bolte and Jendoubi [11]).

1.3. Links with dynamical systems. Algorithm (IFB) is strongly connected with the asymptotic behavior of damped inertial systems. Indeed, algorithm (IFB) comes naturally into play by performing an implicit discretization with respect to Ψ, and an explicit discretization with respect to Φ of the inertial system

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \partial\Psi(x(t)) + \nabla\Phi(x(t)) \ni 0, \tag{3}$$

whence the terminology "Inertial Forward-Backward" algorithm. Taking the damping parameter $\gamma(t)$ that tends to zero as $t \to +\infty$ is a key property for obtaining fast optimization methods. A decisive step in this direction was obtained by Su, Boyd, and Candès in [35], who showed that the FISTA method can be obtained as a discretization of the dynamic system (3) with $\gamma(t) = \frac{3}{t}$. For other recent papers on inertial gradient systems with time-dependent friction, the reader is referred to [3, 16, 17, 23, 27]. On the basis of the correspondance $\alpha_k = 1 - h\gamma_k$ (where h is the time step, and $\gamma_k = \gamma(kh)$), the discrete version of the vanishing damping property is that the sequence (α_k) tends to one from below as $k \to +\infty$. In [1, 2] the first results concerning inertial proximal-based methods were limited to the case $\alpha_k \le \overline{\alpha} < 1$. Then, the condition $\alpha_k \to 1$ is not satisfied, and the convergence rate for the values is no better than for the classical forward-backward methods, namely $\mathcal{O}(\frac{1}{k})$.

1.4. Presentation of the results. Let us briefly introduce the basic ingredients used in the description of the results. The case $\Phi = \Psi = 0$ already reveals the crucial notions. In this case, algorithm (IFB) becomes $x_{k+1} - x_k - \alpha_k(x_k - x_{k-1}) = 0$. It ensues that for every $k \geq 1$,

$$x_k = x_1 + \left(\sum_{i=1}^{k-1} \prod_{j=1}^{i} \alpha_j \right) (x_1 - x_0).$$

Hence, (x_k) converges iff the following condition is satisfied: $\sum_{i=1}^{+\infty} \prod_{j=1}^{i} \alpha_j < +\infty$. This condition has already been identified as a necessary condition of convergence of algorithm (IFB) when the convex potential has a continuum of minima, see [18]. We are naturally led to introduce the sequence (t_k)

$$t_k := 1 + \sum_{i=k}^{+\infty} \prod_{j=k}^{i} \alpha_j,$$

that plays a crucial role. As a model example of our results, let us state the following

THEOREM 1. *Let us make assumptions* (H).
A. Suppose that the sequence (α_k) *satisfies* (K_0) *and* (K_1).

(K_0) $\quad \forall k \geq 1, \quad \sum_{i=k}^{+\infty} \prod_{j=k}^{i} \alpha_j < +\infty,$

(K_1) $\quad \forall k \geq 1, \quad t_{k+1}^2 - t_k^2 \leq t_{k+1}.$

Then, for any sequence (x_k) *generated by algorithm* (IFB)

$$(\Psi + \Phi)(x_k) - \min(\Psi + \Phi) = \mathcal{O}\left(\frac{1}{t_k^2}\right) \quad \text{as } k \to +\infty.$$

B. Assume moreover that there exists $m < 1$ *such that*

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

Then, as $k \to +\infty$,

$$(\Psi + \Phi)(x_k) - \min(\Psi + \Phi) = o\left(\frac{1}{\sum_{i=1}^{k} t_i}\right) \quad \text{and} \quad \|x_k - x_{k-1}\| = o\left(\frac{1}{\sum_{i=1}^{k} t_i}\right)^{\frac{1}{2}}.$$

As a consequence, we have

$$(\Psi + \Phi)(x_k) - \min(\Psi + \Phi) = o\left(\frac{1}{t_k^2}\right) \quad \text{and} \quad \|x_k - x_{k-1}\| = o\left(\frac{1}{t_k}\right) \quad \text{as } k \to +\infty.$$

Finally, if $\alpha_k \in [0,1]$ *for every* $k \geq 1$, *then the sequence* (x_k) *converges weakly toward some* $\overline{x} \in \operatorname{argmin}(\Psi + \Phi)$.

When $\alpha_k = 1 - \frac{\alpha}{k}$ for every $k \geq 1$, one can easily verify that $t_{k+1} = \frac{k}{\alpha - 1}$. In such a case, (K_1) corresponds to $\alpha \geq 3$, (K_1^+) to $\alpha > 3$, which makes it possible to find the results of [4, 5, 19]. Note that (K_1) is satisfied by taking t_{k+1} equal to the positive root of the second-order equation $t_{k+1}^2 - t_{k+1} - t_k^2 = 0$, which is the historical choice by Nesterov. Theorem 1 permits to consider as well the case $\alpha_k = 1 - \frac{\alpha}{k^r}$ for some $0 < r < 1$, thus allowing to recover some results of Aujol-Dossal [7]. In this last article, the authors consider the effect of perturbations on the convergence rate of FISTA and certain variants. As an interesting result, they show that for "large" perturbations, slowing down the over relaxation in the forward-backward iterations (which is the case for $\alpha_k = 1 - \frac{\alpha}{k^r}$, $0 < r < 1$) can ensure the weak convergence of the iterates. This is another justification for studying (IFB) in our general framework.

1.5. Organization of the paper. After introducing some energy estimates and Lyapunov functions in section 2, our main convergence results are established in section 3. Depending on the behavior of the sequence (α_k), we give the corresponding rates of convergence for the sequences (x_k) generated by algorithm (IFB). In section 4, we also examine a few situations where the sequences (x_k) converge strongly. We pay particular attention to the case of a strong minimum. These results are illustrated in section 5 that presents applications to special classes of sequences (α_k). The paper is completed by some technical auxiliary lemmas contained in the Appendix.

2. Algorithm (IFB). Hypotheses and notations. Let us introduce the forward-backward operator $T_s : \mathcal{H} \to \mathcal{H}$, and the operator $G_s : \mathcal{H} \to \mathcal{H}$, respectively given by

$$T_s(y) = \text{prox}_{s\Psi}(y - s\nabla\Phi(y)), \qquad G_s(y) = \frac{1}{s}\left(y - T_s(y)\right).$$

The equality $x_{k+1} = \text{prox}_{s\Psi}(y_k - s\nabla\Phi(y_k))$ can be rewritten as $x_{k+1} = T_s(y_k)$, which is equivalent to $x_{k+1} = y_k - sG_s(y_k)$. The following lemma gives several important classical properties of the operators T_s and G_s. The inequality (4) is known as the descent rule for the forward-backward methods, see [10, Lemma 2.3], [19, Lemma 1]. For the equivalences in (5) see for example [32, section 4.2].

Lemma 2. *Assume hypothesis (H).*

(i) For all x, $y \in \mathcal{H}$, we have

$$\Theta(y - sG_s(y)) \le \Theta(x) + \langle G_s(y), y - x\rangle - \frac{s}{2}\|G_s(y)\|^2. \tag{4}$$

(ii) The operator T_s is nonexpansive, the operator G_s is monotone, and the following equivalences hold true

$$z \in \text{argmin}\,\Theta \iff z = T_s(z) \iff G_s(z) = 0. \tag{5}$$

The energy (potential + kinetic) associated to the iterates generated by algorithm (IFB) is the sequence (W_k)

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

Let us evaluate the energy decay. Recall that the friction effect, and thus the dissipation of energy, is related to $\alpha_k \le 1$.

Proposition 3. *Under hypothesis (H), let (x_k) be a sequence generated by algorithm (IFB). The energy sequence (W_k) satisfies for every $k \ge 1$,*

$$W_{k+1} - W_k \le -\frac{1-\alpha_k^2}{2s}\|x_k - x_{k-1}\|^2. \tag{6}$$

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

Proof. By applying formula (4) with $y = y_k$ and $x = x_k$, we obtain

$$\begin{aligned}\Theta(x_{k+1}) = \Theta(y_k - sG_s(y_k)) &\le \Theta(x_k) + \langle G_s(y_k), y_k - x_k\rangle - \frac{s}{2}\|G_s(y_k)\|^2\\ &= \Theta(x_k) - \frac{1}{2s}\|y_k - sG_s(y_k) - x_k\|^2 + \frac{1}{2s}\|y_k - x_k\|^2.\end{aligned}$$

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

$$\Theta(x_{k+1}) \le \Theta(x_k) - \frac{1}{2s}\|x_{k+1} - x_k\|^2 + \frac{\alpha_k^2}{2s}\|x_k - x_{k-1}\|^2.$$

This can be equivalently rewritten as

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

The last assertion is immediate. □

Let us now fix $x^* \in \mathcal{H}$, and define the sequence (h_k) by $h_k = \frac{1}{2}\|x_k - x^*\|^2$. The next result will be useful for establishing the convergence of the iterates of (IFB).

PROPOSITION 4. *Under* (H), *we have*

$$h_{k+1} - h_k - \alpha_k(h_k - h_{k-1}) = \frac{1}{2}(\alpha_k^2 + \alpha_k)\|x_k - x_{k-1}\|^2 - s\langle G_s(y_k), y_k - x^*\rangle + \frac{s^2}{2}\|G_s(y_k)\|^2. \tag{7}$$

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

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

Proof. Observe that

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

Setting briefly $A_k = h_{k+1} - h_k - \alpha_k(h_k - h_{k-1})$, 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 $x_{k+1} = y_k - sG_s(y_k)$, we obtain (7).
Let us now assume that $x^* \in \operatorname{argmin}\Theta$. Recall from inequality (4) applied with $y = y_k$ and $x = x^*$ that

$$\Theta(x_{k+1}) = \Theta(y_k - sG_s(y_k)) \leq \Theta(x^*) + \langle G_s(y_k), y_k - x^*\rangle - \frac{s}{2}\|G_s(y_k)\|^2.$$

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

$$-s\langle G_s(y_k), y_k - x^*\rangle + \frac{s^2}{2}\|G_s(y_k)\|^2 \leq -s(\Theta(x_{k+1}) - \min\Theta),$$

which completes the proof of Proposition 4. □

3. Asymptotic behavior of the iterates of (IFB)**.** Let us first introduce the sequence (t_k) that will play a central role in the analysis of algorithm (IFB).

