Applying FISTA to optimization problems (with or) without minimizers

Heinz H. Bauschke[*], Minh N. Bui[†], and Xianfu Wang[‡]

July 2, 2019

Abstract

Beck and Teboulle's FISTA method for finding a minimizer of the sum of two convex functions, one of which has a Lipschitz continuous gradient whereas the other may be nonsmooth, is arguably the most important optimization algorithm of the past decade. While research activity on FISTA has exploded ever since, the mathematically challenging case when the original optimization problem has no minimizer has found only limited attention.

In this work, we systematically study FISTA and its variants. We present general results that are applicable, regardless of the existence of minimizers.

2010 Mathematics Subject Classification: Primary 90C25, 65K05; Secondary 49M27

Keywords: convex function, FISTA, forward-backward method, Nesterov acceleration, proximal gradient algorithm

1 Introduction

We assume that

$$\mathcal{H} \text{ is a real Hilbert space} \tag{1}$$

with inner product $\langle \cdot \mid \cdot \rangle$ and associated norm $\|\cdot\|$. We also presuppose throughout the paper that

$$f\colon \mathcal{H} \to \mathbb{R} \quad \text{and} \quad g\colon \mathcal{H} \to \left]-\infty, +\infty\right] \tag{2}$$

satisfy the following:

Assumption 1.1

(A1) f is convex and Fréchet differentiable on $\mathcal{H}$, and ∇f is β-Lipschitz continuous with $\beta \in \left]0, +\infty\right[$;

(A2) g is convex, lower semicontinuous, and proper;

(A3) $\gamma \in \left]0, 1/\beta\right]$ is a parameter.

One fundamental problem in optimization is to

$$\text{minimize } f + g \text{ over } \mathcal{H}. \tag{3}$$

For convenience, we set

$$h := f + g \quad \text{and} \quad T := \operatorname{Prox}_{\gamma g} \circ \left(\operatorname{Id} - \gamma \nabla f\right), \tag{4}$$

[*]Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. Email: heinz.bauschke@ubc.ca.
[†]Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205, USA. Email: mnbui@ncsu.edu.
[‡]Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. Email: shawn.wang@ubc.ca.

and where we follow standard notation in convex analysis (as employed, e.g., in [8]). Then many algorithms designed for solving (3) employ the forward-backward or proximal gradient operator T in some fashion. Since the advent of Nesterov's acceleration [22] (when $g \equiv 0$) and Beck and Teboulle's fast proximal gradient method FISTA [11] (see also [9, Chapter 10]), the literature on algorithms for solving (3) has literally exploded; see, e.g., [22, 11, 7, 1, 3, 5, 18, 2] for a selection of key contributions. Indeed, out of nearly one million mathematical publications that appeared since 2009 and are indexed by *Mathematical Reviews*, the 2009-FISTA paper [11] by Beck and Teboulle takes the *number two spot!* (In passing, we note that it has been cited more than 6,000 times on Google Scholar where it now receives about *3 new citations every day!*) The overwhelming majority of these papers assume that the problem (3) has a solution to start with. Complementing and contributing to these analyses, we follow a path less trodden:

The aim of this paper is to study the behaviour of the fast proximal gradient methods (and monotone variants), in the case when the original problem (3) *does not necessarily have a solution.*

Before we turn to our main results, let us state the FISTA or fast proximal gradient method:

Algorithm 1.2 (FISTA) Let $x_0 \in \mathcal{H}$, set $y_1 := x_0$, and update

$$\begin{aligned} &\text{for } n = 1, 2, \ldots \\ &\left\lfloor \begin{array}{ll} x_n & := Ty_n, \\ y_{n+1} & := x_n + \dfrac{\tau_n - 1}{\tau_{n+1}}(x_n - x_{n-1}), \end{array} \right. \end{aligned} \tag{5}$$

where T is defined in (4), $\mathbb{N}^* := \{1, 2, \ldots\}$, and $(\tau_n)_{n \in \mathbb{N}^*}$ is a sequence of real numbers in $[1, +\infty[$.

Note that when $\tau_n \equiv 1$, one obtains the classical (unaccelerated) proximal gradient method. There are two very popular choices for the sequence $(\tau_n)_{n \in \mathbb{N}^*}$ to achieve acceleration. Firstly, given $\tau_1 := 1$, the classical FISTA [11, 10, 16, 22] update is

$$(\forall n \in \mathbb{N}^*) \quad \tau_{n+1} := \frac{1 + \sqrt{1 + 4\tau_n^2}}{2}. \tag{6}$$

The second update has the explicit formula

$$(\forall n \in \mathbb{N}^*) \quad \tau_n := \frac{n + \rho - 1}{\rho}, \tag{7}$$

where $\rho \in [2, +\infty[$; see, e.g., [3, 5, 15, 27].

Convergence results of the sequence generated by FISTA under a suitable tuning of $(\tau_n)_{n \in \mathbb{N}^*}$ can be found in [5, 1, 15]. The relaxed case was considered in [7] and error-tolerant versions were considered in [3, 2]. In addition, for results concerning the rate of convergence of function values, see [11, 10, 27, 26]. The authors of [16] established a variant of FISTA that covers the strongly convex case. An alternative of the classical proximal gradient algorithm with relaxation and error is presented in [19] (see also [8, 13, 26]). Finally, a new forward-backward splitting scheme (for finding a zero of a sum of two maximally monotone operators) that includes FISTA as a special case was proposed in [18].

The main difference between our work and existing work is that we focus on the minimizing property of the sequences generated by FISTA and MFISTA in the general framework, i.e., when the set $\operatorname{Argmin}(f + g)$ *is possibly empty.* Let us now list our **main results**:

- Theorem 5.3 establishes the behavior of FISTA in the possibly inconsistent case; moreover, our assumption on $(\tau_n)_{n \in \mathbb{N}^*}$ (see (39)) is very mild.
- Theorem 5.5 concerns FISTA when $(\tau_n)_{n \in \mathbb{N}^*}$ behaves similarly to the Beck–Teboulle choice.
- Theorem 5.10 deals with the case when $(\tau_n)_{n \in \mathbb{N}^*}$ is bounded; see, in particular, (ii)(a) and (v)(b).
- Theorem 6.1 considers MFISTA [10], the monotone version of FISTA, when Assumption 4.1 is in force and $(\tau_n)_{n \in \mathbb{N}^*}$ is unbounded.

To the best of our knowledge, Theorem 5.3 is new. The proof of Theorem 5.5, which can be viewed as a "discrete version" of [3, Theorem 2.3], relies on techniques seen in [3, Theorem 2.3] and [1, Proposition 3]; items (ii)–(vi) are new. A result similar to Theorem 5.5(ii) was mentioned in [6, Theorem 4.1]. However, no proof was given, and the parameter sequence there is a special case of the one considered in Theorem 5.5. Items (vii)(a) and (vii)(b) is a slight modification of [4, Proposition 4.3]. Concerning Theorem 5.10, items (i)–(iv) and (v)(b) are new while (v)(a) was proven in [1, Corollary 20(iii)]. Item (i) in the classical case ($\tau_n \equiv 1$) relates to [12, Theorem 4.2] where linesearches were employed. In Theorem 6.1, items (i)–(v) are new. Compared to [10, Theorem 5.1], we allow many possible choices for the parameter sequence in Theorem 6.1(vi); see, e.g., Examples 4.4–4.6. In addition, by adapting the technique of [1, Theorem 9], we improve the convergence rate of MFISTA under the condition (110) in Theorem 6.1.

There are also several **minor results** worth emphasizing: Lemma 2.4 is new. The notion of quasi-Fejér monotonicity is revisited in Lemma 2.7; however, our error sequence need not be positive. The assumptions in Lemma 3.2 and Lemma 3.3 are somewhat minimal, which allow us to establish the minimizing property of FISTA and MFISTA in the case where there are possibly no minimizers in Sections 5 and 6. Example 4.5 is new. Proposition 5.12 describes the behaviour of $(x_n - x_{n-1})_{n\in\mathbb{N}^*}$ in the classical proximal gradient (ISTA) case while Corollary 5.15 provides a sufficient condition for strong convergence of $(x_n)_{n\in\mathbb{N}^*}$ in this case. The new Proposition 5.14 presents some progress towards the still open question regarding the convergence of $(x_n)_{n\in\mathbb{N}^*}$ generated by classical FISTA. The weak convergence part in Corollary 5.15 was considered in [4]; however, our new Fejérian approach allows us to obtain strong convergence when $\operatorname{int}(\operatorname{Argmin} h) \neq \varnothing$.

Let us now turn to the organization of this paper. Classical results on real sequences and new results on the Fejér monotonicity are recorded in Section 2. The "one step" behaviour of both FISTA and MFISTA is carefully examined in Section 3. In Section 4, we investigate properties of the parameter sequence $(\tau_n)_{n\in\mathbb{N}^*}$. Our main results on FISTA and MFISTA are presented in Sections 5 and 6 respectively. The concluding Section 7 contains a discussion of open problems.

A final note on notation is in order. For a sequence $(\xi_n)_{n\in\mathbb{N}^*}$ and an extended real number $\xi \in [-\infty, +\infty]$, the notation $\xi_n \uparrow \xi$ means that $(\xi_n)_{n\in\mathbb{N}^*}$ is increasing (i.e., $\xi_n \leqslant \xi_{n+1}$) and $\xi_n \to \xi$ as $n \to +\infty$. Likewise, $\xi_n \downarrow \xi$ means that $(\xi_n)_{n\in\mathbb{N}^*}$ is decreasing (i.e., $\xi_n \geqslant \xi_{n+1}$) and $\xi_n \to \xi$ as $n \to +\infty$. For any other notation not defined, we refer the reader to [8].

2 Auxiliary results

In this section, we collect results on sequences which will make the proofs in later sections more structured.

Lemma 2.1 *Let $(\tau_n)_{n\in\mathbb{N}^*}$ be an increasing sequence in $[1, +\infty[$ such that $\lim \tau_n = +\infty$. Then*

$$\sum_{n\in\mathbb{N}^*} \left(1 - \left(\frac{\tau_n - 1}{\tau_{n+1}}\right)^2\right) = \sum_{n\in\mathbb{N}^*} \left(1 - \frac{\tau_n^2}{\tau_{n+1}^2}\right) = +\infty. \tag{8}$$

Proof. See Appendix A. ■

Lemma 2.2 *Let $(\alpha_n)_{n\in\mathbb{N}^*}$ and $(\beta_n)_{n\in\mathbb{N}^*}$ be sequences in $\mathbb{R}_+$. Suppose that $\sum_{n\in\mathbb{N}^*} \alpha_n = +\infty$ and that $\sum_{n\in\mathbb{N}^*} \alpha_n\beta_n < +\infty$. Then $\underline{\lim}\, \beta_n = 0$.*

Proof. See Appendix B. ■

The novelty of the following result lies in the fact that the error sequence $(\varepsilon_n)_{n\in\mathbb{N}^*}$ need not lie in $\mathbb{R}_+$.

Lemma 2.3 *Let $(\alpha_n)_{n\in\mathbb{N}^*}$ be a sequence in $\mathbb{R}$, let $(\beta_n)_{n\in\mathbb{N}^*}$ be a sequence in $\mathbb{R}_+$, and let $(\varepsilon_n)_{n\in\mathbb{N}^*}$ be a sequence in $\mathbb{R}$. Suppose that $(\alpha_n)_{n\in\mathbb{N}^*}$ is bounded below, that*

$$(\forall n \in \mathbb{N}^*) \quad \alpha_{n+1} \leqslant \alpha_n - \beta_n + \varepsilon_n, \tag{9}$$

and that the series $\sum_{n\in\mathbb{N}^*} \varepsilon_n$ converges in $\mathbb{R}$. Then the following hold:

(i) *$(\alpha_n)_{n\in\mathbb{N}^*}$ is convergent in $\mathbb{R}$.*

(ii) *$\sum_{n\in\mathbb{N}^*} \beta_n < +\infty$.*

Proof. See Appendix C. ■

Lemma 2.4 *Let $(\alpha_n)_{n\in\mathbb{N}^*}$ be a sequence of real numbers. Consider the following statements:*

(i) *$(n\alpha_n)_{n\in\mathbb{N}^*}$ converges in $\mathbb{R}$.*

(ii) *$\sum_{n\in\mathbb{N}^*} \alpha_n$ converges in $\mathbb{R}$.*

(iii) *$\sum_{n\in\mathbb{N}^*} n(\alpha_n - \alpha_{n+1})$ converges in $\mathbb{R}$.*

Suppose that two of the statements (i)–(iii) hold. Then the remaining one also holds.

Proof. See Appendix D. ■

The following result is stated in [25, Problem 2.6.19]; we provide a proof in Appendix E for completeness.

Lemma 2.5 *Let $(\alpha_n)_{n\in\mathbb{N}^*}$ be a decreasing sequence in $\mathbb{R}_+$. Then*

$$\sum_{n\in\mathbb{N}^*} \alpha_n < +\infty \quad \Leftrightarrow \quad \Big[n\alpha_n \to 0 \text{ as } n \to +\infty \text{ and } \sum_{n\in\mathbb{N}^*} n(\alpha_n - \alpha_{n+1}) < +\infty \Big]. \tag{10}$$

The following variant of Opial's lemma [23] will be required in the sequel.

Lemma 2.6 *Let C be a nonempty subset of $\mathcal{H}$, and let $(u_n)_{n\in\mathbb{N}^*}$ and $(v_n)_{n\in\mathbb{N}^*}$ be sequences in $\mathcal{H}$. Suppose that $u_n - v_n \to 0$, that every weak sequential cluster point of $(v_n)_{n\in\mathbb{N}^*}$ lies in C, and that, for every $c \in C$, $(\|u_n - c\|)_{n\in\mathbb{N}^*}$ converges. Then there exists $w \in C$ such that $u_n \rightharpoonup w$ and $v_n \rightharpoonup w$.*

Proof. For every $c \in C$, since $u_n - v_n \to 0$ and $(\|u_n - c\|)_{n\in\mathbb{N}^*}$ converges, we deduce that $(\|v_n - c\|)_{n\in\mathbb{N}^*}$ converges. In turn, because every weak sequential cluster point of $(v_n)_{n\in\mathbb{N}^*}$ belongs to C, [8, Lemma 2.47] yields the existence of $w \in C$ satisfying $v_n \rightharpoonup w$. Therefore, because $u_n - v_n \to 0$, we conclude that $(u_n)_{n\in\mathbb{N}^*}$ and $(v_n)_{n\in\mathbb{N}^*}$ converge weakly to w. ■

We next revisit the notion of quasi-Fejér monotonicity in the Hilbert spaces setting studied in [17]. This plays a crucial role in our analysis of Proposition 5.14. Nevertheless, to fit our framework of Proposition 5.14, the error sequence $(\varepsilon_n)_{n\in\mathbb{N}^*}$ is not required to be positive in Lemma 2.7. The proof is based on [17, Proposition 3.3(iii) and Proposition 3.10].

Lemma 2.7 *Let C be a nonempty subset of $\mathcal{H}$, let $(u_n)_{n\in\mathbb{N}^*}$ be a sequence in $\mathcal{H}$, and let $(\varepsilon_n)_{n\in\mathbb{N}^*}$ be a sequence in $\mathbb{R}$. Suppose that*

$$(\forall c \in C)(\forall n \in \mathbb{N}^*) \quad \|u_{n+1} - c\|^2 \leqslant \|u_n - c\|^2 + \varepsilon_n, \tag{11}$$

and that $\sum_{n\in\mathbb{N}^*} \varepsilon_n$ converges in $\mathbb{R}$. Then the following hold:

(i) *For every $c \in C$, the sequence $(\|u_n - c\|)_{n\in\mathbb{N}^*}$ converges in $\mathbb{R}$.*

(ii) *Suppose that $\operatorname{int} C \neq \varnothing$. Then $(u_n)_{n\in\mathbb{N}^*}$ converges strongly in $\mathcal{H}$.*

Proof. (i): This is a direct consequence of Lemma 2.3(i).

(ii): We follow along the lines of [17, Proposition 3.10]. Let $v \in \operatorname{int} C$ and $\rho \in]0, +\infty[$ be such that $\mathrm{B}(v;\rho) := \{x \in \mathcal{H} \mid \|x - v\| \leqslant \rho\} \subseteq C$. Define a sequence $(v_n)_{n\in\mathbb{N}^*}$ in C via

$$(\forall n \in \mathbb{N}^*) \quad v_n := \begin{cases} v, & \text{if } u_{n+1} = u_n; \\ v - \rho \dfrac{u_{n+1} - u_n}{\|u_{n+1} - u_n\|}, & \text{otherwise.} \end{cases} \tag{12}$$

We now verify that

$$(\forall n \in \mathbb{N}^*) \quad \|u_{n+1} - v\|^2 \leqslant \|u_n - v\|^2 - 2\rho\|u_{n+1} - u_n\| + \varepsilon_n. \tag{13}$$

Fix $n \in \mathbb{N}^*$. If $u_{n+1} = u_n$, then (11) implies that $\varepsilon_n \geqslant 0$, and therefore (13) holds. Otherwise, because $v_n \in C$, (11) yields $\|u_{n+1} - v_n\|^2 \leqslant \|u_n - v_n\|^2 + \varepsilon_n$. In turn, using (12), we obtain

$$\left\|(u_{n+1} - v) + \rho\frac{u_{n+1} - u_n}{\|u_{n+1} - u_n\|}\right\|^2 \leqslant \left\|(u_n - v) + \rho\frac{u_{n+1} - u_n}{\|u_{n+1} - u_n\|}\right\|^2 + \varepsilon_n, \tag{14}$$

and after expanding both sides and simplifying terms, we get (13). Consequently, owing to (13) and the convergence of $\sum_{n\in\mathbb{N}^*} \varepsilon_n$, we derive from Lemma 2.3(ii) that $\sum_{n\in\mathbb{N}^*} 2\rho\|u_{n+1} - u_n\| < +\infty$. Hence, by completeness of $\mathcal{H}$, $(u_n)_{n\in\mathbb{N}^*}$ converges strongly to a point in $\mathcal{H}$. ■

We conclude this section with a simple identity. If x, y, and z are in $\mathcal{H}$, then

$$\|x - y\|^2 + 2\langle x - y \mid z - x\rangle = \|z - y\|^2 - \|z - x\|^2. \tag{15}$$

3 One-step results

The aim of this section is to present several results on performing just *one step* of FISTA or MFISTA. This allows us to present subsequent convergence results more clearly. Recall that Assumption 1.1 is in force and (see (4)) that

$$h = f + g \quad \text{and} \quad T = \operatorname{Prox}_{\gamma g} \circ \big(\mathrm{Id} - \gamma\nabla f\big)\,. \tag{16}$$

Clearly,

$$\operatorname{ran} T = \operatorname{ran}\big(\operatorname{Prox}_{\gamma g} \circ (\mathrm{Id} - \gamma\nabla f)\big) \subseteq \operatorname{dom}\partial g \subseteq \operatorname{dom} g = \operatorname{dom} h. \tag{17}$$

Lemma 3.1 (Beck–Teboulle) *The following holds:*

$$(\forall (x,y) \in \mathcal{H} \times \mathcal{H}) \quad \gamma^{-1}\langle y - Ty \mid x - y\rangle + (2\gamma)^{-1}\|y - Ty\|^2 \leqslant h(x) - h(Ty). \tag{18}$$

Proof. See Appendix F. ■

Lemma 3.2 (one FISTA step) *Let $(y, x_-) \in \mathcal{H} \times \mathcal{H}$, let τ and τ_+ be in $[1, +\infty[$, and set*

$$x := Ty, \quad y_+ := x + \frac{\tau - 1}{\tau_+}(x - x_-), \quad \textit{and } x_+ := Ty_+. \tag{19}$$

In addition, let $z \in \operatorname{dom} h$, and set

$$\begin{cases} u & := \tau x - (\tau - 1)x_- - z, \\ u_+ & := \tau_+ x_+ - (\tau_+ - 1)x - z, \\ \mu & := h(x) - h(z), \\ \mu_+ & := h(x_+) - h(z). \end{cases} \tag{20}$$

Then the following hold:

(i) $h(x_+) + (2\gamma)^{-1}\|x_+ - x\|^2 \leqslant h(x) + (\tau-1)^2(2\gamma)^{-1}\|x - x_-\|^2/\tau_+^2$.

(ii) $\tau_+^2\mu_+ + (2\gamma)^{-1}\|u_+\|^2 \leqslant \tau_+(\tau_+ - 1)\mu + (2\gamma)^{-1}\|u\|^2$.

(iii) *Suppose that* $\tau \leqslant \tau_+$, *that* $\tau_+(\tau_+ - 1) \leqslant \tau^2$, *and that* $\inf h > -\infty$. *Then*

$$\tau_+^2\mu_+ + (2\gamma)^{-1}\|u_+\|^2 \leqslant \tau^2\mu + (2\gamma)^{-1}\|u\|^2 + \tau_+(h(z) - \inf h). \tag{21}$$

Proof. First, since $z \in \operatorname{dom} h$, we get from (17)&(19)&(20) that $\mu \in \mathbb{R}$ and $\mu_+ \in \mathbb{R}$. Next, because $x_+ = Ty_+$, we derive from (18) (applied to (x, y_+)) that

$$\mu - \mu_+ = h(x) - h(x_+) \geqslant \gamma^{-1}\langle y_+ - x_+ \mid x - y_+\rangle + (2\gamma)^{-1}\|y_+ - x_+\|^2. \tag{22}$$

(i): We derive from (22), (15), and (19) that

$$h(x) - h(x_+) \geqslant (2\gamma)^{-1}\big(\|x - x_+\|^2 - \|y_+ - x_+\|^2 - \|x - y_+\|^2\big) + (2\gamma)^{-1}\|y_+ - x_+\|^2 \tag{23a}$$

$$= (2\gamma)^{-1}\left(\|x - x_+\|^2 - \left(\frac{\tau-1}{\tau_+}\right)^2\|x - x_-\|^2\right), \tag{23b}$$

and thus, since $h(x_+) \in \mathbb{R}$, the conclusion follows.

(ii): Since $x_+ = Ty_+$, applying (18) to (z, y_+) gives

$$-\mu_+ = h(z) - h(x_+) \geqslant \gamma^{-1}\langle y_+ - x_+ \mid z - y_+\rangle + (2\gamma)^{-1}\|y_+ - x_+\|^2. \tag{24}$$

Therefore, because $\tau_+ - 1 \geqslant 0$ by assumption, it follows from (22) and (24) that

$$(\tau_+ - 1)\mu - \tau_+\mu_+ = (\tau_+ - 1)(\mu - \mu_+) + (-\mu_+) \tag{25a}$$

$$\geqslant \gamma^{-1}\langle y_+ - x_+ \mid (\tau_+ - 1)(x - y_+) + (z - y_+)\rangle + (2\gamma)^{-1}\tau_+\|y_+ - x_+\|^2 \tag{25b}$$

$$= \gamma^{-1}\langle y_+ - x_+ \mid (\tau_+ - 1)x - \tau_+y_+ + z\rangle + (2\gamma)^{-1}\tau_+\|y_+ - x_+\|^2. \tag{25c}$$

In turn, on the one hand, multiplying both sides of (25) by $\tau_+ > 0$, we infer from (15) (applied to $(\tau_+y_+, \tau_+x_+, (\tau_+ - 1)x + z)$) and the very definition of u_+ that

$$\tau_+(\tau_+ - 1)\mu - \tau_+^2\mu_+ \geqslant \gamma^{-1}\langle \tau_+y_+ - \tau_+x_+ \mid (\tau_+ - 1)x + z - \tau_+y_+\rangle + (2\gamma)^{-1}\|\tau_+(y_+ - x_+)\|^2 \tag{26a}$$

$$= (2\gamma)^{-1}\big(\|(\tau_+ - 1)x + z - \tau_+x_+\|^2 - \|(\tau_+ - 1)x + z - \tau_+y_+\|^2\big) \tag{26b}$$

$$= (2\gamma)^{-1}\big(\|u_+\|^2 - \|(\tau_+ - 1)x + z - \tau_+y_+\|^2\big). \tag{26c}$$

On the other hand, since $\tau_+y_+ = \tau_+x + (\tau - 1)(x - x_-)$ due to (19), the definition of u yields

$$\tau_+y_+ - (\tau_+ - 1)x - z = \tau_+x + (\tau - 1)(x - x_-) - (\tau_+ - 1)x - z = \tau x - (\tau - 1)x_- - z = u. \tag{27}$$

Altogether, $(2\gamma)^{-1}(\|u_+\|^2 - \|u\|^2) \leqslant \tau_+(\tau_+ - 1)\mu - \tau_+^2\mu_+$, which implies the desired conclusion.

(iii): Since $\mu = h(x) - h(z) \geqslant \inf h - h(z) > -\infty$ and, by assumption, $\tau_+^2 - \tau_+ - \tau^2 \leqslant 0$, we deduce that $(\tau_+^2 - \tau_+ - \tau^2)\mu \leqslant (\tau_+^2 - \tau_+ - \tau^2)(\inf h - h(z)) = (\tau^2 + \tau_+ - \tau_+^2)(h(z) - \inf h)$. Hence, because $0 < \tau \leqslant \tau_+$ and $h(z) - \inf h \geqslant 0$, it follows that $(\tau_+^2 - \tau_+ - \tau^2)\mu \leqslant (\tau^2 + \tau_+ - \tau_+^2)(h(z) - \inf h) \leqslant \tau_+(h(z) - \inf h)$. Consequently, (ii) implies that

$$\tau_+^2\mu_+ + (2\gamma)^{-1}\|u_+\|^2 \leqslant \tau_+(\tau_+ - 1)\mu + (2\gamma)^{-1}\|u\|^2 \tag{28a}$$

$$= \tau^2\mu + (2\gamma)^{-1}\|u\|^2 + (\tau_+^2 - \tau_+ - \tau^2)\mu \tag{28b}$$

$$\leqslant \tau^2\mu + (2\gamma)^{-1}\|u\|^2 + \tau_+(h(z) - \inf h), \tag{28c}$$

as required. ■

The analysis of the following lemma follows the lines of [10, Theorem 5.1].

Lemma 3.3 (one MFISTA step) *Let $(y, x_-) \in \mathcal{H} \times \mathcal{H}$, let τ and τ_+ be in $[1, +\infty[$, and set*

$$\begin{cases} z & := Ty, \\ x & := \begin{cases} x_-, & \text{if } h(x_-) \leqslant h(z); \\ z, & \text{otherwise}, \end{cases} \\ y_+ & := x + \dfrac{\tau}{\tau_+}(z - x) + \dfrac{\tau - 1}{\tau_+}(x - x_-), \\ z_+ & := Ty_+, \\ x_+ & := \begin{cases} x, & \text{if } h(x) \leqslant h(z_+); \\ z_+, & \text{otherwise}. \end{cases} \end{cases} \tag{29}$$

Furthermore, let $w \in \operatorname{dom} h$, and define

$$\begin{cases} u & := \tau z - (\tau - 1)x_- - w, \\ u_+ & := \tau_+ z_+ - (\tau_+ - 1)x - w, \\ \mu & := h(x) - h(w), \\ \mu_+ & := h(x_+) - h(w). \end{cases} \tag{30}$$

Then the following hold:

(i) $h(x_+) + (2\gamma)^{-1}\|z_+ - x\|^2 \leqslant h(x) + (2\gamma)^{-1}\tau^2\|z - x_-\|^2/\tau_+^2$.

(ii) $\tau_+^2\mu_+ + (2\gamma)^{-1}\|u_+\|^2 \leqslant \tau_+(\tau_+ - 1)\mu + (2\gamma)^{-1}\|u\|^2$.

Proof. First, since $z_+ = Ty_+$, using (18) with (x, y_+) and (15) with (y_+, z_+, x) yields

$$h(x) - h(z_+) = h(x) - h(Ty_+) \geqslant \gamma^{-1}\langle y_+ - z_+ \mid x - y_+\rangle + (2\gamma)^{-1}\|y_+ - z_+\|^2. \tag{31a}$$

$$= (2\gamma)^{-1}\big(\|x - z_+\|^2 - \|x - y_+\|^2\big). \tag{31b}$$

(i): On the one hand, by the very definition of x_+ and (31), $h(x) - h(x_+) \geqslant h(x) - h(z_+) \geqslant (2\gamma)^{-1}(\|x - z_+\| - \|x - y_+\|^2)$, and thus,

$$h(x_+) + (2\gamma)^{-1}\|x - z_+\|^2 \leqslant h(x) + (2\gamma)^{-1}\|x - y_+\|^2. \tag{32}$$

On the other hand, due to (29),

$$y_+ - x = \frac{\tau}{\tau_+}(z - x) + \frac{\tau - 1}{\tau_+}(x - x_-) \tag{33a}$$

$$= \begin{cases} \dfrac{\tau}{\tau_+}(z - x_-) + \dfrac{\tau - 1}{\tau_+}(x_- - x_-), & \text{if } h(x_-) \leqslant h(z); \\ \dfrac{\tau}{\tau_+}(z - z) + \dfrac{\tau - 1}{\tau_+}(z - x_-), & \text{otherwise} \end{cases} \tag{33b}$$

$$= \begin{cases} \dfrac{\tau}{\tau_+}(z - x_-), & \text{if } h(x_-) \leqslant h(z); \\ \dfrac{\tau - 1}{\tau_+}(z - x_-), & \text{otherwise}, \end{cases} \tag{33c}$$

and since $\tau \geqslant 1$, it follows that

$$\|y_+ - x\| \leqslant \frac{\tau}{\tau_+}\|z - x_-\|. \tag{34}$$

Altogether, (32) and (34) yield the desired result.