3.1. The sequence (t_k). Given $k \geq 1$, assume that $\sum_{i=k}^{+\infty} \prod_{j=k}^{i} \alpha_j < +\infty$ and set

$$t_k = 1 + \sum_{i=k}^{+\infty} \prod_{j=k}^{i} \alpha_j.$$

Throughout the paper, we will use the convention $\prod_{j=k}^{k-1} \alpha_j = 1$. This allows to write the term t_k under the more compact form

$$t_k = \sum_{i=k-1}^{+\infty} \prod_{j=k}^{i} \alpha_j. \tag{8}$$

If the series defining t_{k+1} is convergent, the series defining t_k is also convergent, and we have the following relation

$$1 + \alpha_k t_{k+1} = 1 + \alpha_k \left(\sum_{i=k}^{+\infty} \prod_{j=k+1}^{i} \alpha_j \right) = 1 + \sum_{i=k}^{+\infty} \prod_{j=k}^{i} \alpha_j = t_k.$$

Conversely, if $\alpha_k \neq 0$ and if t_k is well-defined, then the series defining t_{k+1} is convergent and the above equalities hold true. From now on, we assume that

$$\sum_{i=k}^{+\infty} \prod_{j=k}^{i} \alpha_j < +\infty \qquad \text{for every } k \geq 1. \tag{K_0}$$

Let us summarize the above results.

Lemma 5. *Assume that the nonnegative sequence (α_k) satisfies (K_0). Then the sequence (t_k) is well defined and satisfies for every $k \geq 1$*

$$1 + \alpha_k t_{k+1} = t_k. \tag{9}$$

3.2. Convergence rates for the values and weak convergence of the iterates. Given $x^* \in \operatorname{argmin} \Theta$, let us define the sequence $(\mathcal{E}_k)$ by

$$\mathcal{E}_k = t_k^2(\Theta(x_k) - \min \Theta) + \frac{1}{2s}\|x_{k-1} + t_k(x_k - x_{k-1}) - x^*\|^2. \tag{10}$$

The discrete energy $\mathcal{E}_k$ was introduced in [10]. The next result shows that the sequence $(\mathcal{E}_k)$ is nonincreasing, under some suitable condition named (K_1). The statement is essentially as given in [10, Lemma 4.1]. The only difference is the use of inequality (K_1), in place of the equality considered in [10].

Proposition 6. *Under (H), assume that the nonnegative sequence (α_k) satisfies (K_0). Let (x_k) be a sequence generated by algorithm (IFB), and let $(\mathcal{E}_k)$ be the sequence defined by (10). Then we have*

$$\mathcal{E}_{k+1} - \mathcal{E}_k \leq (t_{k+1}^2 - t_k^2 - t_{k+1})(\Theta(x_k) - \min \Theta). \tag{11}$$

Under the assumption

$$t_{k+1}^2 - t_k^2 \leq t_{k+1} \qquad \textit{for every } k \geq 1, \tag{K_1}$$

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

THEOREM 7. *Under (H), assume that the nonnegative sequence (α_k) satisfies (K_0)-(K_1). Let (x_k) be a sequence generated by algorithm (IFB). Then we have*
(i) *For every $k \geq 1$, $\Theta(x_k) - \min \Theta \leq \dfrac{C}{t_k^2}$ with*

$$C = t_1^2(\Theta(x_1) - \min \Theta) + \frac{1}{s}(d(x_0, S)^2 + t_1^2 \|x_1 - x_0\|^2).$$

(ii) *Assume moreover that there exists $m < 1$ such that*

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

Then we have

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

Proof. (i) From Proposition 6, 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

$$\begin{aligned} (12) \qquad t_k^2\,(\Theta(x_k) - \min \Theta) \leq \mathcal{E}_1 &= t_1^2[\Theta(x_1) - \min \Theta] + \frac{1}{2s}\|x_0 - x^* + t_1(x_1 - x_0)\|^2 \\ &\leq t_1^2[\Theta(x_1) - \min \Theta] + \frac{1}{s}(\|x_0 - x^*\|^2 + t_1^2\|x_1 - x_0\|^2). \end{aligned}$$

Since x^* can be taken arbitrarily in S, we finally obtain $t_k^2\,(\Theta(x_k) - \min \Theta) \leq C$, with

$$C = t_1^2(\Theta(x_1) - \min \Theta) + \frac{1}{s}(d(x_0, S)^2 + t_1^2 \|x_1 - x_0\|^2).$$

(ii) By summing inequality (11) from $k = 1$ to n, we find

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

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

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

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

REMARK 1. *From (12) we have*

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

where

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

As a function of (x_0, x_1), $\mathcal{E}_1$ achieves its minimum when $x_1 \in \operatorname{argmin} \Theta$ and $x_1 - x_0 = \frac{1}{t_1}(x^* - x_0)$. Of course, taking $x_1 \in \operatorname{argmin} \Theta$ 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 forward-backward direction $\operatorname{prox}_{s\Psi}(x_0 - s\nabla\Phi(x_0)) - x_0$ (which is equal to $-sG_s(x_0)$).

PROPOSITION 8. *Under (H), assume that the sequence (α_k) satisfies (K_0)-(K_1^+). Let (x_k) be a sequence generated by algorithm (IFB). Then we have*

$$\sum_{k=1}^{+\infty} t_k \|x_k - x_{k-1}\|^2 < +\infty. \tag{13}$$

Proof. Let us recall the inequality (6) from Proposition 3

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

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

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

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

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

Recalling the expression of W_k, we deduce that

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

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

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

Observing that $t_k \geq 1$, we have

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

and hence

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

Under (K_1^+), we have the estimate $\sum_{k=1}^{+\infty} t_{k+1}(\Theta(x_k) - \min\Theta) < +\infty$, see Theorem 7. The conclusion is obtained by letting n tend to infinity in the above inequality. ∎

REMARK 2. *Observe that assumption (K_1) implies that*

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

Since $t_k \geq 1$, we deduce that $t_{k+1} \leq 2t_k$ for every $k \geq 1$. The estimate (13) then yields

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

Recalling from Theorem 7 that $\sum_{k=1}^{+\infty} t_{k+1}(\Theta(x_k) - \min\Theta) < +\infty$, we infer that $\sum_{k=1}^{+\infty} t_{k+1}W_k < +\infty$.

THEOREM 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 algorithm (IFB), the following holds true*

$$\Theta(x_k) - \min\Theta = o\left(\frac{1}{\sum_{i=1}^k t_i}\right) \quad \text{and} \quad \|x_k - x_{k-1}\| = o\left(\frac{1}{\sum_{i=1}^k t_i}\right)^{\frac{1}{2}} \quad \text{as } k \to +\infty. \tag{14}$$

As a consequence, we have

$$\Theta(x_k) - \min\Theta = o\left(\frac{1}{t_k^2}\right) \quad \text{and} \quad \|x_k - x_{k-1}\| = o\left(\frac{1}{t_k}\right) \quad \text{as } k \to +\infty, \tag{15}$$

and hence $\lim_{k\to+\infty}\Theta(x_k) = \min\Theta$, *and* $\lim_{k\to+\infty}\|x_k - x_{k-1}\| = 0$.

Proof. The energy sequence (W_k) is nonincreasing because $\alpha_k \in [0,1]$ for every $k \geq 1$, see Proposition 3. Recall from Remark 2 that $\sum_{i=1}^{+\infty} t_{i+1}W_i < +\infty$. Since (W_k) is nonincreasing, this implies that $\sum_{i=1}^{+\infty} t_{i+1}W_{i+1} < +\infty$, hence $\sum_{i=1}^{+\infty} t_iW_i < +\infty$. Let us apply Lemma 22 in the appendix, with the sequences (t_k) and (W_k), respectively in place of (τ_k) and (ε_k). We obtain that

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

The estimates (14) follow immediately. 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 estimates (15) and the last assertions. ∎

THEOREM 10. *Under (H), assume that the sequence (α_k) satisfies (K_0)-(K_1^+), together with $\alpha_k \in [0,1]$ for every $k \geq 1$. Then any sequence (x_k) generated by algorithm (IFB) converges weakly, and its limit belongs to* $\operatorname{argmin}\Theta$.

Proof. We apply the Opial lemma, see Lemma 21. Assume that there exist $\overline{x} \in H$ and a sequence (k_n) such that $k_n \to +\infty$, and $x_{k_n} \rightharpoonup \overline{x}$ weakly as $n \to +\infty$. Since the convex function Θ is lower semicontinuous, it is lower semicontinuous for the weak topology, hence satisfies

$$\Theta(\overline{x}) \leq \liminf_{n\to+\infty}\Theta(x_{k_n}) = \lim_{k\to+\infty}\Theta(x_k) = \min\Theta,$$

cf. Theorem 9. It ensues that $\overline{x} \in \operatorname{argmin}\Theta$, which shows the first point.
Let us now fix $x^* \in \operatorname{argmin}\Theta$, and show that $\lim_{k\to+\infty}\|x_k - x^*\|$ exists. For that purpose, let us set $h_k = \frac{1}{2}\|x_k - x^*\|^2$. From Proposition 4, the sequence (h_k) satisfies the following inequalities

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

Taking the positive part, we find

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

From Remark 2, we have $\sum_{k=1}^{+\infty} t_{k+1}\|x_k - x_{k-1}\|^2 < +\infty$. By applying Lemma 23 (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 $\lim_{k\to+\infty} h_k$ exists. The second point of the Opial lemma is shown, which ends the proof. $\square$

Remark 3. *A closer look at the proof of Theorem 10 shows that it suffices to assume that the nonnegative sequence (α_k) is majorized. This allows the sequence (α_k) to take values greater than 1, provided that conditions (K_0) and (K_1^+) are satisfied.*

4. Strong convergence results.

A counterexample due to Baillon [8] shows that the continuous steepest descent trajectories may converge weakly, but not strongly to a minimizer. Based on this idea, Güler [22] constructed an example for which the proximal point algorithm is not strongly convergent. Under suitable additional geometrical or topological assumptions on the potential, the trajectories of the steepest descent dynamical system are known to converge strongly. This has been shown in the case where the function Φ is even (see [15, Theorem 5]) or satisfies $\mathrm{int}(\operatorname{argmin}\Phi) \neq \emptyset$ (see [13, Theorem 3.13]). These results have been extended to inertial gradient systems, see [1] for the case of a constant damping term, and [4] for a damping of the form $\gamma(t) = \frac{\alpha}{t}$, $\alpha > 0$. This has been recently extended in [3] to a general time-dependent viscosity. The purpose of this section is to transpose some of these techniques to the framework of inertial forward-backward methods.

4.1. Even potential.

An interesting situation ensuring strong convergence is the case where the convex potential Θ is even, *i.e.* satisfies $\Theta(-x) = \Theta(x)$ for every $x \in \mathcal{H}$. Notice that we then have $0 \in \operatorname{argmin}\Theta$, since, for every $x \in \mathcal{H}$

$$\Theta(0) \leq \frac{1}{2}\Theta(x) + \frac{1}{2}\Theta(-x) = \Theta(x).$$

Theorem 11. *Under (H), assume that the sequence (α_k) satisfies (K_0)-(K_1^+), together with $\alpha_k \in [0,1]$ for every $k \geq 1$. Suppose moreover that the function Θ is even. Then any sequence (x_k) generated by algorithm (IFB) converges strongly as $k \to +\infty$ toward some $\overline{x} \in \operatorname{argmin}\Theta$.*

Proof. In the continuous case, the argument was initially developed by Bruck [15], who considered the first-order gradient flow. It has been adapted to the case of the proximal algorithms in [14]. Let us fix $n \geq 2$, and define the sequence $(q_k)_{0\leq k\leq n}$ by

$$q_k = \|x_k\|^2 - \|x_n\|^2 - \frac{1}{2}\|x_k - x_n\|^2.$$

By using equality (7) respectively with $x^* = 0$ and $x^* = x_n$, we find successively

$$\|x_{k+1}\|^2 - \|x_k\|^2 - \alpha_k(\|x_k\|^2 - \|x_{k-1}\|^2) = (\alpha_k^2 + \alpha_k)\|x_k - x_{k-1}\|^2 - 2\, s\langle G_s(y_k), y_k\rangle + s^2\|G_s(y_k)\|^2,$$

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

Let us multiply the second equality by $-1/2$, and then add to the first one. We obtain

$$q_{k+1} - q_k - \alpha_k(q_k - q_{k-1}) = \frac{1}{2}(\alpha_k^2 + \alpha_k)\|x_k - x_{k-1}\|^2 - s\langle G_s(y_k), y_k + x_n\rangle + \frac{s^2}{2}\|G_s(y_k)\|^2. \tag{16}$$

Writing inequality (4) at $y = y_k$ and $x = -x_n$ yields

$$\Theta(x_{k+1}) = \Theta(y_k - sG_s(y_k)) \leq \Theta(-x_n) + \langle G_s(y_k), y_k + x_n\rangle - \frac{s}{2}\|G_s(y_k)\|^2.$$

Using the equality $\Theta(-x_n) = \Theta(x_n)$, we then deduce from (16) that

$$q_{k+1} - q_k - \alpha_k(q_k - q_{k-1}) \leq \frac{1}{2}(\alpha_k^2 + \alpha_k)\|x_k - x_{k-1}\|^2 + s(\Theta(x_n) - \Theta(x_{k+1})).$$

Since the sequence (W_k) is nonincreasing, we have for $k \leq n-1$,

$$\Theta(x_n) \leq \frac{1}{2s}\|x_n - x_{n-1}\|^2 + \Theta(x_n) \leq \frac{1}{2s}\|x_{k+1} - x_k\|^2 + \Theta(x_{k+1}).$$

We infer that for every $k \in \{1, \dots, n-1\}$,

$$\begin{aligned} q_{k+1} - q_k - \alpha_k(q_k - q_{k-1}) &\leq \frac{1}{2}(\alpha_k^2 + \alpha_k)\|x_k - x_{k-1}\|^2 + \frac{1}{2}\|x_{k+1} - x_k\|^2 \\ &\leq \|x_k - x_{k-1}\|^2 + \frac{1}{2}\|x_{k+1} - x_k\|^2 \quad \text{since } \alpha_k \in [0,1]. \end{aligned}$$

Setting $a_k = q_k - q_{k-1}$ and $\omega_k = \|x_k - x_{k-1}\|^2 + \frac{1}{2}\|x_{k+1} - x_k\|^2$, we obtain

$$a_{k+1} \leq \alpha_k a_k + \omega_k.$$

An immediate recurrence shows that for every $k \in \{1, \dots, n-1\}$,

$$a_{k+1} \leq \left(\prod_{j=1}^{k} \alpha_j\right) a_1 + \sum_{i=1}^{k}\left[\left(\prod_{j=i+1}^{k} \alpha_j\right)\omega_i\right],$$

with the convention $\prod_{j=k+1}^{k} \alpha_j = 1$. Given $m \leq n-1$, by summing this last inequality from $k = m$ to $n-1$, we deduce that

$$q_n - q_m \leq \sum_{k=m}^{n-1}\left(\prod_{j=1}^{k} \alpha_j\right) a_1 + \sum_{k=m}^{n-1}\sum_{i=1}^{k}\left[\left(\prod_{j=i+1}^{k} \alpha_j\right)\omega_i\right].$$

Therefore, for every $m \leq n-1$, we obtain

$$\frac{1}{2}\|x_m - x_n\|^2 \leq \|x_m\|^2 - \|x_n\|^2 + \sum_{k=m}^{n-1}\left(\prod_{j=1}^{k} \alpha_j\right) a_1 + \sum_{k=m}^{n-1}\sum_{i=1}^{k}\left[\left(\prod_{j=i+1}^{k} \alpha_j\right)\omega_i\right]. \tag{17}$$

In the proof of Theorem 10, we showed that $\lim_{n\to+\infty}\|x_n - x^*\|^2$ exists for all $x^* \in \operatorname{argmin}\Theta$. Taking $x^* = 0$, we deduce that $\lim_{n\to+\infty}\|x_n\|^2$ exists. On the other hand, the series $\sum_{k\geq 1}\left(\prod_{j=1}^{k}\alpha_j\right)$ is convergent in view of assumption (K_0). Recalling the estimates $\sum_{k=1}^{+\infty} t_{k+1}\|x_{k+1} - x_k\|^2 < +\infty$ and $\sum_{k=1}^{+\infty} t_{k+1}\|x_k - x_{k-1}\|^2 < +\infty$, that follow respectively from Proposition 8 and Remark 2, we have $\sum_{k=1}^{+\infty} t_{k+1}\omega_k < +\infty$. By using Fubini Theorem and the expression (8) of the sequence (t_k), we then get

$$\sum_{k=1}^{+\infty}\sum_{i=1}^{k}\left[\left(\prod_{j=i+1}^{k}\alpha_j\right)\omega_i\right] = \sum_{i=1}^{+\infty}\omega_i\left[\sum_{k=i}^{+\infty}\left(\prod_{j=i+1}^{k}\alpha_j\right)\right] = \sum_{i=1}^{+\infty}\omega_i\, t_{i+1} < +\infty.$$

Coming back to inequality (17), we conclude that the sequence (x_n) has the Cauchy property as $n \to +\infty$, and hence converges strongly toward some $\overline{x} \in \mathcal{H}$. In view of the minimization property $\lim_{n\to+\infty}\Theta(x_n) = \min\Theta$, we conclude that $\overline{x} \in \operatorname{argmin}\Theta$. ∎

4.2. Case of a strong minimum. In this subsection, we assume that the convex function Θ has a strong minimum, that is, there exists $x^* \in \mathcal{H}$ and $\eta > 0$ such that for all $x \in \mathcal{H}$,

$$\Theta(x) \geq \Theta(x^*) + \frac{\eta}{2}\|x - x^*\|^2. \tag{18}$$

This clearly implies that $\operatorname{argmin} \Theta = \{x^*\}$. Under this condition, we are able to precisely determine the decay rate of the energy sequence (W_k) associated with the algorithm (IFB).

Theorem 12. *Under (H), assume that $\Theta : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ admits a strong minimum x^*, in the sense of (18). Suppose that the sequence $(1-\alpha_k)$ is nonincreasing, converges to 0 and satisfies $\sum_{k=1}^{+\infty}(1-\alpha_k) = +\infty$. Let (x_k) be a sequence generated by (IFB), and let (W_k) be the associated energy sequence. The following holds true*

(i) If $\alpha_{k+1} - \alpha_k = o(1-\alpha_k)$ as $k \to +\infty$, then for any $m \in]0, 2/3[$, we have for k large enough

$$W_k = \mathcal{O}\left(e^{-m\sum_{i=1}^{k}(1-\alpha_i)}\right). \tag{19}$$

Hence, $\Theta(x_k) - \min \Theta = \mathcal{O}\left(e^{-m\sum_{i=1}^{k}(1-\alpha_i)}\right)$, $\|x_k - x^*\|^2 = \mathcal{O}\left(e^{-m\sum_{i=1}^{k}(1-\alpha_i)}\right)$, $\|x_k - x_{k-1}\|^2 = \mathcal{O}\left(e^{-m\sum_{i=1}^{k}(1-\alpha_i)}\right)$. As a consequence, the iterates x_k converge strongly to the unique minimizer x^*.