(ii): Applying (18) to the pair (w, y_+) and noticing that $z_+ = Ty_+$, we get

$$h(w) - h(z_+) \geqslant \gamma^{-1}\langle y_+ - z_+ \mid w - y_+\rangle + (2\gamma)^{-1}\|y_+ - z_+\|^2. \tag{35}$$

In turn, since $\tau_+ \geqslant 1$, the very definition of x_+, (31), and (35) imply that

$$\begin{aligned}
(\tau_+ - 1)\mu - \tau_+\mu_+ &= (\tau_+ - 1)(h(x) - h(w)) - \tau_+(h(x_+) - h(w)) && \text{(36a)}\\
&= (\tau_+ - 1)h(x) + h(w) - \tau_+ h(x_+) && \text{(36b)}\\
&\geqslant (\tau_+ - 1)h(x) + h(w) - \tau_+ h(z_+) && \text{(36c)}\\
&= (\tau_+ - 1)(h(x) - h(z_+)) + h(w) - h(z_+) && \text{(36d)}\\
&\geqslant (2\gamma)^{-1}\tau_+\|y_+ - z_+\|^2 + \gamma^{-1}\langle y_+ - z_+ \mid (\tau_+ - 1)(x - y_+) + w - y_+\rangle && \text{(36e)}\\
&= (2\gamma)^{-1}\tau_+\|y_+ - z_+\|^2 + \gamma^{-1}\langle y_+ - z_+ \mid w + (\tau_+ - 1)x - \tau_+ y_+\rangle. && \text{(36f)}
\end{aligned}$$

Thus, since $\tau_+ > 0$, it follows from (15) (applied to $(\tau_+ y_+, \tau_+ z_+, w + (\tau_+ - 1)x)$) that

$$\begin{aligned}
\tau_+(\tau_+ - 1)\mu - \tau_+^2\mu_+ &\geqslant (2\gamma)^{-1}\tau_+^2\|y_+ - z_+\|^2 + \gamma^{-1}\langle \tau_+ y_+ - \tau_+ z_+ \mid w + (\tau_+ - 1)x - \tau_+ y_+\rangle && \text{(37a)}\\
&= (2\gamma)^{-1}\big(\|\tau_+ z_+ - (\tau_+ - 1)x - w\|^2 - \|\tau_+ y_+ - (\tau_+ - 1)x - w\|^2\big). && \text{(37b)}
\end{aligned}$$

Furthermore, by the definition of y_+, we have $\tau_+ y_+ = \tau_+ x + \tau(z - x) + (\tau - 1)(x - x_-) = (\tau_+ - 1)x + \tau z - (\tau - 1)x_-$, which asserts that $\tau_+ y_+ - (\tau_+ - 1)x = \tau z - (\tau - 1)x_-$. Combining this and (37) entails that

$$\tau_+(\tau_+ - 1)\mu - \tau_+^2\mu_+ \geqslant (2\gamma)^{-1}\big(\|\tau_+ z_+ - (\tau_+ - 1)x - w\|^2 - \|\tau z - (\tau - 1)x_- - w\|^2\big), \tag{38}$$

which completes the proof. ■

4 The parameter sequence

A central ingredient of FISTA and MFISTA is the parameter sequence $(\tau_n)_{n\in\mathbb{N}^*}$. In this section, we present various properties of the parameter sequence as well as examples. From this point onwards, we will assume the following:

Assumption 4.1 We assume that $(\tau_n)_{n\in\mathbb{N}^*}$ is a sequence of real numbers such that

$$\tau_1 \in [1, +\infty[, \quad (\forall n \in \mathbb{N}^*)\ \tau_{n+1} \in \left[\tau_n, \frac{1 + \sqrt{1 + 4\tau_n^2}}{2}\right], \quad \text{and} \quad \tau_\infty := \sup_{k\in\mathbb{N}^*} \tau_k. \tag{39}$$

Remark 4.2 A few observations regarding Assumption 4.1 are in order.

(i) It is clear from (39) that

$$(\forall n \in \mathbb{N}^*) \quad \tau_n \geqslant 1. \tag{40}$$

(ii) Because $(\tau_n)_{n\in\mathbb{N}^*}$ is increasing,

$$\tau_n \uparrow \tau_\infty \overset{(40)}{\in} [1, +\infty]. \tag{41}$$

(iii) Due to (40) and the assumption that $(\forall n \in \mathbb{N}^*)\ \tau_{n+1} \leqslant \big(1 + \sqrt{1 + 4\tau_n^2}\big)/2$, it is straightforward to verify that

$$(\forall n \in \mathbb{N}^*) \quad \tau_{n+1}^2 - \tau_{n+1} \leqslant \tau_n^2. \tag{42}$$

(iv) For every $n \in \mathbb{N}^*$, since $\tau_n \leqslant \tau_{n+1} \leqslant (1 + \sqrt{1 + 4\tau_n^2})/2$ by (39), it follows from (42) and (40) that

$$\tau_{n+1} - \tau_n = \frac{\tau_{n+1}^2 - \tau_n^2}{\tau_{n+1} + \tau_n} \leqslant \frac{\tau_{n+1}}{\tau_{n+1} + \tau_n} \leqslant \frac{1 + \sqrt{1 + 4\tau_n^2}}{2(\tau_n + \tau_n)} \leqslant \frac{\tau_n + \sqrt{\tau_n^2 + 4\tau_n^2}}{4\tau_n} = \frac{1 + \sqrt{5}}{4} < 0.81. \tag{43}$$

Lemma 4.3 *The following hold:*

(i) $\overline{\lim}(\tau_n/n) \leqslant \tau_1/2$.

(ii) *Using the convention that* $\frac{1}{+\infty} = 0$, *we have*

$$\frac{1-1/\tau_\infty}{1+1/\tau_\infty} - \frac{1}{\tau_\infty(\tau_\infty+1)} \leqslant \underline{\lim} \frac{\tau_n - 1}{\tau_{n+1}} \leqslant \overline{\lim} \frac{\tau_n - 1}{\tau_{n+1}} \leqslant 1 - \frac{1}{\tau_\infty}. \tag{44}$$

(iii) *Suppose that* $\lim \tau_n = +\infty$. *Then*

$$\lim \frac{\tau_n - 1}{\tau_{n+1}} = 1. \tag{45}$$

Proof. (i): We claim that $(\forall n \in \mathbb{N}^*)\ \tau_n \leqslant \tau_1(n+\sqrt{n})/2$. The inequality is clear when $n = 1$. Assume that, for some integer $n \geqslant 1$, we have $\tau_n \leqslant \tau_1(n+\sqrt{n})/2$. Then, on the one hand, we derive from (39) that

$$\tau_{n+1} \leqslant \frac{1+\sqrt{1+4\tau_n^2}}{2} \leqslant \frac{\tau_1 + \sqrt{\tau_1^2 + \tau_1^2(n+\sqrt{n})^2}}{2} = \frac{\tau_1\big(1+\sqrt{1+(n+\sqrt{n})^2}\big)}{2}. \tag{46}$$

On the other hand, since $(n+\sqrt{n+1})^2 - (1+(n+\sqrt{n})^2) = 2n(\sqrt{n+1}-\sqrt{n}) > 0$, we obtain $\sqrt{1+(n+\sqrt{n})^2} < n+\sqrt{n+1}$. Altogether, $\tau_{n+1} < \tau_1(n+1+\sqrt{n+1})/2$, which concludes the induction argument. Consequently, $\overline{\lim}(\tau_n/n) \leqslant \lim \tau_1(n+\sqrt{n})/(2n) = \tau_1/2$.

(ii): First, since $(\forall n \in \mathbb{N}^*)\ (\tau_n - 1)/\tau_{n+1} \leqslant (\tau_{n+1}-1)/\tau_{n+1} = 1 - 1/\tau_{n+1}$ by (39), we infer from (41) that $\overline{\lim}(\tau_n - 1)/\tau_{n+1} \leqslant 1 - 1/\tau_\infty$. Next, by (42) and (39), we have

$$(\forall n \in \mathbb{N}^*) \quad \frac{\tau_n - 1}{\tau_{n+1}} = \frac{\tau_n^2 - 1}{\tau_{n+1}(\tau_n+1)} \geqslant \frac{\tau_{n+1}^2 - \tau_{n+1} - 1}{\tau_{n+1}(\tau_n+1)} \tag{47a}$$

$$= \frac{\tau_{n+1}-1}{\tau_n+1} - \frac{1}{\tau_{n+1}(\tau_n+1)} \tag{47b}$$

$$\geqslant \frac{\tau_n-1}{\tau_n+1} - \frac{1}{\tau_{n+1}(\tau_n+1)} \tag{47c}$$

$$= \frac{1-1/\tau_n}{1+1/\tau_n} - \frac{1}{\tau_{n+1}(\tau_n+1)}, \tag{47d}$$

and hence, we get from (41) that $\underline{\lim}(\tau_n - 1)/\tau_{n+1} \geqslant (1-1/\tau_\infty)/(1+1/\tau_\infty) - 1/(\tau_\infty(\tau_\infty+1))$, as desired.

(iii): Follows from (ii) and (41). ■

Example 4.4 The condition

$$\sup_{n\in\mathbb{N}^*} (n/\tau_n) < +\infty \tag{48}$$

and the quotient

$$\frac{\tau_n - 1}{\tau_{n+1}} \tag{49}$$

play significant roles in subsequent convergence results. Here are the two popular examples of sequences that satisfy Assumption 4.1 as well as (48) already seen in Section 1:

(i) [10, 11, 16, 22] Set $\tau_1 := 1$, and set $(\forall n \in \mathbb{N}^*)\ \tau_{n+1} := \big(1+\sqrt{1+4\tau_n^2}\big)/2$. Then, it is straightforward to verify that $(\forall n \in \mathbb{N}^*)\ \tau_n^2 - \tau_{n+1}^2 + \tau_{n+1} = 0$ and that $(\tau_n)_{n\in\mathbb{N}^*}$ is an increasing sequence in $[1, +\infty[$. Moreover, an inductive argument shows that $(\forall n \in \mathbb{N}^*)\ \tau_n \geqslant (n+1)/2$, from which we obtain $\tau_\infty = +\infty$ and $\sup_{n\in\mathbb{N}^*}(n/\tau_n) \leqslant 2$. This and Lemma 4.3(i) guarantee that $\lim(\tau_n/n) = 1/2$.

Furthermore, it is part of the folklore that

$$\frac{\tau_n - 1}{\tau_{n+1}} = 1 - \frac{3}{n} + \mathrm{o}\left(\frac{1}{n}\right); \tag{50}$$

for completeness, a proof is provided in Appendix G.

(ii) [3, 5, 15, 27] Let $\rho \in [2, +\infty[$, and define $(\forall n \in \mathbb{N}^*)$ $\tau_n := (n + \rho - 1)/\rho$. Then, clearly $(\tau_n)_{n\in\mathbb{N}^*}$ is an increasing sequence in $[1, +\infty[$ with $\tau_\infty = +\infty$ and, for every $n \in \mathbb{N}^*$, we have $n/\tau_n = n\rho/(n + \rho - 1) \leqslant \rho$,

$$\tau_n^2 - \tau_{n+1}^2 + \tau_{n+1} = \left(\frac{n+\rho-1}{\rho}\right)^2 - \left(\frac{n+\rho}{\rho}\right)^2 + \frac{n+\rho}{\rho} = \frac{(\rho-2)n + (\rho-1)^2}{\rho^2} \geqslant \frac{1}{4}, \tag{51}$$

and

$$\frac{\tau_n - 1}{\tau_{n+1}} = \frac{n-1}{n+\rho} = 1 - \frac{1+\rho}{n} + \mathrm{O}\left(\frac{1}{n^2}\right). \tag{52}$$

We now turn to examples of the condition

$$(\exists \delta \in \,]0,1[)(\forall n \in \mathbb{N}^*) \quad \tau_{n+1}^2 - \tau_n^2 \leqslant \delta \tau_{n+1}, \tag{53}$$

which is of some interest in Section 5 (see (107)) and Section 6. Further examples of sequences that satisfy (53) can be found in [1, Section 5].

Example 4.5 Let $\rho \in \left]1, +\infty\right[$ and set

$$(\forall n \in \mathbb{N}^*) \quad \mu_n := \frac{\tau_n + \rho - 1}{\rho}. \tag{54}$$

Then

$$(\forall n \in \mathbb{N}^*) \quad \mu_{n+1}^2 - \mu_n^2 \leqslant \frac{1+\sqrt{5}}{2\rho}\mu_{n+1}. \tag{55}$$

If $\rho > (1+\sqrt{5})/2$, then the sequence $(\mu_n)_{n\in\mathbb{N}^*}$ satisfies (53) with $\delta = (1+\sqrt{5})/(2\rho) \in \,]0,1[$.

Proof. Indeed, since $(1+\sqrt{5})/2 > 1$, we derive from (54), (42), and (43) that

$$(\forall n \in \mathbb{N}^*) \quad \mu_{n+1}^2 - \mu_n^2 = \frac{\tau_{n+1}^2 - \tau_n^2 + 2(\rho-1)(\tau_{n+1} - \tau_n)}{\rho^2} \leqslant \frac{\tau_{n+1} + \frac{1+\sqrt{5}}{2}(\rho-1)}{\rho^2} \tag{56a}$$

$$< \frac{\frac{1+\sqrt{5}}{2}\tau_{n+1} + \frac{1+\sqrt{5}}{2}(\rho-1)}{\rho^2} = \frac{1+\sqrt{5}}{2\rho}\mu_{n+1}, \tag{56b}$$

as claimed. The remaining implication follows readily. ∎

Example 4.6 [7] Let $(a, d) \in \,]0, +\infty[\times \mathbb{R}_+$, set

$$(\forall n \in \mathbb{N}^*) \quad \tau_n := \left(\frac{n+a-1}{a}\right)^d, \tag{57}$$

and suppose that one of the following holds:

(i) $d = 0$.

(ii) $d \in \,]0,1]$ and $a > \max\{1, (2d)^{1/d}\}$.

Aujol and Dossal's [7, Lemma 3.2] yields

$$(\forall n \in \mathbb{N}^*) \quad \frac{1}{a^d} - \frac{2d}{a^{2d}} > 0 \quad \text{and} \quad \tau_n^2 - \tau_{n+1}^2 + \tau_{n+1} \geqslant \left(\frac{1}{a^d} - \frac{2d}{a^{2d}}\right)(n+a)^d > 0. \tag{58}$$

Let us add to their analysis by pointing out that if (ii) holds, then (53) holds with $\delta = (2d)/a^d \in]0,1[$. Indeed, (57) and (58) assert that

$$(\forall n \in \mathbb{N}^*) \quad \tau_{n+1}^2 - \tau_n^2 \leqslant \tau_{n+1} - \Big(\frac{1}{a^d} - \frac{2d}{a^{2d}}\Big)(n+a)^d = \tau_{n+1} - \Big(\frac{1}{a^d} - \frac{2d}{a^{2d}}\Big)a^d\tau_{n+1} = \delta\tau_{n+1}. \tag{59}$$

Also, note that if $d \in]0,1[$, then $\sup_{n\in\mathbb{N}^*}(n/\tau_n) = +\infty$ (by L'Hôpital's rule) in contrast to Example 4.4.

5 FISTA

In this section, we present three main results on FISTA. We again recall that Assumption 1.1 is in force and (see (4)) that

$$h = f + g \quad \text{and} \quad T = \operatorname{Prox}_{\gamma g} \circ \big(\operatorname{Id} - \gamma\nabla f\big). \tag{60}$$

Algorithm 5.1 (FISTA) Let $x_0 \in \mathcal{H}$, set $y_1 := x_0$, and update

$$\begin{array}{l} \text{for } n = 1,2,\ldots \\ \left\lfloor \begin{array}{ll} x_n & := Ty_n, \\ y_{n+1} & := x_n + \dfrac{\tau_n - 1}{\tau_{n+1}}(x_n - x_{n-1}), \end{array}\right. \end{array} \tag{61}$$

where T is as in (60) and $(\tau_n)_{n\in\mathbb{N}^*}$ satisfies (39).

We assume for the remainder of this section that

$$(x_n)_{n\in\mathbb{N}^*} \text{ is a sequence generated by Algorithm 5.1.} \tag{62}$$

We also set

$$(\forall n \in \mathbb{N}^*) \quad \sigma_n := h(x_n) + \frac{1}{2\gamma}\|x_n - x_{n-1}\|^2 \quad \text{and} \quad \alpha_n := \frac{\tau_n - 1}{\tau_{n+1}}. \tag{63}$$

Note that, by (40) and (39),

$$(\forall n \in \mathbb{N}^*) \quad 0 \leqslant \alpha_n = \frac{\tau_n - 1}{\tau_{n+1}} \leqslant \frac{\tau_n - 1}{\tau_n} < 1. \tag{64}$$

The first two items of the following result are due to Attouch and Cabot; see [1, Proposition 3].

Lemma 5.2 *The following holds:*

(i) $(\forall n \in \mathbb{N}^*)\ (2\gamma)^{-1}(1-\alpha_n^2)\|x_n - x_{n-1}\|^2 \leqslant \sigma_n - \sigma_{n+1}$.

(ii) *The sequence* $(\sigma_n)_{n\in\mathbb{N}^*}$ *is decreasing and convergent to a point in* $[-\infty, +\infty[$.

(iii) *Suppose that* $\inf_{n\in\mathbb{N}^*}\sigma_n > -\infty$. *Then the following hold:*

  (a) $\sum_{n\in\mathbb{N}^*}(1-\alpha_n^2)\|x_n - x_{n-1}\|^2 < +\infty$.

  (b) *Suppose that* $\sup_{n\in\mathbb{N}^*}\tau_n < +\infty$. *Then* $\inf_{n\in\mathbb{N}^*}(1-\alpha_n^2) > 0$ *and* $\sum_{n\in\mathbb{N}^*}\|x_n - x_{n-1}\|^2 < +\infty$.

Proof. (i): For every $n \in \mathbb{N}^*$, Lemma 3.2(i) (applied to $(y, x_-, \tau, \tau_+) = (y_n, x_{n-1}, \tau_n, \tau_{n+1})$) asserts that $\sigma_{n+1} \leqslant h(x_n) + \alpha_n^2(2\gamma)^{-1}\|x_n - x_{n-1}\|^2 = \sigma_n - (1-\alpha_n^2)(2\gamma)^{-1}\|x_n - x_{n-1}\|^2$, from which the desired inequality follows. (ii): A consequence of (i) and (64).

(iii)(a): By (i) and (63),

$$(\forall n \in \mathbb{N}^*) \quad \sum_{k=1}^{n} \frac{1-\alpha_k^2}{2\gamma}\|x_k - x_{k-1}\|^2 \leqslant \sum_{k=1}^{n}(\sigma_k - \sigma_{k+1}) = \sigma_1 - \sigma_{n+1} \leqslant \sigma_1 - \inf_{k\in\mathbb{N}^*}\sigma_k < +\infty. \tag{65}$$

Thus, $\sum_{n\in\mathbb{N}^*}(1-\alpha_n^2)\|x_n-x_{n-1}\|^2<+\infty$, as claimed.

(iii)(b): Because the function $]0,+\infty[\,\to\mathbb{R}\colon\xi\mapsto(\xi-1)/\xi$ is increasing and $(\forall n\in\mathbb{N}^*)$ $0<\tau_n\leqslant\tau_{n+1}\leqslant\tau_\infty$, we see that $(\forall n\in\mathbb{N}^*)$ $\alpha_n=(\tau_n-1)/\tau_{n+1}\leqslant(\tau_n-1)/\tau_n\leqslant(\tau_\infty-1)/\tau_\infty\in[0,1[$; therefore,

$$(\forall n\in\mathbb{N}^*)\quad 1-\alpha_n^2\geqslant 1-\left(\frac{\tau_\infty-1}{\tau_\infty}\right)^2>0. \tag{66}$$

Combining (66) and (iii)(a) yields the conclusion. ∎

We are ready for our first main result which establishes a minimizing property of the sequence $(x_n)_{n\in\mathbb{N}^*}$ generated by Algorithm 5.1 in the general setting.

Theorem 5.3 *The following holds:*

$$(\forall m\in\mathbb{N}^*)\quad \inf_{n\geqslant m}h(x_n)=\lim_n\min_{1\leqslant k\leqslant n}h(x_k)=\varliminf_n h(x_n)=\inf h. \tag{67}$$

Proof. Let us first establish that

$$(\forall m\in\mathbb{N}^*)\quad \inf_{n\geqslant m}h(x_n)=\inf h. \tag{68}$$

To do so, we proceed by contradiction: assume that there exists $N\in\mathbb{N}^*$ such that $\inf_{n\geqslant N}h(x_n)>\inf h$. Then, there exists $z\in\operatorname{dom}h$ satisfying

$$-\infty<h(z)<\inf_{n\geqslant N}h(x_n). \tag{69}$$

In turn, set $(\forall n\in\mathbb{N}^*)$ $\mu_n:=h(x_n)-h(z)$ and $u_n:=\tau_nx_n-(\tau_n-1)x_{n-1}-z$. For every $n\geqslant N$, in the light of Lemma 3.2(ii) (applied to $(y,x_-,\tau,\tau_+)=(y_n,x_{n-1},\tau_n,\tau_{n+1})$), we get

$$\begin{aligned}\tau_{n+1}^2\mu_{n+1}+(2\gamma)^{-1}\|u_{n+1}\|^2&\leqslant\tau_{n+1}(\tau_{n+1}-1)\mu_n+(2\gamma)^{-1}\|u_n\|^2 &(70a)\\ &=\tau_n^2\mu_n+(2\gamma)^{-1}\|u_n\|^2-\big(\tau_n^2-\tau_{n+1}^2+\tau_{n+1}\big)\mu_n. &(70b)\end{aligned}$$

Furthermore, due to (69),

$$(\forall n\geqslant N)\quad \mu_n=h(x_n)-h(z)>0. \tag{71}$$

Let us consider the following two possible cases.

(a) $\tau_\infty=+\infty$: By (41), $\tau_n\to+\infty$. Next, we derive from (70), (42), and (71) that $(\forall n\geqslant N)$ $\tau_n^2\mu_n\leqslant\tau_N^2\mu_N+(2\gamma)^{-1}\|u_N\|^2$ or, equivalently, by the very definition of $(\mu_n)_{n\in\mathbb{N}^*}$,

$$(\forall n\geqslant N)\quad h(x_n)\leqslant h(z)+\frac{\tau_N^2}{\tau_n^2}\mu_N+\frac{1}{2\gamma\tau_n^2}\|u_N\|^2. \tag{72}$$

Consequently, since $\tau_n\uparrow+\infty$, taking the limit superior in (72) gives $\inf_{n\geqslant N}h(x_n)\leqslant\varlimsup h(x_n)\leqslant h(z)$, which contradicts (69).

(b) $\tau_\infty<+\infty$: Set $(\forall n\geqslant N)$ $\xi_n:=\tau_n^2\mu_n+(2\gamma)^{-1}\|u_n\|^2$ and $\eta_n:=(\tau_n^2-\tau_{n+1}^2+\tau_{n+1})\mu_n$. Then, by (71), $\{\xi_n\}_{n\geqslant N}\subseteq\,]0,+\infty[$ and, by (69)&(42), $\{\eta_n\}_{n\geqslant N}\subseteq\mathbb{R}_+$. In turn, on the one hand, combining (70) and Lemma 2.3(ii), we infer that $\sum_{n\geqslant N}(\tau_n^2-\tau_{n+1}^2+\tau_{n+1})\mu_n=\sum_{n\geqslant N}\eta_n<+\infty$. On the other hand, because $(\forall n\in\mathbb{N}^*)$ $\tau_n^2\leqslant(\sup_{k\in\mathbb{N}^*}\tau_k)^2<+\infty$ and $\{\tau_n\}_{n\in\mathbb{N}^*}\subseteq[1,+\infty[$ by our assumption and (40),

$$(\forall p\in\mathbb{N}^*)\quad \sum_{n=N}^{N+p}(\tau_n^2-\tau_{n+1}^2+\tau_{n+1})=\tau_N^2-\tau_{N+p+1}^2+\sum_{n=N}^{N+p}\tau_{n+1}\geqslant\tau_N^2-\Big(\sup_{n\in\mathbb{N}^*}\tau_n\Big)^2+p+1, \tag{73}$$

from which we deduce that $\sum_{n\geqslant N}(\tau_n^2-\tau_{n+1}^2+\tau_{n+1})=+\infty$. Altogether, Lemma 2.2 and (71) guarantee that $\varliminf(h(x_n)-h(z))=\varliminf\mu_n=0$, i.e., $\varliminf h(x_n)=h(z)$. Consequently, due to the inequality $\inf_{n\geqslant N}h(x_n)\leqslant\varliminf h(x_n)$, it follows from (69) that $h(z)<h(z)$, which is absurd.

To summarize, we have reached a contradiction in each case, and therefore (68) holds. Thus, because $\min_{1\leqslant k\leqslant n} h(x_k) \to \inf_{m\in\mathbb{N}^*} h(x_m)$ as $n \to +\infty$, we infer from (68) that $\min_{1\leqslant k\leqslant n} h(x_k) \to \inf h$ as $n \to +\infty$. Finally, (68) guarantees that $\underline{\lim}\, h(x_n) = \sup_{n\in\mathbb{N}^*}\big(\inf_{k\geqslant n} h(x_k)\big) = \sup_{n\in\mathbb{N}^*}(\inf h) = \inf h$, which completes the proof. ■

Remark 5.4 In Theorem 5.3, we do not know whether or not $(h(x_n))_{n\in\mathbb{N}^*}$ converges to $\inf h$. However, Theorem 5.5, Theorem 5.10, and Proposition 5.8 suggest a positive answer.

We are now ready for our second main result (Theorem 5.5), which is a discrete version of Attouch et al.'s [3, Theorem 2.3]. When $(\tau_n)_{n\in\mathbb{N}^*}$ is as in Example 4.4(ii) with $\rho = 2$, items (ii) and (iv) were mentioned (without a detailed proof) in [6, Theorem 4.1]. The analysis of Theorem 5.5(iii) was motivated by Attouch and Cabot's [1, Proposition 3]. Furthermore, the boundedness of the sequences $(x_n)_{n\in\mathbb{N}^*}$ and $(n\|x_n - x_{n-1}\|)_{n\in\mathbb{N}^*}$ in the consistent case was first obtained in Attouch et al.'s [4, Proposition 4.3]; here, we slightly modified the proof of this result to obtain the boundedness of $(x_n)_{n\in\mathbb{N}^*}$ in a more general setting.

Theorem 5.5 *Suppose that*

$$\inf h > -\infty \quad \text{and} \quad \sup_{n\in\mathbb{N}^*}(n/\tau_n) < +\infty. \tag{74}$$

For every $z \in \operatorname{dom} h$, set $\beta_z := \tau_1^2(h(x_1) - h(z)) + (2\gamma)^{-1}\|\tau_1 x_1 - (\tau_1 - 1)x_0 - z\|^2$. Then the following hold:

(i) *For every $z \in \operatorname{dom} h$, we have*

$$\tau_n^2(h(x_n) - h(z)) + (2\gamma)^{-1}\|\tau_n x_n - (\tau_n - 1)x_{n-1} - z\|^2 \leqslant \beta_z + \tau_n^2\big(h(z) - \inf h\big) \sup_{k\in\mathbb{N}^*}(k/\tau_k) \tag{75}$$

and

$$(\forall n \in \mathbb{N}^*) \quad h(x_n) - h(z) \leqslant \frac{\beta_z}{\tau_n^2} + \big(h(z) - \inf h\big) \sup_{k\in\mathbb{N}^*}(k/\tau_k). \tag{76}$$

(ii) $h(x_n) \to \inf h$.

(iii) *$(x_n)_{n\in\mathbb{N}^*}$ is asymptotically regular, i.e., $x_n - x_{n-1} \to 0$.*

(iv) *Every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}^*}$ belongs to* $\operatorname{Argmin} h$.

(v) *Suppose that $(x_n)_{n\in\mathbb{N}^*}$ has a bounded subsequence. Then* $\operatorname{Argmin} h \neq \varnothing$.

(vi) *Suppose that* $\operatorname{Argmin} h = \varnothing$. *Then* $\|x_n\| \to +\infty$.

(vii) *Suppose that* $\operatorname{Argmin} h \neq \varnothing$. *Then the following hold:*

   (a) (Beck–Teboulle [11]) $h(x_n) - \min h = O(1/n^2)$ *as $n \to +\infty$; more precisely, for every $z \in \operatorname{Argmin} h$,*

$$(\forall n \in \mathbb{N}^*) \quad h(x_n) - \min h \leqslant \frac{\beta_z\big(\sup_{k\in\mathbb{N}^*}(k/\tau_k)\big)^2}{n^2}. \tag{77}$$

   (b) *The sequences $(x_n)_{n\in\mathbb{N}^*}$ and $(\tau_n(x_n - x_{n-1}))_{n\in\mathbb{N}^*}$ are bounded.*

Proof. Set $\kappa := \sup_{n\in\mathbb{N}^*}(n/\tau_n) \in \,]0, +\infty[$. Since $(\forall n \in \mathbb{N}^*)\ \tau_n \geqslant n/\kappa$, we see that

$$\tau_n \to +\infty. \tag{78}$$

(i): Take $z \in \operatorname{dom} h$, and set

$$(\forall n \in \mathbb{N}^*) \quad \mu_n := h(x_n) - h(z) \quad \text{and} \quad u_n := \tau_n x_n - (\tau_n - 1)x_{n-1} - z. \tag{79}$$

Now, for every $n \in \mathbb{N}^*$, since $\inf h > -\infty$, $\tau_n \leqslant \tau_{n+1}$, and $\tau_{n+1}(\tau_{n+1}-1) \leqslant \tau_n^2$, applying Lemma 3.2(iii) to $(y, x_-, \tau, \tau_+) = (y_n, x_{n-1}, \tau_n, \tau_{n+1})$ yields $\tau_{n+1}^2\mu_{n+1} + (2\gamma)^{-1}\|u_{n+1}\|^2 \leqslant \tau_n^2\mu_n + (2\gamma)^{-1}\|u_n\|^2 + \tau_{n+1}(h(z) - \inf h)$. Hence, because $(\tau_n)_{n\in\mathbb{N}^*}$ is increasing and $h(z) - \inf h \geqslant 0$, an inductive argument gives

$$(\forall n \in \mathbb{N}^*) \quad \tau_{n+1}^2\mu_{n+1} + (2\gamma)^{-1}\|u_{n+1}\|^2 \leqslant \tau_1^2\mu_1 + (2\gamma)^{-1}\|u_1\|^2 + \big(h(z) - \inf h\big)\sum_{k=2}^{n+1}\tau_k \tag{80a}$$

$$\leqslant \tau_1^2\mu_1 + (2\gamma)^{-1}\|u_1\|^2 + n\tau_{n+1}\big(h(z) - \inf h\big) \tag{80b}$$

$$\leqslant \beta_z + \kappa\tau_{n+1}^2\big(h(z) - \inf h\big). \tag{80c}$$

Therefore, since (75) trivially holds when $n = 1$, we obtain the conclusion. Consequently, (76) readily follows from (75).