(ii) If $\sum_{k=1}^{+\infty}(1-\alpha_k)^2 < +\infty$, then the estimates of (i) are satisfied for $m = 2/3$.

The proof of Theorem 12, which is quite technical, is given in the appendix. Let us illustrate the above results in some classical situations.

Corollary 13. *Under (H), assume that $\Theta : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ admits a strong minimum $x^* \in \mathcal{H}$.*

(i) If there exists $\alpha > 0$ such that $\alpha_k = 1 - \frac{\alpha}{k}$ for every $k \geq 1$, then

$$W_k = \mathcal{O}\left(k^{-\frac{2\alpha}{3}}\right) \quad \text{as } k \to +\infty.$$

(ii) If there exist $\alpha > 0$ and $r \in]1/2, 1[$ such that $\alpha_k = 1 - \frac{\alpha}{k^r}$ for every $k \geq 1$, then

$$W_k = \mathcal{O}\left(e^{-\frac{2\alpha}{3(1-r)}k^{1-r}}\right) \quad \text{as } k \to +\infty.$$

(iii) If there exist $\alpha > 0$ and $r \in]0, 1[$ such that $\alpha_k = 1 - \frac{\alpha}{k^r}$ for every $k \geq 1$, then for any $m \in]0, 2/3[$, we have

$$W_k = \mathcal{O}\left(e^{-\frac{m\alpha}{1-r}k^{1-r}}\right) \quad \text{as } k \to +\infty.$$

Proof. (i) The condition $\sum_{k=1}^{+\infty}(1-\alpha_k)^2 < +\infty$ is clearly satisfied. It is then sufficient to apply Theorem 12 (ii) and to recall that

$$\sum_{i=1}^{k} \frac{1}{i} = \ln k + \gamma + o(1) \quad \text{as } k \to +\infty,$$

for some $\gamma \in \mathbb{R}$ (this constant is known as the Euler constant).
(ii) The condition $\sum_{k=1}^{+\infty}(1-\alpha_k)^2 < +\infty$ is guaranteed by the assumption $r \in]1/2, 1[$. Then apply Theorem 12 (ii), together with the following asymptotic expansion

$$\sum_{i=1}^{k} \frac{1}{i^r} = \frac{k^{1-r}}{1-r} + l + o(1) \quad \text{as } k \to +\infty, \tag{20}$$

for some $l \in \mathbb{R}$.
(iii) The condition $\alpha_{k+1} - \alpha_k = o(1-\alpha_k)$ is satisfied as $k \to +\infty$, hence the announced estimate is a consequence of Theorem 12 (i), combined with the equality (20). ∎

REMARK 4. *At the best knowledge of the authors, the results stated in Theorem 12 and Corollary 13 are new in the framework of forward-backward algorithms. Under a strong convexity assumption on* Θ*, a partial result is obtained by Su, Boyd, and Candès [35] in the case* $\alpha_k = \frac{k-1}{k+\alpha-1}$*. These authors show that* $\mathcal{O}(1/k^3)$ *convergence rate holds true as* $k \to +\infty$*, provided that* $\alpha \geq 9/2$*. Such a result can be immediately recovered from Theorem 12* (ii)*, without requiring the strong convexity of* Θ *(it is enough to assume that we have a strong minimum). This being said, the results obtained in Theorem 12 and Corollary 13 are at this stage theoretical, and need further developments.*

5. Application to special classes of sequences (α_k)**.** In this section, we assume that there exists $c \in [0, 1[$ such that for every $k \geq 1$,

$$\frac{1}{1-\alpha_{k+1}} - \frac{1}{1-\alpha_k} \leq c. \tag{21}$$

PROPOSITION 14. *Let* $c \in [0, 1[$ *and let* (α_k) *be a sequence satisfying* $\alpha_k \in [0, 1[$ *together with inequality (21) for every* $k \geq 1$*. Then the following holds true*

(i) *Condition* (K_0) *is satisfied and we have for every* $k \geq 1$*,*

$$t_{k+1} \leq \frac{1}{(1-c)(1-\alpha_k)}.$$

(ii) *If* $c \leq 1/3$ *(resp.* $c < 1/3$*), then condition* (K_1) *(resp.* (K_1^+)*) is fulfilled.*

Proof. (i) Let us fix $k \geq 1$. Observe that for every $i \geq k$

$$\begin{aligned} \frac{1}{1-\alpha_{i+1}} \prod_{j=k+1}^{i+1} \alpha_j - \frac{1}{1-\alpha_i} \prod_{j=k+1}^{i} \alpha_j &= \left(\frac{\alpha_{i+1}}{1-\alpha_{i+1}} - \frac{1}{1-\alpha_i}\right) \prod_{j=k+1}^{i} \alpha_j \\ &= \left(-1 + \frac{1}{1-\alpha_{i+1}} - \frac{1}{1-\alpha_i}\right) \prod_{j=k+1}^{i} \alpha_j \\ &\leq -(1-c) \prod_{j=k+1}^{i} \alpha_j \quad \text{in view of (21).} \end{aligned} \tag{22}$$

By summing this inequality from $i = k$ to $n-1$, we find

$$(1-c) \sum_{i=k}^{n-1} \prod_{j=k+1}^{i} \alpha_j \leq \frac{1}{1-\alpha_k} - \frac{1}{1-\alpha_n} \prod_{j=k+1}^{n} \alpha_j \leq \frac{1}{1-\alpha_k}.$$

Taking the limit as $n \to +\infty$, we deduce that

$$(1-c) \sum_{i=k}^{+\infty} \prod_{j=k+1}^{i} \alpha_j \leq \frac{1}{1-\alpha_k} < +\infty.$$

In view of the expression (8) of the sequence (t_k), we infer that $(1-c)\, t_{k+1} \leq 1/(1-\alpha_k)$ for every $k \geq 1$, and hence condition (K_0) holds true.

(ii) From what precedes, we obtain that $t_{k+1}(1-\alpha_k) \leq 1/(1-c)$ for every $k \geq 1$. If $c \leq 1/3$ (resp. $c < 1/3$), then we have $1/(1-c) \leq 3/2$ (resp. $< 3/2$) and we deduce from Lemma 24 that (K_1) (resp. (K_1^+)) holds true. ∎

The following Proposition provides a criterion for simply obtaining an asymptotic equivalent of t_k. Its proof, a little technical, is given in appendix.

PROPOSITION 15. *Let (α_k) be a sequence such that $\alpha_k \in [0,1[$ for every $k \geq 1$. Given $c \in [0,1[$, assume that*

$$\lim_{k\to+\infty} \frac{1}{1-\alpha_{k+1}} - \frac{1}{1-\alpha_k} = c. \tag{23}$$

Then we have

$$t_{k+1} \sim \frac{1}{(1-c)(1-\alpha_k)} \quad \text{as } k \to +\infty.$$

Let us now apply the results of the previous sections to the class of sequences (α_k) satisfying inequality (21).

THEOREM 16. *Under hypothesis (H), assume that the sequence (α_k) is such that $\alpha_k \in [0,1[$ for every $k \geq 1$. Given $c \in [0,1/3]$, suppose that $1/(1-\alpha_{k+1}) - 1/(1-\alpha_k) \leq c$ for every $k \geq 1$. Let (x_k) be a sequence generated by (IFB). Then we have*

(i) $\Theta(x_k) - \min\Theta = \mathcal{O}\left(\frac{1}{t_k^2}\right)$ *as $k \to +\infty$;*

Assuming now that $c \in [0,1/3[$, the following holds true

(ii) $\sum_{k=1}^{+\infty} t_{k+1}(\Theta(x_k) - \min\Theta) < +\infty$;

(iii) $\sum_{k=1}^{+\infty} t_k\|x_k - x_{k-1}\|^2 < +\infty$;

(iv) $\Theta(x_k) - \min\Theta = o\left(\frac{1}{\sum_{i=1}^k t_i}\right)$ *and* $\|x_k - x_{k-1}\| = o\left(\frac{1}{\sum_{i=1}^k t_i}\right)^{\frac{1}{2}}$ *as $k \to +\infty$;*

(v) The sequence (x_k) converges weakly as $k \to +\infty$ toward some $\overline{x} \in \operatorname{argmin}\Theta$. If the function Θ is even, then the convergence is strong in $\mathcal{H}$.

Proof. (i) Assume that there exists $c \in [0,1/3]$ such that $1/(1-\alpha_{k+1}) - 1/(1-\alpha_k) \leq c$ for every $k \geq 1$. Proposition 14 shows that conditions (K_0) and (K_1) are satisfied. It is then sufficient to apply Theorem 7 (i).
(ii) Assuming that $c < 1/3$, the stronger condition (K_1^+) is satisfied, see Proposition 14 (ii). We then use Theorem 7 (ii).
(iii) The announced estimate is furnished by Proposition 8.
(iv) This point is a consequence of Theorem 9.
(v) The weak convergence of (x_k) as $k \to +\infty$ follows from Theorem 10. The strong convergence result follows from Theorem 11. ∎

It is easy to see that the conclusions of Theorem 16 remain unchanged if we assume only that $1/(1-\alpha_{k+1}) - 1/(1-\alpha_k) \leq c$ for k large enough. This elementary fact will be used in the examples of subsections 5.2 and 5.3. Let us now consider several particular cases of Theorem 16.