(ii): For every $z \in \operatorname{dom} h$, taking the limit superior over n in (76) and using (78) yields $\overline{\lim}\, h(x_n) \leqslant h(z) + \kappa(h(z) - \inf h)$. Consequently, letting $h(z) \downarrow \inf h$, we conclude that $\overline{\lim}\, h(x_n) \leqslant \inf h$, as desired.

(iii): First, due to (63), $(\forall n \in \mathbb{N}^*)$ $\sigma_n \geqslant h(x_n) \geqslant \inf h > -\infty$, and thus,

$$\inf_{n\in\mathbb{N}^*} \sigma_n > -\infty. \tag{81}$$

Hence, we conclude via Lemma 5.2(ii) that $(\sigma_n)_{n\in\mathbb{N}^*}$ is convergent in $\mathbb{R}$. In turn, on the one hand, (ii) and (63) imply that

$$\big(\|x_n - x_{n-1}\|^2\big)_{n\in\mathbb{N}^*} \text{ converges in } \mathbb{R}. \tag{82}$$

On the other hand, (81) and Lemma 5.2(iii)(a) yield $\sum_{n\in\mathbb{N}^*}(1-\alpha_n^2)\|x_n - x_{n-1}\|^2 < +\infty$, and since $\sum_{n\in\mathbb{N}^*}(1-\alpha_n^2) = +\infty$ due to Lemma 2.1 and (78), we get from Lemma 2.2 that

$$\underline{\lim}\|x_n - x_{n-1}\|^2 = 0. \tag{83}$$

Altogether, combining (82) and (83) yields $x_n - x_{n-1} \to 0$, as announced.

(iv): Let x be a weak sequential cluster point of $(x_n)_{n\in\mathbb{N}^*}$, say $x_{k_n} \rightharpoonup x$. Then, since h is convex and lower semicontinuous, it is weakly sequentially lower semicontinuous by [8, Theorem 9.1]. Hence, (ii) entails that $h(x) \leqslant \underline{\lim}\, h(x_{k_n}) = \inf h$, which ensures that $x \in \operatorname{Argmin} h$.

(v): Combine (iv) and [8, Lemma 2.45].

(vi): This is the contrapositive of (v).

(vii)(a): Clear from (i) and (74).

(vii)(b): Fix $z \in \operatorname{Argmin} h$. For every $n \geqslant 2$, because $h(z) = \min h$, we derive from (75) that

$$(2\gamma)^{-1}\|\tau_n x_n - (\tau_n - 1)x_{n-1} - z\|^2 \leqslant \tau_n^2(h(x_n) - \min h) + (2\gamma)^{-1}\|\tau_n x_n - (\tau_n - 1)x_{n-1} - z\|^2 \leqslant \beta_z, \tag{84}$$

and now a simple expansion gives

$$2\gamma\beta_z \geqslant \|\tau_n x_n - (\tau_n - 1)x_{n-1} - z\|^2 \tag{85a}$$

$$= \|(x_n - z) + (\tau_n - 1)(x_n - x_{n-1})\|^2 \tag{85b}$$

$$= \|x_n - z\|^2 + 2(\tau_n - 1)\langle x_n - z \mid x_n - x_{n-1}\rangle + (\tau_n - 1)^2\|x_n - x_{n-1}\|^2 \tag{85c}$$

$$\geqslant \|x_n - z\|^2 + 2(\tau_n - 1)\langle x_n - z \mid x_n - x_{n-1}\rangle \tag{85d}$$

$$\overset{(15)}{=} \|x_n - z\|^2 + (\tau_n - 1)\big(\|x_n - z\|^2 - \|x_{n-1} - z\|^2 + \|x_n - x_{n-1}\|^2\big) \tag{85e}$$

$$= \tau_n\|x_n - z\|^2 - (\tau_n - 1)\|x_{n-1} - z\|^2 + (\tau_n - 1)\|x_n - x_{n-1}\|^2 \tag{85f}$$

$$\overset{(43)}{\geqslant} \tau_n\|x_n - z\|^2 - \tau_{n-1}\|x_{n-1} - z\|^2. \tag{85g}$$

In turn,

$$(\forall n \geqslant 2)\ \tau_n\|x_n - z\|^2 - \tau_1\|x_1 - z\|^2 = \sum_{k=2}^{n}\big(\tau_k\|x_k - z\|^2 - \tau_{k-1}\|x_{k-1} - z\|^2\big) \leqslant \sum_{k=2}^{n} 2\gamma\beta_z \leqslant 2\gamma\beta_z n. \tag{86}$$

Hence, since $\kappa = \sup_{n\in\mathbb{N}^*}(n/\tau_n) < +\infty$ and $(\forall n \in \mathbb{N}^*)\ \tau_1 \leqslant \tau_n$, we get

$$(\forall n \geqslant 2) \quad \|x_n - z\|^2 \leqslant 2\gamma\beta_z\frac{n}{\tau_n} + \frac{\tau_1}{\tau_n}\|x_1 - z\|^2 \leqslant 2\gamma\beta_z\kappa + \|x_1 - z\|^2, \tag{87}$$

from which the boundedness of $(x_n)_{n\in\mathbb{N}^*}$ follows. Consequently, because $(\tau_n(x_n - x_{n-1}))_{n\in\mathbb{N}^*} = (\tau_n x_n - (\tau_n - 1)x_{n-1} - z)_{n\in\mathbb{N}^*} - (x_{n-1} - z)_{n\in\mathbb{N}^*}$ and both sequences on the right-hand side are bounded due to (84) and (87), we conclude that $(\tau_n(x_n - x_{n-1}))_{n\in\mathbb{N}^*}$ is bounded, as announced. ■

Remark 5.6 By choosing the sequence $(\tau_n)_{n\in\mathbb{N}^*}$ as in Example 4.4(i), we shall see in Proposition 5.8 that Theorem 5.5(ii) is still valid even when the assumption that $\inf h > -\infty$ is omitted. Therefore, it is appealing to conjecture that this assumption can be left out in Theorem 5.5(ii). In stark contrast, it is crucial to assume that h is bounded from below in Theorem 5.5(iii), as illustrated in Example 5.7.

Example 5.7 Suppose that $\mathcal{H} = \mathbb{R}$, that $f\colon \mathcal{H} \to \mathbb{R} : x \mapsto -x$, that $g = 0$, that $\gamma = 1$, and that $\tau_n \uparrow \tau_\infty = +\infty$. Then, since $\operatorname{Prox}_g = \operatorname{Id}$ and $(\forall x \in \mathcal{H})\ \nabla f(x) = -1$, we see that (61) turns into

$$\begin{array}{l} \text{for } n = 1, 2, \ldots \\ \left\lfloor \begin{array}{ll} x_n & := y_n + 1, \\ y_{n+1} & := x_n + \dfrac{\tau_n - 1}{\tau_{n+1}}(x_n - x_{n-1}). \end{array} \right. \end{array} \tag{88}$$

Hence, $(\forall n \in \mathbb{N}^*)\ x_{n+1} - 1 = y_{n+1} = x_n + (\tau_n - 1)(x_n - x_{n-1})/\tau_{n+1}$, and upon setting $(\forall n \in \mathbb{N}^*)\ z_n := x_n - x_{n-1}$, we obtain

$$(\forall n \in \mathbb{N}^*) \quad z_{n+1} = 1 + \frac{\tau_n - 1}{\tau_{n+1}} z_n. \tag{89}$$

Let us establish that $z_n \to +\infty$. First, since $y_1 = x_0$ by Algorithm 5.1, we get from (88) that $z_1 = x_1 - x_0 = x_1 - y_1 = 1$. In turn, by induction and (89), $(\forall n \in \mathbb{N}^*)\ z_n \geqslant 1$. We now suppose to the contrary that $\xi := \varliminf z_n \in \mathbb{R}_+$. Then, taking the limit inferior over n in (89) and using Lemma 4.3 yield $\xi = 1 + 1\cdot\xi = 1 + \xi$, which is absurd. Therefore, $\xi = +\infty$, and it follows that $x_n - x_{n-1} = z_n \to +\infty$.

Proposition 5.8 *Suppose that the sequence $(\tau_n)_{n\in\mathbb{N}^*}$ is as in Example 4.4(i). Then $h(x_n) \to \inf h \in [-\infty, +\infty[$.*

Proof. First, as seen in Example 4.4(i),

$$(\forall n \in \mathbb{N}^*) \quad \tau_{n+1}^2 - \tau_{n+1} = \tau_n^2. \tag{90}$$

Now it is sufficient to show that $\varlimsup h(x_n) \leqslant \inf h$. To do so, fix $z \in \operatorname{dom} h$, and set $(\forall n \in \mathbb{N}^*)\ \mu_n := h(x_n) - h(z)$ and $u_n := \tau_n x_n - (\tau_n - 1)x_{n-1} - z$. Then, according to Lemma 3.2(ii) and (90),

$$(\forall n \in \mathbb{N}^*) \quad \tau_{n+1}^2\mu_{n+1} + (2\gamma)^{-1}\|u_{n+1}\|^2 \leqslant \tau_{n+1}(\tau_{n+1} - 1)\mu_n + (2\gamma)^{-1}\|u_n\|^2 = \tau_n^2\mu_n + (2\gamma)^{-1}\|u_n\|^2. \tag{91}$$

Thus,

$$(\forall n \in \mathbb{N}^*) \quad h(x_n) - h(z) = \mu_n \leqslant \frac{\tau_n^2\mu_n + (2\gamma)^{-1}\|u_n\|^2}{\tau_n^2} \leqslant \frac{\tau_1^2\mu_1 + (2\gamma)^{-1}\|u_1\|^2}{\tau_n^2}. \tag{92}$$

Hence, because $\lim \tau_n = +\infty$, taking the limit superior over n yields $\varlimsup h(x_n) \leqslant h(z)$. Consequently, since z is an arbitrary element of $\operatorname{dom} h$, we conclude that $\varlimsup h(x_n) \leqslant \inf h$, as required. ■

Remark 5.9 Proposition 5.8 is a special case of the *accelerated inexact forward-backward splitting* developed in [28]; see [28, Theorem 4.3 and Remark 3].

We now turn to our third main result, which concerns the case where the parameter sequence $(\tau_n)_{n\in\mathbb{N}^*}$ in Assumption 4.1 is bounded.

Theorem 5.10 *Suppose that $\tau_\infty < +\infty$. Then the following hold:*

(i) $\lim \sigma_n = \lim h(x_n) = \inf h \in [-\infty, +\infty[$.

(ii) *Assume that* $\inf h > -\infty$. *Then the following hold:*

  (a) $\sum_{n\in\mathbb{N}^*} \|x_n - x_{n-1}\|^2 < +\infty$.

  (b) *Every weak sequential cluster point of* $(x_n)_{n\in\mathbb{N}^*}$ *lies in* $\operatorname{Argmin} h$.

(iii) *Assume that* $(x_n)_{n\in\mathbb{N}^*}$ *has a bounded subsequence. Then* $\operatorname{Argmin} h \neq \varnothing$.

(iv) *Assume that* $\operatorname{Argmin} h = \varnothing$. *Then* $\|x_n\| \to +\infty$.

(v) *Assume that* $\operatorname{Argmin} h \neq \varnothing$. *Then the following hold:*

  (a) (Attouch–Cabot [1]) $h(x_n) - \min h = o(1/n)$ *as* $n \to +\infty$.

  (b) $\sum_{n\in\mathbb{N}^*} n\|x_n - x_{n-1}\|^2 < +\infty$. *As a consequence,* $\|x_n - x_{n-1}\| = o(1/\sqrt{n})$ *as* $n \to +\infty$.

Proof. (i): Since, by (63), $(\forall n \in \mathbb{N}^*)$ $\inf h \leqslant h(x_n) \leqslant \sigma_n$ and, by Lemma 5.2(ii), $(\sigma_n)_{n\in\mathbb{N}^*}$ converges to a point $\sigma \in [-\infty, +\infty[$, it is enough to verify that $\sigma = \lim \sigma_n = \inf h$. Assume to the contrary that

$$-\infty \leqslant \inf h < \sigma. \tag{93}$$

It then follows that $\inf_{n\in\mathbb{N}^*} \sigma_n > -\infty$, and Lemma 5.2(iii)(b) thus yields $\|x_n - x_{n-1}\|^2 \to 0$, from which and (63) we deduce that $h(x_n) \to \sigma$. This and Theorem 5.3 imply that $\sigma = \inf h$. This and (93) yield a contradiction.

(ii)(a): Our assumption ensures that $\inf_{n\in\mathbb{N}^*} \sigma_n > -\infty$, and therefore, thanks to the boundedness of $(\tau_n)_{n\in\mathbb{N}^*}$, Lemma 5.2(iii)(b) yields $\sum_{n\in\mathbb{N}^*} \|x_n - x_{n-1}\|^2 < +\infty$.

(ii)(b)&(iii)&(iv): Similar to Theorem 5.5(iv)&(v)&(vi), respectively.

(v): Fix $z \in \operatorname{Argmin} h$, and set $(\forall n \in \mathbb{N}^*)$ $\mu_n := h(x_n) - h(z) = h(x_n) - \min h \geqslant 0$ and $u_n := \tau_n x_n - (\tau_n - 1)x_{n-1} - z$. By (41), we have

$$\tau_n \uparrow \tau_\infty, \tag{94}$$

which implies that $\tau_n^2 - \tau_{n+1}^2 + \tau_{n+1} \to \tau_\infty$. Therefore, because $\tau_\infty \in]0, +\infty[$, there exists $N \in \mathbb{N}^*$ such that

$$\inf_{n\geqslant N} (\tau_n^2 - \tau_{n+1}^2 + \tau_{n+1}) \geqslant \frac{\tau_\infty}{2}. \tag{95}$$

Next, for every $n \geqslant N$, using Lemma 3.2(ii) with $(y, x_-, \tau, \tau_+) = (y_n, x_{n-1}, \tau_n, \tau_{n+1})$, we get $\tau_{n+1}^2 \mu_{n+1} + (2\gamma)^{-1}\|u_{n+1}\|^2 \leqslant \tau_n^2 \mu_n + (2\gamma)^{-1}\|u_n\|^2 - (\tau_n^2 - \tau_{n+1}^2 + \tau_{n+1})\mu_n$. Hence, because $\{\tau_n^2 \mu_n + (2\gamma)^{-1}\|u_n\|^2\}_{n\geqslant N} \subseteq \mathbb{R}_+$ and, by (42), $\{(\tau_n^2 - \tau_{n+1}^2 + \tau_{n+1})\mu_n\}_{n\geqslant N} \subseteq \mathbb{R}_+$, Lemma 2.3(ii) and (95) give $(\tau_\infty/2) \sum_{n\geqslant N} \mu_n \leqslant \sum_{n\geqslant N} (\tau_n^2 - \tau_{n+1}^2 + \tau_{n+1})\mu_n < +\infty$. This, (ii)(a), and (63) ensure that

$$\sum_{n\in\mathbb{N}^*} (\sigma_n - \min h) = \sum_{n\in\mathbb{N}^*} \big(\mu_n + (2\gamma)^{-1}\|x_n - x_{n-1}\|^2\big) < +\infty. \tag{96}$$

Furthermore, Lemma 5.2(ii) and (i) yield

$$\sigma_n - \min h \downarrow 0. \tag{97}$$

(v)(a): Appealing to (96) and (97), Lemma 2.5 guarantees that $n(\sigma_n - \min h) \to 0$. Consequently, since $(\forall n \in \mathbb{N}^*)$ $\sigma_n - \min h = (h(x_n) - \min h) + (2\gamma)^{-1}\|x_n - x_{n-1}\|^2 \geqslant h(x_n) - \min h \geqslant 0$, the conclusion follows.

(v)(b): Thanks to (96) and (97), we derive from Lemma 2.5 that

$$\sum_{n\in\mathbb{N}^*} n(\sigma_n - \sigma_{n+1}) = \sum_{n\in\mathbb{N}^*} n[(\sigma_n - \min h) - (\sigma_{n+1} - \min h)] < +\infty, \tag{98}$$

and hence, by Lemma 5.2(i), $\sum_{n\in\mathbb{N}^*} n(1-\alpha_n^2)\|x_n - x_{n-1}\|^2 < +\infty$. Thus, because $\inf_{n\in\mathbb{N}^*}(1-\alpha_n^2) > 0$ due to the boundedness of $(\tau_n)_{n\in\mathbb{N}^*}$ and Lemma 5.2(iii)(b), we conclude that $\sum_{n\in\mathbb{N}^*} n\|x_n - x_{n-1}\|^2 < +\infty$. This gives $n\|x_n - x_{n-1}\|^2 \to 0$, i.e., $\|x_n - x_{n-1}\| = \mathrm{o}(1/\sqrt{n})$ as $n \to +\infty$, as desired. ■

Remark 5.11

- (i) In the case of the classical forward-backward algorithm (without the extrapolation step) with line-searches, results similar to Theorem 5.10(i)&(iv) were established in [12, Theorem 4.2] by Bello Cruz and Nghia. To the best of our knowledge, Theorem 5.10(i) is new in the setting of Algorithm 5.1.
- (ii) Theorem 5.10(v)(a) was obtained by Attouch and Cabot [1, Corollary 20(iii)]. Here we provide a proof based on the technique developed in [1] to be self-contained.
- (iii) The summabilities established in Theorem 5.10(ii)(a)&(v)(b) are new. Nevertheless, in the case of the forward-backward algorithm, i.e., when $\tau_n \equiv 1$, Theorem 5.10(v)(b) appears implicitly in the Beck and Teboulle's proof of [11, Theorem 3.1].

In the case of the classical forward-backward algorithm, by applying [14, Corollary 1.5] to the forward-backward operator $\mathrm{Prox}_{\gamma g} \circ (\mathrm{Id} - \gamma\nabla f)$, we obtain further information on the sequence $(x_n)_{n\in\mathbb{N}^*}$ as follows.

Proposition 5.12 *Suppose that* $(\forall n \in \mathbb{N}^*)\ \tau_n = 1$, *and set*[1,2] $v := \mathrm{P}_{\overline{\mathrm{ran}(\mathrm{Id}-T)}} 0$. *Then* $x_n - x_{n-1} \to v$.

Proof. By assumption, Algorithm 5.1 becomes $(\forall n \in \mathbb{N}^*)\ x_n = T^n x_0$. Next, we learn from [21, Proposition 3.2 and Corollary 4.2] that T is *averaged*, i.e., there exists $\alpha \in]0,1[$ and a nonexpansive operator $R\colon \mathcal{H} \to \mathcal{H}$ such that $T = (1-\alpha)\,\mathrm{Id} + \alpha R$. Hence, we conclude via [14, Proposition 1.3 and Corollary 1.5] that $x_n - x_{n-1} = T^n x_0 - T^{n-1} x_0 \to v$. For an alternative proof of [14, Corollary 1.2] in the Hilbert space setting, see [21, Proposition 2.1]. ■

Remark 5.13 Some comments are in order.

- (i) In stark contrast to Proposition 5.12 and Theorem 5.10, if $\tau_\infty = +\infty$, then it may happen that $\|x_n - x_{n-1}\| \to +\infty$ (see Example 5.7).
- (ii) For a recent study on the forward-backward operator T, we refer the reader to [21].

Proposition 5.14 *Suppose that* $\mathrm{Argmin}\, h \neq \varnothing$, *that* $(\tau_n^2(h(x_n) - \min h))_{n\in\mathbb{N}^*}$ *converges in* $\mathbb{R}$, *and that* $\tau_n\|x_n - x_{n-1}\| \to 0$. *Then the following hold:*

- (i) $h(x_n) \to \min h$.
- (ii) *The sequence* $(x_n)_{n\in\mathbb{N}^*}$ *converges weakly to a point in* $\mathrm{Argmin}\, h$.
- (iii) *Suppose that* $\mathrm{int}(\mathrm{Argmin}\, h) \neq \varnothing$. *Then* $(x_n)_{n\in\mathbb{N}^*}$ *converges strongly to a point in* $\mathrm{Argmin}\, h$.

Proof. Set

$$(\forall n \in \mathbb{N}^*) \quad z_n := \tau_n x_n - (\tau_n - 1)x_{n-1} \text{ and } \varepsilon_n := 2\gamma\big(\tau_n^2(h(x_n) - \min h) - \tau_{n+1}^2(h(x_{n+1}) - \min h)\big). \tag{99}$$

Since, by (40) and (99), $(\forall n \in \mathbb{N}^*)\ \|z_n - x_n\| = (\tau_n - 1)\|x_n - x_{n-1}\| \leqslant \tau_n\|x_n - x_{n-1}\|$ and since, by our assumption, $\tau_n\|x_n - x_{n-1}\| \to 0$, we see that

$$z_n - x_n \to 0. \tag{100}$$

[1] For a nonempty set C, P_C denotes the projector associated with C.

[2] The set $\overline{\mathrm{ran}(\mathrm{Id} - T)}$ is closed and convex by [21, Corollary 4.2] and [24, Lemma 4].

Next, due to our assumption and

$$(\forall n \in \mathbb{N}^*) \quad \sum_{k=1}^{n} \varepsilon_k = 2\gamma \sum_{k=1}^{n} \left[\tau_k^2(h(x_k) - \min h) - \tau_{k+1}^2(h(x_{k+1}) - \min h)\right] \tag{101a}$$

$$= 2\gamma\big(\tau_1^2(h(x_1) - \min h) - \tau_{n+1}^2(h(x_{n+1}) - \min h)\big), \tag{101b}$$

we see that

$$\sum_{n\in\mathbb{N}^*} \varepsilon_n \text{ is convergent in } \mathbb{R}. \tag{102}$$

Let us now establish that

$$(\forall z \in \operatorname{Argmin} h)(\forall n \in \mathbb{N}^*) \quad \|z_{n+1} - z\|^2 \leqslant \|z_n - z\|^2 + \varepsilon_n. \tag{103}$$

Fix $z \in \operatorname{Argmin} h$ and $n \in \mathbb{N}^*$. Applying Lemma 3.2(ii) to $(y, x_-, \tau, \tau_+) = (y_n, x_{n-1}, \tau_n, \tau_{n+1})$ and invoking (42) yields

$$\tau_{n+1}^2(h(x_{n+1}) - \min h) + (2\gamma)^{-1}\|z_{n+1} - z\|^2 \leqslant \tau_{n+1}(\tau_{n+1} - 1)\underbrace{(h(x_n) - \min h)}_{\geqslant 0} + (2\gamma)^{-1}\|z_n - z\|^2 \tag{104a}$$

$$\leqslant \tau_n^2(h(x_n) - \min h) + (2\gamma)^{-1}\|z_n - z\|^2, \tag{104b}$$

from which and (99) we obtain (103).

(i): Since, by assumption, $(\tau_n^2(h(x_n) - \min h))_{n\in\mathbb{N}^*}$ converges and since, by (41), $(1/\tau_n^2)_{n\in\mathbb{N}^*}$ converges in $\mathbb{R}$, it follows that $(h(x_n) - \min h)_{n\in\mathbb{N}^*}$ is convergent in $\mathbb{R}$. Therefore, due to Theorem 5.3, $h(x_n) - \min h \to 0$.

(ii): In the light of (i), arguing similarly to the proof of Theorem 5.5(iv), we conclude that

$$\text{every weak sequential cluster point of } (x_n)_{n\in\mathbb{N}^*} \text{ belongs to } \operatorname{Argmin} h. \tag{105}$$

In turn, appealing to (102) and (103), Lemma 2.7(i) implies that

$$(\forall z \in \operatorname{Argmin} h) \quad (\|z_n - z\|)_{n\in\mathbb{N}^*} \text{ is convergent in } \mathbb{R}. \tag{106}$$

Thus, combining (100)&(105)&(106), we get via Lemma 2.6 that $(x_n)_{n\in\mathbb{N}^*}$ converges weakly to a point in $\operatorname{Argmin} h$.

(iii): Since $\operatorname{int}(\operatorname{Argmin} h) \neq \varnothing$, owing to Lemma 2.7(ii), we derive from (102) and (103) that there exists $z \in \mathcal{H}$ such that $z_n \to z$. Hence, by (100), $x_n \to z$, and (ii) implies that $z \in \operatorname{Argmin} h$. To sum up, $(x_n)_{n\in\mathbb{N}^*}$ converges strongly to a minimizer of h. ∎

Corollary 5.15 *Suppose that* $\operatorname{Argmin} h \neq \varnothing$ *and that* $\sup_{n\in\mathbb{N}^*} \tau_n < +\infty$. *Then* $(x_n)_{n\in\mathbb{N}^*}$ *converges weakly to a point in* $\operatorname{Argmin} h$. *Moreover, if* $\operatorname{int}(\operatorname{Argmin} h) \neq \varnothing$, *then the convergence is strong.*

Proof. By Theorem 5.10(v), we see that $h(x_n) - \min h \to 0$ and $\|x_n - x_{n-1}\| \to 0$, and since $\sup_{n\in\mathbb{N}^*} \tau_n < +\infty$, it follows that $\tau_n^2(h(x_n) - \min h) \to 0$ and $\tau_n\|x_n - x_{n-1}\| \to 0$. The conclusion thus follows from Proposition 5.14. ∎

Remark 5.16 Consider the setting of Corollary 5.15. Although the weak convergence of the sequence $(x_n)_{n\in\mathbb{N}^*}$ has been shown in [1, Corollary 20(iv)], our Fejér-based proof here is new and may suggest other approaches to tackle the convergence of $(x_n)_{n\in\mathbb{N}^*}$ in the setting Theorem 5.5(vii).

We conclude this section with an instance where the assumption of Proposition 5.14 holds.

Example 5.17 Suppose, in addition to Assumption 4.1, that there exists $\delta \in \left]0,1\right[$ such that

$$(\forall n \in \mathbb{N}^*) \quad \tau_{n+1}^2 - \tau_n^2 \leqslant \delta\tau_{n+1} \tag{107}$$

(see Examples 4.5 and 4.6). Then Attouch and Cabot's [1, Theorem 9] yields $\tau_n^2(h(x_n) - \min h) \to 0$ and $\tau_n\|x_n - x_{n-1}\| \to 0$.

6 MFISTA

In this section, we discuss the minimizing property of the sequence generated by MFISTA. The monotonicity of function values allows us to overcome the issue stated in Remark 5.4. Compared to Beck and Teboulle's [10, Theorem 5.1] (see also [9, Theorem 10.40]), we allow other possibilities for the choice of $(\tau_n)_{n\in\mathbb{N}^*}$ in Theorem 6.1(vi). Furthermore, we provide in item (vii), which was motivated by [1, Theorem 9], a better rate of convergence.

Theorem 6.1 *In addition to Assumption 4.1, suppose that $\tau_\infty = +\infty$. Let $x_0 \in \mathcal{H}$, set $y_1 := x_0$, and update*

$$\text{for } n = 1,2,\ldots \quad \left\lfloor \begin{array}{ll} z_n & := Ty_n, \\ x_n & := \begin{cases} x_{n-1}, & \text{if } h(x_{n-1}) \leqslant h(z_n); \\ z_n, & \text{otherwise}, \end{cases} \\ y_{n+1} & := x_n + \dfrac{\tau_n}{\tau_{n+1}}(z_n - x_n) + \dfrac{\tau_n - 1}{\tau_{n+1}}(x_n - x_{n-1}), \end{array} \right. \tag{108}$$

where T is as in (16). Furthermore, set

$$(\forall n \in \mathbb{N}^*) \quad \sigma_n := h(x_n) + \frac{1}{2\gamma}\|z_n - x_{n-1}\|^2. \tag{109}$$

Then the following hold:

(i) *$(h(x_n))_{n\in\mathbb{N}^*}$ is decreasing and $h(x_n) \downarrow \inf h$.*

(ii) *$(\sigma_n)_{n\in\mathbb{N}^*}$ is decreasing and $\sigma_n \downarrow \inf h$.*

(iii) *Suppose that $\inf h > -\infty$. Then $z_n - x_{n-1} \to 0$ and $x_n - x_{n-1} \to 0$.*

(iv) *Suppose that $(x_n)_{n\in\mathbb{N}^*}$ has a bounded subsequence. Then $\operatorname{Argmin} h \neq \varnothing$.*

(v) *Suppose that $\operatorname{Argmin} h = \varnothing$. Then $\|x_n\| \to +\infty$.*

(vi) *Suppose that $\operatorname{Argmin} h \neq \varnothing$. Then $h(x_n) - \min h = \mathrm{O}(1/\tau_n^2)$ as $n \to +\infty$.*

(vii) *Suppose that $\operatorname{Argmin} h \neq \varnothing$ and that there exists $\delta \in \left]0,1\right[$ such that*

$$(\forall n \in \mathbb{N}^*) \quad \tau_{n+1}^2 - \tau_n^2 \leqslant \delta\tau_{n+1}. \tag{110}$$