5.1. Case $\alpha_k = 1 - \frac{\alpha}{k}$ for some $\alpha > 0$. Observe that, for every $k \geq 1$,

$$\frac{1}{1-\alpha_{k+1}} - \frac{1}{1-\alpha_k} = \frac{1}{\alpha}(k+1) - \frac{1}{\alpha}k = \frac{1}{\alpha},$$

hence we are in the framework of Theorem 16 with $c = \frac{1}{\alpha}$. If $\alpha \geq 3$ (resp. $\alpha > 3$), we have $c \in]0,1/3]$ (resp. $c \in]0,1/3[$). By arguing as in Proposition 15 (with an equality in place of the equivalence), we easily obtain, for $k \geq 1$

$$t_{k+1} = \frac{1}{(1-c)(1-\alpha_k)} = \frac{\alpha}{\alpha-1}\,\frac{k}{\alpha} = \frac{k}{\alpha-1}.$$

By applying Theorem 16, we obtain the following statement.

Corollary 17. *Assume hypothesis (H). Given $\alpha \geq 3$, suppose that $\alpha_k = 1 - \frac{\alpha}{k}$ for every $k \geq 1$. Let (x_k) be a sequence generated by (IFB). Then we have*

(i) $\Theta(x_k) - \min \Theta = \mathcal{O}\left(\frac{1}{k^2}\right)$ *as $k \to +\infty$;*

Assuming now that $\alpha > 3$, the following holds true

(ii) $\sum_{k=1}^{+\infty} k(\Theta(x_k) - \min \Theta) < +\infty$*;*

(iii) $\sum_{k=1}^{+\infty} k\|x_k - x_{k-1}\|^2 < +\infty$*;*

(iv) $\Theta(x_k) - \min \Theta = o\left(\frac{1}{k^2}\right)$ *and* $\|x_k - x_{k-1}\| = o\left(\frac{1}{k}\right)$ *as $k \to +\infty$;*

(v) The sequence (x_k) converges weakly as $k \to +\infty$ toward some $\overline{x} \in \operatorname{argmin} \Theta$. If the function Θ is even, then the convergence is strong in $\mathcal{H}$.

Remark 5. *By taking a sequence (α_k) of the form $\alpha_k = \frac{k}{k+\alpha}$, we obtain exactly the same results as in Corollary 17, thus allowing to recover the results of [4, 5, 19].*

5.2. Case $\alpha_k = 1 - \frac{(\ln k)^\theta}{k}$ for some $\theta > 0$. Observe that

$$\frac{1}{1-\alpha_{k+1}} - \frac{1}{1-\alpha_k} = \frac{k+1}{(\ln(k+1))^\theta} - \frac{k}{(\ln k)^\theta} = f(k+1) - f(k),$$

where $f(t) = \frac{t}{(\ln t)^\theta}$. After computing the derivative of f, let $f'(t) = \frac{1}{(\ln t)^\theta} - \frac{\theta}{(\ln t)^{\theta+1}}$, the mean value theorem gives

$$|f(k+1) - f(k)| \leq \sup_{t \in [k,k+1]} |f'(t)| \leq \frac{1}{(\ln k)^\theta} + \frac{\theta}{(\ln k)^{\theta+1}},$$

that tends to zero as $k \to +\infty$. For each $c > 0$, the condition $1/(1-\alpha_{k+1}) - 1/(1-\alpha_k) \leq c$ is satisfied for k large enough. We deduce from Proposition 15 that $t_k \sim \frac{k}{(\ln k)^\theta}$ as $k \to +\infty$. Fix $i_0 \geq 2$ such that $f'(t)$ is nonnegative for every $t \geq i_0$. By summing the previous equivalences, we obtain

$$\sum_{i=1}^{k} t_i \sim \sum_{i=i_0}^{k} \frac{i}{(\ln i)^\theta} = \sum_{i=i_0}^{k} f(i) \quad \text{as } k \to +\infty. \tag{24}$$

Since the function f is nondecreasing on $[i_0, +\infty[$, we have

$$f(i_0) + \int_{i_0}^{k} f(t)\,dt \leq \sum_{i=i_0}^{k} f(i) \leq \int_{i_0}^{k+1} f(t)\,dt. \tag{25}$$

Using an integration by parts, it can be easily shown that

$$\int_{i_0}^{k} f(t)\,dt \sim \frac{k^2}{2(\ln k)^\theta} \quad \text{as } k \to +\infty.$$

From the inequalities (25), we deduce that $\sum_{i=i_0}^{k} f(i) \sim \frac{k^2}{2(\ln k)^\theta}$ as $k \to +\infty$. In view of (24), this implies in turn that $\sum_{i=1}^{k} t_i \sim \frac{k^2}{2(\ln k)^\theta}$ as $k \to +\infty$. From this, and by applying Theorem 16, we obtain directly the following statement.

Corollary 18. *Assume hypothesis (H). Given some $\theta > 0$, let us assume that $\alpha_k = 1 - \frac{(\ln k)^\theta}{k}$. Let (x_k) be a sequence generated by (IFB). Then we have* *(i)* $\displaystyle\sum_{k=1}^{+\infty} \frac{k}{(\ln k)^\theta}(\Theta(x_k) - \min \Theta) < +\infty$*;*

(ii) $\displaystyle\sum_{k=1}^{+\infty} \frac{k}{(\ln k)^\theta}\|x_k - x_{k-1}\|^2 < +\infty$*;*

(iii) $\Theta(x_k) - \min \Theta = o\left(\frac{(\ln k)^\theta}{k^2}\right)$ *and* $\|x_k - x_{k-1}\| = o\left(\frac{(\ln k)^\theta}{k^2}\right)^{\frac{1}{2}}$ *as* $k \to +\infty$*;*

(iv) The sequence (x_k) *converges weakly as* $k \to +\infty$ *toward some* $\overline{x} \in \operatorname{argmin} \Theta$*.*
If the function Θ *is even, then the convergence is strong in* $\mathcal{H}$*.*

5.3. Case $\alpha_k = 1 - \frac{\alpha}{k^r}$ **for some** $\alpha > 0$ **and** $r \in]0,1[$**.** Observe that

$$\frac{1}{1-\alpha_{k+1}} - \frac{1}{1-\alpha_k} = \frac{1}{\alpha}(k+1)^r - \frac{1}{\alpha}k^r = \frac{k^r}{\alpha}\left((1+1/k)^r - 1\right) \sim \frac{r}{\alpha}\, k^{r-1} \to 0 \quad \text{as } k \to +\infty.$$

For each $c > 0$, the condition $1/(1-\alpha_{k+1}) - 1/(1-\alpha_k) \le c$ is satisfied for k large enough. On the other hand, we deduce from Proposition 15 that $t_k \sim \frac{k^r}{\alpha}$ as $k \to +\infty$. This implies immediately that $\sum_{i=1}^k t_i \sim \frac{1}{\alpha(1+r)}\, k^{1+r}$ as $k \to +\infty$. By applying Theorem 16, we obtain directly the following statement.

Corollary 19. *Assume hypothesis* (H)*. Given* $\alpha > 0$ *and* $r \in]0,1[$*, suppose that* $\alpha_k = 1 - \frac{\alpha}{k^r}$ *for every* $k \ge 1$*. Let* (x_k) *be a sequence generated by* (IFB)*. Then we have*

(i) $\displaystyle\sum_{k=1}^{+\infty} k^r(\Theta(x_k) - \min \Theta) < +\infty$*;*

(ii) $\displaystyle\sum_{k=1}^{+\infty} k^r\|x_k - x_{k-1}\|^2 < +\infty$*;*

(iii) $\Theta(x_k) - \min \Theta = o\left(\frac{1}{k^{1+r}}\right)$ *and* $\|x_k - x_{k-1}\| = o\left(\frac{1}{k^{1+r}}\right)^{\frac{1}{2}}$ *as* $k \to +\infty$*;*

(iv) The sequence (x_k) *converges weakly as* $k \to +\infty$ *toward some* $\overline{x} \in \operatorname{argmin} \Theta$*.*
If the function Θ *is even, then the convergence is strong in* $\mathcal{H}$*.*

5.4. Case $\alpha_k \in [m, M]$ **with** $0 \le m \le M < 1$**.** This contains the case where the sequence is constant, thus allowing to recover the Polyak's heavy ball method as a particular case.

Corollary 20. *Under hypothesis* (H)*, assume that the sequence* (α_k) *satisfies* $\alpha_k \in [m, M]$ *for every* $k \ge 1$*, with* $0 \le m \le M < 1$*. Suppose moreover that* $(1-m)/(1-M) < 3/2$*. Let* (x_k) *be a sequence generated by* (IFB)*. Then we have*

(i) $\displaystyle\sum_{k=1}^{+\infty}(\Theta(x_k) - \min \Theta) < +\infty$*;*

(ii) $\displaystyle\sum_{k=1}^{+\infty} \|x_k - x_{k-1}\|^2 < +\infty$*;*

(iii) $\Theta(x_k) - \min \Theta = o\left(\frac{1}{k}\right)$ *and* $\|x_k - x_{k-1}\| = o\left(\frac{1}{k^{1/2}}\right)$ *as* $k \to +\infty$*;*

(iv) The sequence (x_k) *converges weakly as* $k \to +\infty$ *toward some* $\overline{x} \in \operatorname{argmin} \Theta$*.*
If the function Θ *is even, then the convergence is strong in* $\mathcal{H}$*.*

Proof. Since $\alpha_k \le M < 1$ for every $k \ge 1$, we deduce from the expression of t_k that for every $k \ge 1$

$$t_k \le 1 + \sum_{i=k}^{+\infty} M^{i-k+1} = \frac{1}{1-M} < +\infty,$$

and hence condition (K_0) holds true. On the other hand, since $\alpha_k \geq m$ for $k \geq 1$, we obtain

$$t_{k+1}(1-\alpha_k) \leq \frac{1-m}{1-M}.$$

Recalling that the right member above is less than 3/2 by assumption, we infer from Lemma 24 that condition (K_1^+) is satisfied. Using that $t_k \geq 1$ for every $k \geq 1$, we have $\sum_{i=1}^{k} t_i \geq k$, and items (i)-(iv) are deduced from the corresponding results of the previous sections. ∎

REMARK 6. *If the nonnegative sequence (α_k) converges toward some $l \in [0,1[$, then the assumptions of Corollary 20 are satisfied for k large enough. Indeed, for $\varepsilon > 0$, we take $m = \max\{0, l-\varepsilon\}$, $M = l + \varepsilon$, and we choose ε sufficiently small so as to have $M < 1$ and $(1-m)/(1-M) < 3/2$.*

The following table gives a synthetic summary about the speed of convergence in the different situations studied above. Recall that $W_k \to 0$ reflects both the rate of convergence to zero of the values $\Theta(x_k) - \min\Theta$, and of the velocities $\|x_k - x_{k-1}\|$. The parameters α and θ are supposed to be positive. From left to right, the table is ordered by decreasing rate of convergence for W_k.

α_k	$\alpha_k = 1 - \frac{\alpha}{k},\ \alpha > 3$	$\alpha_k = 1 - \frac{3}{k}$	$\alpha_k = 1 - \frac{(\ln k)^\theta}{k}$	$\alpha_k = 1 - \frac{\alpha}{k^r},\ \ r \in]0,1[$	$\alpha_k \to l \in [0,1[$
t_k	$\frac{k-1}{\alpha-1}$	$\frac{k-1}{2}$	$\sim \frac{k}{(\ln k)^\theta}$	$\sim \frac{k^r}{\alpha}$	$\sim \frac{1}{1-l}$
W_k	$o\left(\frac{1}{k^2}\right)$	$\mathcal{O}\left(\frac{1}{k^2}\right)$	$o\left(\frac{(\ln k)^\theta}{k^2}\right)$	$o\left(\frac{1}{k^{r+1}}\right)$	$o\left(\frac{1}{k}\right)$

6. Appendix.

6.1. Some auxiliary results. In this section, we present some auxiliary lemmas that are used throughout the paper. To establish the weak convergence of the iterates of (IFB), we apply Opial's Lemma [31], that we recall in its discrete form.

LEMMA 21. *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 $k \to +\infty$, belongs to S;
(ii) for every $z \in S$, $\lim_{k\to+\infty} \|x_k - z\|$ exists.

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

Owing to the next lemma, we are able to estimate the rate of convergence of a sequence (ε_k) supposed to be nonincreasing and summable with respect to weight coefficients.

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*

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

Proof. For $k \geq 1$, let us set

$$a_k = \max\left\{n \geq 1,\ \sum_{i=1}^{n} \tau_i \leq \frac{1}{2}\sum_{i=1}^{k} \tau_i\right\}.$$

Due to the assumption $\sum_{i=1}^{+\infty} \tau_i = +\infty$, the term a_k is well-defined for k large enough. The sequence (a_k) is nondecreasing and satisfies $\lim_{k\to+\infty} a_k = +\infty$. By definition, we have

$$\sum_{i=1}^{a_k} \tau_i \leq \frac{1}{2}\sum_{i=1}^{k} \tau_i, \quad \text{hence} \quad \sum_{i=a_k+1}^{k} \tau_i \geq \frac{1}{2}\sum_{i=1}^{k} \tau_i.$$

Recalling that the sequence (ε_k) is nonincreasing, we obtain

$$\sum_{i=a_k+1}^{k} \tau_i\varepsilon_i \geq \varepsilon_k \sum_{i=a_k+1}^{k} \tau_i \geq \frac{1}{2}\varepsilon_k \sum_{i=1}^{k} \tau_i.$$

By assumption, we have $\sum_{i=1}^{+\infty} \tau_i\varepsilon_i < +\infty$. The sequence (a_k) satisfies $\lim_{k\to+\infty} a_k = +\infty$. We deduce that $\lim_{k\to+\infty} \sum_{i=a_k+1}^{k} \tau_i\varepsilon_i = 0$. The conclusion follows from the above inequality. ∎

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

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*

$$a_{k+1} \leq \alpha_k a_k + \omega_k, \tag{26}$$

for all $k \geq 0$. If $\sum_{k=0}^{+\infty} t_{k+1}\omega_k < +\infty$, then $\sum_{k=0}^{+\infty} a_k < +\infty$.

Proof. By Lemma 5, we have $t_{k+1}\alpha_k = t_k - 1$. Multiplying inequality (26) by t_{k+1} gives

$$t_{k+1}a_{k+1} \leq (t_k - 1)a_k + t_{k+1}\omega_k,$$

or equivalently $a_k \leq (t_k a_k - t_{k+1}a_{k+1}) + t_{k+1}\omega_k$. By summing from $k = 0$ to n, we deduce that

$$\begin{aligned}\sum_{k=0}^{n} a_k &\leq t_0 a_0 - t_{n+1}a_{n+1} + \sum_{k=0}^{n} t_{k+1}\omega_k \\ &\leq t_0 a_0 + \sum_{k=0}^{+\infty} t_{k+1}\omega_k < +\infty \quad \text{by assumption.}\end{aligned}$$

The conclusion follows by letting n tend to $+\infty$. ∎

The following lemma gives practical conditions ensuring that the assumptions (K_1) and (K_1^+) are satisfied.

Lemma 24. *Assumption (K_1) (resp. (K_1^+)) is satisfied if there exists $\mu \leq 3/2$ (resp. $\mu < 3/2$) such that for every $k \geq 1$,*

$$t_{k+1}(1 - \alpha_k) \leq \mu.$$

Proof. Let $m \leq 1$. Since $t_k = \alpha_k t_{k+1} + 1$, the assumption $t_{k+1}^2 - t_k^2 \leq m\, t_{k+1}$ can be rewritten as

$$t_{k+1}^2 - (\alpha_k t_{k+1} + 1)^2 \leq m\, t_{k+1}$$

$$\Updownarrow$$

$$t_{k+1}^2(1 - \alpha_k^2) - t_{k+1}(m + 2\alpha_k) - 1 \leq 0.$$

If $t_{k+1}(1 - \alpha_k) \leq 1 + m/2$, then we have

$$\begin{aligned}t_{k+1}^2(1 - \alpha_k^2) - t_{k+1}(m + 2\alpha_k) - 1 &\leq t_{k+1}(1 + \alpha_k)(1 + m/2) - t_{k+1}(m + 2\alpha_k) - 1 \\ &= (1 - m/2)t_{k+1}(1 - \alpha_k) - 1 \\ &\leq (1 - m/2)(1 + m/2) - 1 = -m^2/4 \leq 0,\end{aligned}$$

thus implying that $t_{k+1}^2 - t_k^2 \leq m\, t_{k+1}$. The conclusion follows by taking $m = 1$ (resp. $m < 1$). ∎

6.2. Proof of Theorem 12. Recall from Proposition 4 that the sequence (h_k) defined by $h_k = \frac{1}{2}\|x_k - x^*\|^2$ satisfies for every $k \geq 1$,

$$\begin{aligned} h_{k+1} - h_k - \alpha_k(h_k - h_{k-1}) &\leq \frac{1}{2}(\alpha_k^2 + \alpha_k)\|x_k - x_{k-1}\|^2 - s(\Theta(x_{k+1}) - \min\Theta) \\ &\leq \|x_k - x_{k-1}\|^2 - s(\Theta(x_{k+1}) - \min\Theta). \end{aligned}$$

From the definition of (W_k), it follows that

$$h_{k+1} - h_k - \alpha_k(h_k - h_{k-1}) + sW_{k+1} \leq \|x_k - x_{k-1}\|^2 + \frac{1}{2}\|x_{k+1} - x_k\|^2. \tag{27}$$

Recalling the expression (6) of the decay of (W_k), we have

$$\|x_k - x_{k-1}\|^2 \leq \frac{2s}{1-\alpha_k^2}(W_k - W_{k+1}) \quad \text{and} \quad \|x_{k+1} - x_k\|^2 \leq \frac{2s}{1-\alpha_{k+1}^2}(W_{k+1} - W_{k+2}).$$

Multiplying inequality (27) by $1 - \alpha_{k+1}^2$ and using that $\alpha_k \leq \alpha_{k+1}$, we obtain

$$(1-\alpha_{k+1}^2)[h_{k+1} - h_k - \alpha_k(h_k - h_{k-1})] + s(1-\alpha_{k+1}^2)W_{k+1} \leq 2s(W_k - W_{k+1}) + s(W_{k+1} - W_{k+2}). \tag{28}$$

Let us now define the sequence $(\widehat{W}_k)$ by $\widehat{W}_k = \frac{2}{3}W_k + \frac{1}{3}W_{k+1}$. Since the sequence (W_k) is nonnegative and nonincreasing, we have

$$\frac{2}{3}W_k \leq \widehat{W}_k \leq W_k. \tag{29}$$

We immediately deduce from (28) that for every $k \geq 1$,

$$(1-\alpha_{k+1}^2)[h_{k+1} - h_k - \alpha_k(h_k - h_{k-1})] + s\left[(1-\alpha_{k+1}^2)\widehat{W}_{k+1} + 3(\widehat{W}_{k+1} - \widehat{W}_k)\right] \leq 0. \tag{30}$$