Then

$$h(x_n) - \min h = \mathrm{o}\left(\frac{1}{\sum_{k=1}^n \tau_k}\right) \text{ as } n \to +\infty \tag{111}$$

and

$$h(x_n) - \min h = \mathrm{o}\left(\frac{1}{\tau_n^2}\right) \text{ as } n \to +\infty. \tag{112}$$

Proof. (i): By (108), the sequence $(h(x_n))_{n\in\mathbb{N}^*}$ is decreasing, from which we have $h(x_n) \downarrow \inf_{k\in\mathbb{N}^*} h(x_k)$. Therefore, it suffices to prove that $\inf_{n\in\mathbb{N}^*} h(x_n) = \inf h$. To this end, assume to the contrary that $\inf_{n\in\mathbb{N}^*} h(x_n) > \inf h$. This yields the existence of a point $w \in \operatorname{dom} h$ such that

$$\inf_{n\in\mathbb{N}^*} h(x_n) > h(w). \tag{113}$$

Set

$$(\forall n \in \mathbb{N}^*) \quad \mu_n := h(x_n) - h(w) \quad \text{and} \quad u_n := \tau_n z_n - (\tau_n - 1)x_{n-1} - w. \tag{114}$$

In turn, for every $n \in \mathbb{N}^*$, because, by (42), $\tau_{n+1}(\tau_{n+1} - 1) \leqslant \tau_n^2$ and, by (113), $\mu_n > 0$, it follows from Lemma 3.3(ii) (applied to $(y, x_-, \tau, \tau_+) = (y_n, x_{n-1}, \tau_n, \tau_{n+1})$) that

$$\tau_{n+1}^2 \mu_{n+1} + (2\gamma)^{-1}\|u_{n+1}\|^2 \leqslant \tau_{n+1}(\tau_{n+1} - 1)\mu_n + (2\gamma)^{-1}\|u_n\|^2 \leqslant \tau_n^2 \mu_n + (2\gamma)^{-1}\|u_n\|^2. \tag{115}$$

Hence,

$$(\forall n \in \mathbb{N}^*) \quad h(x_n) - h(w) = \mu_n \leqslant \frac{1}{\tau_n^2}\big(\tau_n^2 \mu_n + (2\gamma)^2\|u_n\|^2\big) \leqslant \frac{1}{\tau_n^2}\big(\tau_1^2 \mu_1 + (2\gamma)^2\|u_1\|^2\big). \tag{116}$$

Consequently, since $h(x_n) \downarrow \inf_{k\in\mathbb{N}^*} h(x_k)$ and $\tau_n \to +\infty$, we derive from (116) that $\inf_{n\in\mathbb{N}^*} h(x_n) \leqslant h(w)$, which contradicts (113).

(ii): Let us first show that $(\sigma_n)_{n\in\mathbb{N}^*}$ is decreasing. Towards this end, for every $n \in \mathbb{N}^*$, we deduce from Lemma 3.3(i) that $\sigma_{n+1} = h(x_{n+1}) + (2\gamma)^{-1}\|z_{n+1} - x_n\|^2 \leqslant h(x_n) + (2\gamma)^{-1}\tau_n^2\|z_n - x_{n-1}\|^2/\tau_{n+1}^2 = \sigma_n - (2\gamma)^{-1}(1 - \tau_n^2/\tau_{n+1}^2)\|z_n - x_{n-1}\|^2$. Therefore,

$$(\forall n \in \mathbb{N}^*) \quad \frac{1}{2\gamma}\left(1 - \frac{\tau_n^2}{\tau_{n+1}^2}\right)\|z_n - x_{n-1}\|^2 \leqslant \sigma_n - \sigma_{n+1}, \tag{117}$$

and because $(\forall n \in \mathbb{N}^*)\ 0 < \tau_n/\tau_{n+1} \leqslant 1$, we conclude that

$$(\sigma_n)_{n\in\mathbb{N}^*} \text{ is decreasing.} \tag{118}$$

It remains to show that $\sigma_n \to \inf h$. Set $\sigma := \inf_{n\in\mathbb{N}^*} \sigma_n$. Due to (118),

$$\sigma_n \downarrow \sigma \tag{119}$$

and it therefore suffices to prove that $\sigma = \inf h$. Let us argue by contradiction: assume that $\sigma > \inf h \geqslant -\infty$. By (117),

$$(\forall n \in \mathbb{N}^*) \quad \frac{1}{2\gamma}\sum_{k=1}^{n}\left(1 - \frac{\tau_k^2}{\tau_{k+1}^2}\right)\|z_k - x_{k-1}\|^2 \leqslant \sum_{k=1}^{n}(\sigma_k - \sigma_{k+1}) = \sigma_1 - \sigma_{n+1} \leqslant \sigma_1 - \sigma < +\infty, \tag{120}$$

which implies that $\sum_{n\in\mathbb{N}^*}(1 - \tau_n^2/\tau_{n+1}^2)\|z_n - x_{n-1}\|^2 < +\infty$. Thus, since $\sum_{n\in\mathbb{N}^*}(1 - \tau_n^2/\tau_{n+1}^2) = +\infty$ by Lemma 2.1, Lemma 2.2 guarantees that $\underline{\lim}\|z_n - x_{n-1}\|^2 = 0$, i.e., $\underline{\lim}\|z_n - x_{n-1}\| = 0$. In turn, let $(k_n)_{n\in\mathbb{N}^*}$ be a strictly increasing sequence in $\mathbb{N}^*$ such that $\|z_{k_n} - x_{k_n-1}\| \to \underline{\lim}\|z_n - x_{n-1}\| = 0$. It follows from (i) and (119) that $\sigma \leftarrow \sigma_{k_n} = h(x_{k_n}) + (2\gamma)^{-1}\|z_{k_n} - x_{k_n-1}\|^2 \to \inf h + 0 = \inf h$. Consequently, $\sigma = \inf h$, which violates the assumption that $\sigma > \inf h$. To summarize, we have shown that $\sigma_n \downarrow \inf h$.

(iii): Since $\inf h > -\infty$, combining (i), (ii), and (109) gives $z_n - x_{n-1} \to 0$. To show that $x_n - x_{n-1} \to 0$, we infer from (108) that, for every $n \in \mathbb{N}^*$, $x_n - x_{n-1} = x_{n-1} - x_{n-1} = 0$ if $h(x_{n-1}) \leqslant h(z_n)$, and $x_n - x_{n-1} = z_n - x_{n-1}$ otherwise; therefore, $(\forall n \in \mathbb{N}^*)\ \|x_n - x_{n-1}\| \leqslant \|z_n - x_{n-1}\|$. Consequently, because $z_n - x_{n-1} \to 0$, it follows that $x_n - x_{n-1} \to 0$, as required.

(iv)&(v): Straightforward.

(vi): Fix $w \in \operatorname{Argmin} h$ and define $(\forall n \in \mathbb{N}^*)\ \mu_n := h(x_n) - h(w) = h(x_n) - \min h \geqslant 0$ and $u_n := \tau_n z_n - (\tau_n - 1)x_{n-1} - w$. Due to (42) and the fact that $\{\mu_n\}_{n\in\mathbb{N}^*} \subseteq \mathbb{R}_+$, Lemma 3.3(ii) entails that $(\forall n \in \mathbb{N}^*)\ \tau_{n+1}^2\mu_{n+1} + (2\gamma)^{-1}\|u_{n+1}\|^2 \leqslant \tau_{n+1}(\tau_{n+1} - 1)\mu_n + (2\gamma)^{-1}\|u_n\|^2 \leqslant \tau_n^2\mu_n + (2\gamma)^{-1}\|u_n\|^2$. Hence,

$$(\forall n \in \mathbb{N}^*) \quad h(x_n) - \min h = \mu_n \leqslant \frac{1}{\tau_n^2}\big(\tau_n^2\mu_n + (2\gamma)^{-1}\|u_n\|^2\big) \leqslant \frac{1}{\tau_n^2}\big(\tau_1^2\mu_1 + (2\gamma)^{-1}\|u_1\|^2\big), \tag{121}$$

which verifies the claim.

(vii): Let us adapt the notation of (vi). Since $\{\mu_n\}_{n\in\mathbb{N}^*} \subseteq \mathbb{R}_+$, we derive from Lemma 3.3(ii) and (110) that

$$(\forall n \in \mathbb{N}^*) \quad \tau_{n+1}^2 \mu_{n+1} + (2\gamma)^{-1} \|u_{n+1}\|^2 \leqslant \tau_{n+1}(\tau_{n+1} - 1)\mu_n + (2\gamma)^{-1}\|u_n\|^2 \tag{122a}$$

$$= \tau_n^2 \mu_n + (2\gamma)^{-1}\|u_n\|^2 - \big(\tau_n^2 - \tau_{n+1}^2 + \tau_{n+1}\big)\mu_n \tag{122b}$$

$$\leqslant \tau_n^2 \mu_n + (2\gamma)^{-1}\|u_n\|^2 - (1-\delta)\tau_{n+1}\mu_n. \tag{122c}$$

On the other hand, since $\delta \in \,]0,1[$ and $\{\mu_n\}_{n\in\mathbb{N}^*} \subseteq \mathbb{R}_+$, it follows that $\{(1-\delta)\tau_{n+1}\mu_n\}_{n\in\mathbb{N}^*} \subseteq \mathbb{R}_+$. Combining this, (122), and Lemma 2.3(ii), we infer that $(1-\delta)\sum_{n\in\mathbb{N}^*} \tau_{n+1}\mu_n < +\infty$. In turn, since $(\tau_n)_{n\in\mathbb{N}^*}$ is increasing and $1-\delta > 0$, it follows that $\sum_{n\in\mathbb{N}^*} \tau_n\mu_n < +\infty$. Consequently, since $(\mu_n)_{n\in\mathbb{N}^*}$ is decreasing due to (i) and since clearly $\sum_{n\in\mathbb{N}^*} \tau_n = +\infty$, [1, Lemma 22] ensures that

$$h(x_n) - \min h = \mu_n = \mathrm{o}\left(\frac{1}{\sum_{k=1}^n \tau_k}\right) \text{ as } n \to +\infty, \tag{123}$$

which establishes (111). In turn, we deduce from (110), (123), and (i) that

$$0 \leqslant \tau_{n+1}^2 (h(x_{n+1}) - \min h) = (h(x_{n+1}) - \min h)\left(\tau_1^2 + \sum_{k=1}^n \big(\tau_{k+1}^2 - \tau_k^2\big)\right) \tag{124a}$$

$$\leqslant (h(x_{n+1}) - \min h)\left(\tau_1^2 + \delta \sum_{k=1}^n \tau_{k+1}\right) \tag{124b}$$

$$\leqslant (h(x_{n+1}) - \min h)\left(\tau_1^2 - \delta\tau_1 + \delta \sum_{k=1}^{n+1} \tau_k\right) \tag{124c}$$

$$\to 0 \text{ as } n \to +\infty, \tag{124d}$$

which verifies (112). ■

Remark 6.2 In Theorem 6.1, the assumption that $\tau_\infty = +\infty$ is actually not needed in items (i) and (iv)–(vii). For clarity, let us sketch the proof of (i) under the assumption that $\tau_\infty < +\infty$. Assume that $\tau_\infty < +\infty$. We infer from the first inequality in (115) that

$$(\forall n \in \mathbb{N}^*) \quad \tau_{n+1}^2 \mu_{n+1} + (2\gamma)^{-1}\|u_{n+1}\|^2 \leqslant \tau_n^2 \mu_n + (2\gamma)^{-1}\|u_n\|^2 - \big(\tau_n^2 - \tau_{n+1}^2 + \tau_{n+1}\big)\mu_n \tag{125}$$

and it follows from Lemma 2.3(ii) that $\sum_{n\in\mathbb{N}^*}(\tau_n^2 - \tau_{n+1}^2 + \tau_{n+1})\mu_n < +\infty$. One may argue similarly to the case (b) in the proof of Theorem 5.3 to obtain $\underline{\lim}\, \mu_n = 0$ or, equivalently, $\underline{\lim}\, h(x_n) = h(w)$, which contradicts (113). Therefore $\inf_{n\in\mathbb{N}^*} h(x_n) = \inf h$ and we get $h(x_n) \downarrow \inf_{n\in\mathbb{N}^*} h(x_n) = \inf h$. Items (iv) and (v) follow from this. In addition, note that we did not use the assumption that $\tau_\infty = +\infty$ in the proof of (vi) and (vii). It is, however, worth pointing out that the conclusion of Theorem 6.1(vi) is not so interesting when $\tau_\infty < +\infty$.

7 Open problems

We conclude this paper with a few open problems.

P1 In Theorem 5.3, is it true that $h(x_n) \to \inf h$?

P2 What can be said about the conclusions of Theorem 5.5(iii)&(vii)(b) if $\sup_{n\in\mathbb{N}^*}(n/\tau_n) = +\infty$?

P3 Suppose that $\operatorname{Argmin} h \neq \varnothing$. Do the sequences generated by Algorithm 5.1 and (108) always converge weakly to a point in $\operatorname{Argmin} h$?

Acknowledgments

We thank two referees for their very careful reading and constructive comments. HHB and XW were partially supported by NSERC Discovery Grants while MNB was partially supported by a Mitacs Globalink Graduate Fellowship Award.

References

[1] H. Attouch and A. Cabot, *Convergence rates of inertial forward-backward algorithms*, SIAM J. Optim., 28 (2018), pp. 849–874.

[2] H. Attouch, A. Cabot, Z. Chbani, and H. Riahi, *Inertial forward–backward algorithms with perturbations: Application to Tikhonov regularization*, J. Optim. Theory Appl., 179 (2018), pp. 1–36.

[3] H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont, *Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity*, Math. Program. Ser. B, 168 (2018), pp. 123–175.

[4] H. Attouch, Z. Chbani, and H. Riahi, *Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leqslant 3$*, ESAIM Control Optim. Calc. Var., 25 (2019). https://doi.org/10.1051/cocv/2017083.

[5] H. Attouch and J. Peypouquet, *The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$*, SIAM J. Optim., 26 (2016), pp. 1824–1834.

[6] H. Attouch, J. Peypouquet, and P. Redont, *Fast convergence of an inertial gradient-like system with vanishing viscosity*. https://arxiv.org/abs/1507.04782, 2015.

[7] J.-F. Aujol and C. Dossal, *Stability of over-relaxations for the forward-backward algorithm, application to FISTA*, SIAM J. Optim., 25 (2015), pp. 2408–2433.

[8] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, second ed., 2017.

[9] A. Beck, *First-Order Methods in Optimization*, SIAM, 2017.

[10] A. Beck and M. Teboulle, *Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems*, IEEE Trans. Image Process., 18 (2009), pp. 2419–2434.

[11] ——, *A fast iterative shrinkage-thresholding algorithm for linear inverse problem*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[12] J. Y. Bello Cruz and T. T. A. Nghia, *On the convergence of the forward-backward splitting method with linesearches*, Optim. Methods Softw., 31 (2016), pp. 1209–1238.

[13] K. Bredies, *A forward-backward splitting algorithm for the minimization of non-smooth convex functionals in Banach space*, Inverse Problems, 25 (2009), pp. 015005, 20.