Now observe that

$$\begin{aligned} (1-\alpha_{k+1}^2)[h_{k+1} - h_k - \alpha_k(h_k - h_{k-1})] &= (1-\alpha_{k+1}^2)(h_{k+1} - h_k) - (1-\alpha_{k+2}^2)\alpha_{k+1}(h_{k+1} - h_k) \\ &\quad +(1-\alpha_{k+2}^2)\alpha_{k+1}(h_{k+1} - h_k) - (1-\alpha_{k+1}^2)\alpha_k(h_k - h_{k-1}) \\ &= \big(1-\alpha_{k+1}^2 - (1-\alpha_{k+2}^2)\alpha_{k+1}\big)(h_{k+1} - h_k) \\ &\quad +(1-\alpha_{k+2}^2)\alpha_{k+1}(h_{k+1} - h_k) - (1-\alpha_{k+1}^2)\alpha_k(h_k - h_{k-1}). \end{aligned} \tag{31}$$

Let us introduce the sequence $(\widetilde{W}_k)$ given by

$$\widetilde{W}_k = \widehat{W}_k + \frac{1}{3s}(1-\alpha_{k+1}^2)\alpha_k(h_k - h_{k-1}). \tag{32}$$

Dividing inequality (30) by $3s$ and using (31), we infer that

$$\frac{1}{3}(1-\alpha_{k+1}^2)\widehat{W}_{k+1} + \widetilde{W}_{k+1} - \widetilde{W}_k \leq \frac{1}{3s}\left[(1-\alpha_{k+2}^2)\alpha_{k+1} - (1-\alpha_{k+1}^2)\right](h_{k+1} - h_k),$$

which can be rewritten as

$$\begin{aligned} &\frac{1}{3}(1-\alpha_{k+1}^2)\widetilde{W}_{k+1} + \widetilde{W}_{k+1} - \widetilde{W}_k \leq \\ &\qquad \frac{1}{3s}\left[\frac{1}{3}(1-\alpha_{k+1}^2)(1-\alpha_{k+2}^2)\alpha_{k+1} + (1-\alpha_{k+2}^2)\alpha_{k+1} - (1-\alpha_{k+1}^2)\right](h_{k+1} - h_k). \end{aligned} \tag{33}$$

Using that the sequence (α_k) is nondecreasing and satisfies $\alpha_k \in [0,1]$ for every $k \geq 1$, we have

$$0 \leq (1-\alpha_{k+1}^2)(1-\alpha_{k+2}^2)\alpha_{k+1} \leq (1-\alpha_{k+1}^2)^2\alpha_{k+1} \leq 4(1-\alpha_{k+1})^2$$

and

$$\begin{aligned} 0 \leq (1-\alpha_{k+1}^2) - (1-\alpha_{k+2}^2)\alpha_{k+1} &= (1-\alpha_{k+1}^2)(1-\alpha_{k+1}) + (\alpha_{k+2}^2 - \alpha_{k+1}^2)\alpha_{k+1} \\ &\leq 2(1-\alpha_{k+1})^2 + 2(\alpha_{k+2}-\alpha_{k+1}). \end{aligned}$$

It ensues that the term between brackets in (33) is comprised between $-2(1-\alpha_{k+1})^2 - 2(\alpha_{k+2}-\alpha_{k+1})$ and $\frac{4}{3}(1-\alpha_{k+1})^2$. This implies that its absolute value is majorized by $2(1-\alpha_{k+1})^2 + 2(\alpha_{k+2}-\alpha_{k+1})$. We then deduce from (33) that

$$\frac{1}{3}(1-\alpha_{k+1}^2)\widetilde{W}_{k+1} + \widetilde{W}_{k+1} - \widetilde{W}_k \leq \frac{2}{3s}\left[(1-\alpha_{k+1})^2 + (\alpha_{k+2}-\alpha_{k+1})\right]|h_{k+1}-h_k|. \tag{34}$$

Now observe that

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

Since $|\langle x_{k+1}-x_k, x_{k+1}-x^*\rangle| \leq \frac{1}{2}\|x_{k+1}-x_k\|^2 + \frac{1}{2}\|x_{k+1}-x^*\|^2$, we infer that

$$|h_{k+1}-h_k| \leq \|x_{k+1}-x_k\|^2 + \frac{1}{2}\|x_{k+1}-x^*\|^2.$$

Recalling that $\frac{\eta}{2}\|x_{k+1}-x^*\|^2 \leq \Theta(x_{k+1}) - \min\Theta$ by assumption, we find

$$\begin{aligned} |h_{k+1}-h_k| &\leq \|x_{k+1}-x_k\|^2 + \frac{1}{\eta}(\Theta(x_{k+1}) - \min\Theta) \\ &\leq 2sC\,W_{k+1} \quad \text{with } C = \max\{1, 1/(2s\eta)\} \\ &\leq 3sC\,\widehat{W}_{k+1} \quad \text{in view of (29)}. \end{aligned} \tag{35}$$

Since $\lim_{k\to+\infty}\alpha_k = 1$, the expression (32) of $\widetilde{W}_k$ and inequality (35) show that

$$\widetilde{W}_k = \widehat{W}_k + o(\widehat{W}_k) \quad \text{as } k \to +\infty. \tag{36}$$

Let $C' > C$. By combining (34), (35) and (36), we obtain the existence of $k_0 \geq 1$ such that for every $k \geq k_0$,

$$\frac{1}{3}(1-\alpha_{k+1}^2)\widetilde{W}_{k+1} + \widetilde{W}_{k+1} - \widetilde{W}_k \leq 2C'\,\widetilde{W}_{k+1}\left[(1-\alpha_{k+1})^2 + (\alpha_{k+2}-\alpha_{k+1})\right].$$

Noting that $1-\alpha_{k+1}^2 = 2(1-\alpha_{k+1}) - (1-\alpha_{k+1})^2$, the above inequality can be rewritten as

$$\widetilde{W}_{k+1}\left(1 + \frac{2}{3}(1-\alpha_{k+1}) - u_{k+1}\right) \leq \widetilde{W}_k,$$

where the sequence (u_k) is defined by

$$u_k = \left(\frac{1}{3} + 2C'\right)(1-\alpha_k)^2 + 2C'(\alpha_{k+1}-\alpha_k).$$

Let $n \geq k_0 + 1$. By multiplying the above inequality from $k = k_0$ to $n - 1$, we derive that

$$\widetilde{W}_n \leq \frac{\widetilde{W}_{k_0}}{\prod_{k=k_0+1}^{n}\left(1 + \frac{2}{3}(1-\alpha_k) - u_k\right)} = \widetilde{W}_{k_0}\, e^{-\left[\sum_{k=k_0+1}^{n} \ln\left(1+\frac{2}{3}(1-\alpha_k)-u_k\right)\right]}. \tag{37}$$

(i) Let us now fix $m \in]0, 2/3[$ and assume that $\alpha_{k+1} - \alpha_k = o(1-\alpha_k)$ as $k \to +\infty$. Since $\lim_{k\to+\infty} \alpha_k = 1$, the expression of u_k shows that $u_k = o(1-\alpha_k)$ as $k \to +\infty$. It ensues that

$$\ln\left(1 + \frac{2}{3}(1-\alpha_k) - u_k\right) = \frac{2}{3}(1-\alpha_k) + o(1-\alpha_k) \quad \text{as } k \to +\infty,$$

and hence for k large enough,

$$\ln\left(1 + \frac{2}{3}(1-\alpha_k) - u_k\right) \geq m\,(1-\alpha_k).$$

Coming back to (37), we easily deduce that

$$\widetilde{W}_n = \mathcal{O}\left(e^{-m\sum_{k=1}^{n}(1-\alpha_k)}\right) \quad \text{as } n \to +\infty.$$

In view of (29) and (36), we immediately derive the estimate (19). The other estimates follow directly.
(ii) Let us now assume that $\sum_{k=1}^{+\infty}(1-\alpha_k)^2 < +\infty$. Observe that $\sum_{k=1}^{+\infty}(\alpha_{k+1}-\alpha_k) < +\infty$, since the sequence (α_k) is nondecreasing and tends to 1 as $k \to +\infty$. The expression of (u_k) then shows that $\sum_{k=1}^{+\infty} u_k < +\infty$. Using that $\lim_{k\to+\infty}\alpha_k = 1$ and $\lim_{k\to+\infty} u_k = 0$, we have for k large enough,

$$\ln\left(1 + \frac{2}{3}(1-\alpha_k) - u_k\right) = \frac{2}{3}(1-\alpha_k) - u_k - \frac{1}{2}\left[\frac{2}{3}(1-\alpha_k) - u_k\right]^2 + o\left(\left[\frac{2}{3}(1-\alpha_k) - u_k\right]^2\right).$$

Since $\sum_{k=1}^{+\infty}(1-\alpha_k)^2 < +\infty$ and $\sum_{k=1}^{+\infty} u_k^2 < +\infty$ (recall that (u_k) is summable from what precedes), we obtain

$$\sum_{k=1}^{+\infty}\left[\frac{2}{3}(1-\alpha_k) - u_k\right]^2 < +\infty.$$

Defining the sequence (v_k) by

$$v_k = \ln\left(1 + \frac{2}{3}(1-\alpha_k) - u_k\right) - \frac{2}{3}(1-\alpha_k),$$

we deduce from what precedes that the series $\sum_k v_k$ is convergent. It ensues that

$$\begin{aligned}\sum_{k=k_0+1}^{n} \ln\left(1 + \frac{2}{3}(1-\alpha_k) - u_k\right) &= \frac{2}{3}\sum_{k=k_0+1}^{n}(1-\alpha_k) + \sum_{k=k_0+1}^{n} v_k \\ &= \frac{2}{3}\sum_{k=k_0+1}^{n}(1-\alpha_k) + \sum_{k=k_0+1}^{+\infty} v_k + o(1) \quad \text{as } n \to +\infty.\end{aligned}$$

Coming back to (37), we easily deduce that

$$\widetilde{W}_n = \mathcal{O}\left(e^{-\frac{2}{3}\sum_{k=1}^{n}(1-\alpha_k)}\right) \quad \text{as } n \to +\infty.$$

The conclusion then follows from (29) and (36).

6.3. Proof of Proposition 15. Fix $\varepsilon \in]0, 1-c[$. In view of (23), there exists $i_0 \geq 1$ such that for every $i \geq i_0$,

$$c - \varepsilon \leq \frac{1}{1-\alpha_{i+1}} - \frac{1}{1-\alpha_i} \leq c + \varepsilon. \tag{38}$$

We deduce from equality (22) that for every $i \geq i_0$,

$$(1-c-\varepsilon)\prod_{j=k+1}^{i}\alpha_j \leq \frac{1}{1-\alpha_i}\prod_{j=k+1}^{i}\alpha_j - \frac{1}{1-\alpha_{i+1}}\prod_{j=k+1}^{i+1}\alpha_j \leq (1-c+\varepsilon)\prod_{j=k+1}^{i}\alpha_j.$$

Take integers k, n such that $i_0 \leq k \leq n-1$ and let us sum the above inequalities from $i = k$ to $n-1$. We find

$$(1-c-\varepsilon)\sum_{i=k}^{n-1}\prod_{j=k+1}^{i}\alpha_j \leq \frac{1}{1-\alpha_k} - \frac{1}{1-\alpha_n}\prod_{j=k+1}^{n}\alpha_j \leq (1-c+\varepsilon)\sum_{i=k}^{n-1}\prod_{j=k+1}^{i}\alpha_j. \tag{39}$$

Let us now prove that $\lim_{n\to+\infty}\frac{1}{1-\alpha_n}\prod_{j=k+1}^{n}\alpha_j = 0$. If there exists $n_0 \geq k+1$ such that $\alpha_{n_0} = 0$, then the sequence $\left(\frac{1}{1-\alpha_n}\prod_{j=k+1}^{n}\alpha_j\right)_{n\geq k}$ is stationary and equal to 0 for $n \geq n_0$. Without loss of generality, we can assume that $\alpha_n > 0$ for every $n \geq k+1$. For every $i \geq k$, we have

$$\begin{aligned}
\frac{1}{1-\alpha_{i+1}}\prod_{j=k+1}^{i+1}\alpha_j \Bigg/ \left(\frac{1}{1-\alpha_i}\prod_{j=k+1}^{i}\alpha_j\right) &= \frac{1-\alpha_i}{1-\alpha_{i+1}}\alpha_{i+1}\\
&\leq \left(1 + (c+\varepsilon)(1-\alpha_i)\right)\alpha_{i+1} \quad \text{in view of (38)}\\
&\leq \exp((c+\varepsilon)(1-\alpha_i))\exp(\alpha_{i+1}-1)\\
&= \exp(-(1-c-\varepsilon)(1-\alpha_i) + \alpha_{i+1} - \alpha_i).
\end{aligned}$$

By multiplying from $i = k$ to $n-1$, we find

$$\frac{1}{1-\alpha_n}\prod_{j=k+1}^{n}\alpha_j \leq \frac{1}{1-\alpha_k}\exp\left(-(1-c-\varepsilon)\sum_{i=k}^{n-1}(1-\alpha_i) + \alpha_n - \alpha_k\right). \tag{40}$$

By summing the right inequality of (38), we obtain $1/(1-\alpha_i) \leq 1/(1-\alpha_{i_0}) + (c+\varepsilon)(i-i_0)$ for every $i \geq i_0$. Setting $d = 1/(1-\alpha_{i_0})$, we deduce immediately that $1-\alpha_i \geq 1/\big(d + (c+\varepsilon)(i-i_0)\big)$, thus implying that $\sum_{i=1}^{+\infty}(1-\alpha_i) = +\infty$. Since $\varepsilon \in]0, 1-c[$ and since $\alpha_n \in [0, 1[$ for every $n \geq 1$, we infer from inequality (40) that

$$\lim_{n\to+\infty}\frac{1}{1-\alpha_n}\prod_{j=k+1}^{n}\alpha_j = 0.$$

Let us return to (39) and take the limit as $n \to +\infty$. We then obtain for every $k \geq i_0$,

$$(1-c-\varepsilon)\sum_{i=k}^{+\infty}\prod_{j=k+1}^{i}\alpha_j \leq \frac{1}{1-\alpha_k} \leq (1-c+\varepsilon)\sum_{i=k}^{+\infty}\prod_{j=k+1}^{i}\alpha_j,$$

or equivalently,

$$(1-c-\varepsilon)\,t_{k+1} \leq \frac{1}{1-\alpha_k} \leq (1-c+\varepsilon)\,t_{k+1}.$$

This double inequality expresses that $t_{k+1} \sim \frac{1}{(1-c)(1-\alpha_k)}$ as $k \to +\infty$.

Acknowledgments. The authors express their gratitude to the referees for their careful reading of the paper and their valuable suggestions.

REFERENCES

[1] F. Alvarez, *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. Alvarez, 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.

[3] H. Attouch, A. Cabot, *Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity*, J. Differential Equations, 263 (2017), pp. 5412-5458.

[4] H. Attouch, Z. Chbani, J. Peypouquet, P. Redont, *Fast convergence of inertial dynamics and algorithms with asymptotic vanishing damping*, to appear in Math. Program. DOI: 10.1007/s10107-016-0992-8.

[5] 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.

[6] H. Attouch, J. Peypouquet, P. Redont, *A dynamical approach to an inertial forward-backward algorithm for convex minimization*, SIAM J. Optim., 24 (2014), No. 1, pp. 232–256.

[7] 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.

[8] J.-B. Baillon, *Un exemple concernant le comportement asymptotique de la solution du problème* $\frac{du}{dt} + \partial\phi(u) \ni 0$, Journal of Functional Analysis, 28 (1978), pp. 369-376.

[9] H. Bauschke, P. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert spaces*, CMS Books in Mathematics, Springer, (2011).

[10] A. Beck, M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), No. 1, pp. 183–202.

[11] P. Bégout, J. Bolte, M. A. Jendoubi, *On damped second-order gradient systems*, Journal of Differential Equations, 259 (2015), pp. 3115–3143.

[12] R. I. Bot, E. R. Csetnek, *Second order forward-backward dynamical systems for monotone inclusion problems*, SIAM J. Control Optim., 54 (2016), pp. 1423-1443.

[13] H. Brézis, *Opérateurs maximaux monotones dans les espaces de Hilbert et équations d'évolution*, Lecture Notes 5, North Holland, (1972).

[14] H. Brézis, P.L. Lions, *Produits infinis de résolvantes*, Israel Journal of Mathematics, 29 (1978), pp. 329-345.

[15] R.E. Bruck, *Asymptotic convergence of nonlinear contraction semigroups in Hilbert spaces*, J. Funct. Anal., 18 (1975), pp. 15–26.

[16] A. Cabot, H. Engler, S. Gadat, *On the long time behavior of second order differential equations with asymptotically small dissipation*, Transactions of the American Mathematical Society, 361 (2009), pp. 5983–6017.

[17] A. Cabot, H. Engler, S. Gadat, *Second order differential equations with asymptotically small dissipation and piecewise flat potentials*, Electronic Journal of Differential Equations, 17 (2009), pp. 33–38.

[18] A. Cabot, P. Frankel, *Asymptotics for some proximal-like method involving inertia and memory aspects*, Set Valued and Variational Analysis, 19 (2011), pp. 59-74.

[19] 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.

[20] P.L. Combettes, V.R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200.

[21] M. Ghisi, M. Gobbino, A. Haraux, *The remarkable effectiveness of time-dependent damping terms for second order evolution equations*, SIAM J. Control Optim., 54 (2016), No. 3, pp. 1266-1294.

[22] O. Güler, *On the convergence of the proximal point algorithm for convex minimization*, SIAM J. Control Optim. 29 (1991), No. 2, pp. 403-419.

[23] M. A. Jendoubi, R. May, *Asymptotics for a second-order differential equation with nonautonomous damping and an integrable source term*, Applicable Analysis, 94 (2015), pp. 435-443.

[24] D. Kim, J.A. Fessler, *Optimized first-order methods for smooth convex minimization*, Math. Program., 159 (2016), No. 1, pp. 81-107.

[25] 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.

[26] D.A. Lorenz, T. Pock, *An inertial forward-backward algorithm for monotone inclusions*, J. Math. Imaging and Vision, 51 (2015), pp. 311-325.

[27] R. May, *Asymptotic for a second order evolution equation with convex potential and vanishing damping term*, Turk. J. Math., 41 (2017), pp. 681-685.

[28] Y. Nesterov, *A method of solving a convex programming problem with convergence rate O(1/k2)*, Soviet Mathematics Doklady, 27 (1983), pp. 372–376.

[29] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, volume 87 of Applied Optimization.

Kluwer Academic Publishers, Boston, MA, 2004.

[30] B. O'Donoghue, E. Candès, *Adaptive Restart for Accelerated Gradient Schemes*, Foundations of Computational Mathematics, 15 (2015), No. 3, pp. 715–732.

[31] Z. Opial, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. Amer. Math. Soc., 73 (1967), pp. 591–597.

[32] N. Parikh, S. Boyd, *Proximal algorithms*, Foundations and trends in optimization, volume 1, (2013), pp. 123-231.

[33] B.T. Polyak, *Some methods of speeding up the convergence of iteration methods*, USSR Computational Mathematics and Mathematical Physics, 4 (1964), No. 5, pp. 1-17.

[34] B.T. Polyak, *Introduction to optimization.* New York: Optimization Software. (1987).

[35] 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.

[36] S. Villa, S. Salzo, L. Baldassarres, A. Verri, *Accelerated and inexact forward-backward*, SIAM J. Optim., 23 (2013), No. 3, pp. 1607–1633.