[14] R. E. Bruck and S. Reich, *Nonexpansive projections and resolvents of accretive operators in Banach spaces*, Houston J. Math., 3 (1977), pp. 459–470.

[15] A. Chambolle and C. Dossal, *On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm"*, J. Optim. Theory Appl., 166 (2015), pp. 968–982.

[16] A. Chambolle and T. Pock, *An introduction to continuous optimization for imaging*, Acta Numer., 25 (2016), pp. 161–319.

[17] P. L. Combettes, *Quasi-Fejérian analysis of some optimization algorithms*, in Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications, D. Butnariu, Y. Censor, and S. Reich, eds., vol. 8, North-Holland, Amsterdam, 2001, pp. 115–152.

[18] P. L. Combettes and L. E. Glaudin, *Quasinonexpansive iterations on the affine hull of orbits: From Mann's mean value algorithm to inertial methods*, SIAM J. Optim., 27 (2017), pp. 2356–2380.

[19] P. L. Combettes, S. Salzo, and S. Villa, *Consistent learning by composite proximal thresholding*, Math. Program. Ser. B, 167 (2018), pp. 99–127.

[20] W. J. KACZOR AND M. T. NOWAK, *Problems in Mathematical Analysis. I. Real Numbers, Sequences and Series*, American Mathematical Society, Providence, RI, 2000.

[21] W. M. MOURSI, *The forward–backward algorithm and the normal problem*, J. Optim. Theory Appl., 176 (2018), pp. 605–624.

[22] Y. E. NESTEROV, *A method for solving the convex programming problem with convergence rate* $O(1/k^2)$, Dokl. Akad. Nauk, 269 (1983), pp. 543–547.

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

[24] A. PAZY, *Asymptotic behavior of contractions in Hilbert space*, Israel J. Math., 9 (1971), pp. 235–240.

[25] T.-L. RĂDULESCU, V. D. RĂDULESCU, AND T. ANDREESCU, *Problems in Real Analysis: Advanced Calculus on the Real Axis*, Springer, New York, 2009.

[26] M. SCHMIDT, N. L. ROUX, AND F. BACH, *Convergence rates of inexact proximal-gradient methods for convex optimization*, in Advances in Neural Information Processing Systems 24, J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, eds., Curran Associates, Inc., 2011, pp. 1458–1466.

[27] W. SU, S. BOYD, AND E. J. CANDÈS, *A differential equation for modeling Nesterov's accelerated gradient method: theory and insights*, J. Mach. Learn. Res., 17 (2016), pp. 1–43.

[28] S. VILLA, S. SALZO, L. BALDASSARRE, AND A. VERRI, *Accelerated and inexact forward-backward algorithms*, SIAM J. Optim., 23 (2013), pp. 1607–1633.

Appendices

Appendix A

For the sake of completeness, we provide the following proof of Lemma 2.1 based on [20, Problem 3.2.43].

Proof of Lemma 2.1. Because $(\forall n \in \mathbb{N}^*)\ 1-(\tau_n-1)^2/\tau_{n+1}^2 \geqslant 1-\tau_n^2/\tau_{n+1}^2$ due to the assumption that $\{\tau_n\}_{n\in\mathbb{N}^*} \subseteq [1,+\infty[$, it is sufficient to establish that

$$\sum_{n\in\mathbb{N}^*}\left(1-\frac{\tau_n^2}{\tau_{n+1}^2}\right)=+\infty. \tag{126}$$

Indeed, since $\tau_n \to +\infty$, there exists $N \in \mathbb{N}^*$ such that

$$(\forall n \geqslant N)\quad \tau_n^2 \geqslant 2\tau_1^2. \tag{127}$$

Now, set $(\forall n \in \mathbb{N}^*)\ \xi_n := \tau_{n+1}^2-\tau_n^2$, and $(\forall n \in \mathbb{N}^*)\ \sigma_n := \sum_{k=1}^n \xi_k$. Then, on the one hand, since $(\tau_n)_{n\in\mathbb{N}^*}$ is increasing and positive, we have $(\forall n \in \mathbb{N}^*)\ \xi_n = \tau_{n+1}^2-\tau_n^2 \geqslant 0$, and $(\sigma_n)_{n\in\mathbb{N}^*}$ is therefore an increasing sequence in $\mathbb{R}_+$; moreover, due to (127), $(\forall n \geqslant N)\ \sigma_n = \sum_{k=1}^n(\tau_{k+1}^2-\tau_k^2) = \tau_{n+1}^2-\tau_1^2 \geqslant \tau_1^2 \geqslant 1$. On the other hand, because $\tau_n \to +\infty$, we have $\sigma_n = \tau_{n+1}^2-\tau_1^2 \to +\infty$. Altogether, since

$$(\forall n \geqslant N)(\forall p \in \mathbb{N}^*)\quad \sum_{k=1}^{p}\frac{\xi_{n+k}}{\sigma_{n+k}} \geqslant \sum_{k=1}^{p}\frac{\xi_{n+k}}{\sigma_{n+p}} = \frac{\sigma_{n+p}-\sigma_n}{\sigma_{n+p}} = 1-\frac{\sigma_n}{\sigma_{n+p}} \tag{128}$$

by the fact that $(\sigma_n)_{n\geqslant N}$ is increasing, we see that $(\forall n \geqslant N)\ \underline{\lim}_p \sum_{k=1}^p(\xi_{n+k}/\sigma_{n+k}) \geqslant 1$. It follows that the partial sums of $\sum_{n\geqslant N}(\xi_n/\sigma_n)$ do not satisfy the Cauchy property. Hence, since $(\forall n \geqslant N)\ \xi_n/\sigma_n \geqslant 0$ and $\sigma_n = \tau_{n+1}^2-\tau_1^2$, we obtain

$$\sum_{n\geqslant N}\frac{\tau_{n+1}^2-\tau_n^2}{\tau_{n+1}^2-\tau_1^2} = \sum_{n\geqslant N}\frac{\xi_n}{\sigma_n} = +\infty. \tag{129}$$

Consequently, in the light of (127),

$$\sum_{n\geqslant N}\left(1-\frac{\tau_n^2}{\tau_{n+1}^2}\right)=\sum_{n\geqslant N}\frac{\tau_{n+1}^2-\tau_n^2}{\tau_{n+1}^2}\geqslant\sum_{n\geqslant N}\frac{\tau_{n+1}^2-\tau_n^2}{2\big(\tau_{n+1}^2-\tau_1^2\big)}=+\infty, \tag{130}$$

and (126) follows. ■

Appendix B

Proof of Lemma 2.2. Let us argue by contradiction. Towards this goal, assume that $\underline{\lim}\,\beta_n\in\,]0,+\infty]$ and fix $\beta\in\,]0,\underline{\lim}\,\beta_n[$. Then, there exists $N\in\mathbb{N}^*$ such that $(\forall n\geqslant N)\ \beta_n\geqslant\beta$, and hence, because $\{\alpha_n\}_{n\in\mathbb{N}^*}\subseteq\mathbb{R}_+$, we have $(\forall n\geqslant N)\ \alpha_n\beta_n\geqslant\beta\alpha_n$. Consequently, since $\sum_{n\in\mathbb{N}^*}\alpha_n=+\infty$, it follows that $\sum_{n\geqslant N}\alpha_n\beta_n\geqslant\sum_{n\geqslant N}\beta\alpha_n=+\infty$, which violates our assumption. To sum up, $\underline{\lim}\,\beta_n=0$. ■

Appendix C

The following self-contained proof of Lemma 2.3 follows [17, Lemma 3.1] in the case $\chi=1$; however, we do not require the error sequence $(\varepsilon_n)_{n\in\mathbb{N}^*}$ to be positive.

Proof of Lemma 2.3. (i): Set $\alpha:=\underline{\lim}_n\,\alpha_n\in[\inf_{n\in\mathbb{N}^*}\alpha_n,+\infty]$ and let $(\alpha_{k_n})_{n\in\mathbb{N}^*}$ be a subsequence of $(\alpha_n)_{n\in\mathbb{N}^*}$ that converges to α. We first show that $\alpha<+\infty$. Since $\{\beta_n\}_{n\in\mathbb{N}^*}\subseteq\mathbb{R}_+$, it follows from (9) that $(\forall n\in\mathbb{N}^*)\ \alpha_{n+1}-\alpha_n\leqslant\varepsilon_n$. Thus, $(\forall n\geqslant 2)\ \alpha_n=\alpha_1+\sum_{k=1}^{n-1}(\alpha_{k+1}-\alpha_k)\leqslant\alpha_1+\sum_{k=1}^{n-1}\varepsilon_k$; in particular, $(\forall n\geqslant 2)\ \alpha_{k_n}\leqslant\alpha_1+\sum_{k=1}^{k_n-1}\varepsilon_k$. Hence, since $\alpha_{k_n}\to\alpha$ and $\sum_{n\in\mathbb{N}^*}\varepsilon_n$ converges, it follows that $\alpha\leqslant\alpha_1+\sum_{k\in\mathbb{N}}\varepsilon_k<+\infty$, as claimed. In turn, to establish the convergence of $(\alpha_n)_{n\in\mathbb{N}^*}$, it suffices to verify that $\overline{\lim}_n\,\alpha_n\leqslant\underline{\lim}_n\,\alpha_n$. Towards this goal, let δ be in $]0,+\infty[$. Then, on the one hand, Cauchy's criterion ensures the existence of $k_{n_0}\in\mathbb{N}^*$ such that $\alpha_{k_{n_0}}-\alpha\leqslant\delta/2$ and that $(\forall n\geqslant k_{n_0})(\forall m\in\mathbb{N}^*)\ \sum_{k=n}^{n+m}\varepsilon_k\leqslant\delta/2$. On the other hand, because $\{\beta_n\}_{n\in\mathbb{N}^*}\subseteq\mathbb{R}_+$, (9) implies that $(\forall n\geqslant k_{n_0}+1)\ \alpha_n-\alpha_{k_{n_0}}=\sum_{k=k_{n_0}}^{n-1}(\alpha_{k+1}-\alpha_k)\leqslant\sum_{k=k_{n_0}}^{n-1}\varepsilon_k$. Altogether, $(\forall n\geqslant k_{n_0}+1)\ \alpha_n\leqslant\alpha_{k_{n_0}}+\sum_{k=k_{n_0}}^{n-1}\varepsilon_k\leqslant(\alpha+\delta/2)+\delta/2=\alpha+\delta$, from which we deduce that $\overline{\lim}_n\,\alpha_n\leqslant\alpha+\delta$. Consequently, since δ is arbitrarily chosen in $]0,+\infty[$, it follows that $\overline{\lim}_n\,\alpha_n\leqslant\alpha=\underline{\lim}_n\,\alpha_n$, and therefore, $(\alpha_n)_{n\in\mathbb{N}}$ converges to α.

(ii): We derive from (9) that $(\forall N\in\mathbb{N}^*)\ \sum_{n=1}^N\beta_n\leqslant\sum_{n=1}^N(\alpha_n-\alpha_{n+1})+\sum_{n=1}^N\varepsilon_n=\alpha_1-\alpha_{N+1}+\sum_{n=1}^N\varepsilon_n$. Hence, since $\sum_{n\in\mathbb{N}^*}\varepsilon_n$ is convergent and, by (i), $\lim_n\alpha_n=\alpha$, letting $N\to+\infty$ yields $\sum_{n\in\mathbb{N}}\beta_n\leqslant\alpha_1-\alpha+\sum_{n\in\mathbb{N}}\varepsilon_n<+\infty$, and so $\sum_{n\in\mathbb{N}^*}\beta_n<+\infty$, as required. ■

Appendix D

Proof of Lemma 2.4. Indeed, since $(\forall n\in\mathbb{N}^*)$

$$\sum_{k=1}^n k(\alpha_k-\alpha_{k+1})=\sum_{k=1}^n\big(k\alpha_k-(k+1)\alpha_{k+1}+\alpha_{k+1}\big) \tag{131a}$$

$$=\sum_{k=1}^n\big(k\alpha_k-(k+1)\alpha_{k+1}\big)+\sum_{k=1}^n\alpha_{k+1}=\alpha_1-(n+1)\alpha_{n+1}+\sum_{k=1}^n\alpha_{k+1}, \tag{131b}$$

we readily obtain the conclusion. ■

Appendix E

Proof of Lemma 2.5. "⇒": Since $(\alpha_n)_{n\in\mathbb{N}^*}$ is a decreasing sequence in $\mathbb{R}_+$ and $\sum_{n\in\mathbb{N}^*}\alpha_n < +\infty$, it follows that $n\alpha_n \to 0$ (see, e.g., [20, Problem 3.2.35]). Invoking the assumption that $\sum_{n\in\mathbb{N}^*}\alpha_n < +\infty$ once more, we infer from Lemma 2.4 that $\sum_{n\in\mathbb{N}^*} n(\alpha_n - \alpha_{n+1}) < +\infty$, as desired.

"⇐": A consequence of Lemma 2.4. ∎

Appendix F

Proof of Lemma 3.1. This is similar to the one found in [11, Lemma 2.3] and included for completeness; see also [15, Lemma 3.1]. Fix $(x,y) \in \mathcal{H}\times\mathcal{H}$. On the one hand, by **(A1)** and **(A3)** in Assumption 1.1, ∇f is Lipschitz continuous with constant γ^{-1}, from which, the Descent Lemma (see, e.g., [8, Lemma 2.64]), and the convexity of f we infer that

$$f(Ty) \leqslant f(y) + \langle \nabla f(y) \mid Ty - y\rangle + (2\gamma)^{-1}\|Ty-y\|^2 \tag{132a}$$
$$= f(y) + \langle \nabla f(y) \mid x - y\rangle + \langle \nabla f(y) \mid Ty - x\rangle + (2\gamma)^{-1}\|Ty-y\|^2 \tag{132b}$$
$$\leqslant f(x) + \langle \nabla f(y) \mid Ty - x\rangle + (2\gamma)^{-1}\|Ty-y\|^2. \tag{132c}$$

On the other hand, because $Ty = \operatorname{Prox}_{\gamma g}(y - \gamma\nabla f(y))$, [8, Proposition 12.26] asserts that

$$g(Ty) \leqslant g(x) - \gamma^{-1}\langle (y-\gamma\nabla f(y)) - Ty \mid x - Ty\rangle \tag{133a}$$
$$\leqslant g(x) + \langle \gamma^{-1}(y - Ty) - \nabla f(y) \mid Ty - x\rangle. \tag{133b}$$

Altogether, upon adding (132) and (133), it follows that

$$h(Ty) \leqslant h(x) + \gamma^{-1}\langle y - Ty \mid Ty - x\rangle + (2\gamma)^{-1}\|Ty-y\|^2 \tag{134a}$$
$$= h(x) + \gamma^{-1}\langle y - Ty \mid y - x\rangle + \gamma^{-1}\langle y - Ty \mid Ty - y\rangle + (2\gamma)^{-1}\|Ty-y\|^2 \tag{134b}$$
$$= h(x) + \gamma^{-1}\langle y - Ty \mid y - x\rangle - (2\gamma)^{-1}\|Ty-y\|^2, \tag{134c}$$

which yields (18). ∎

Appendix G

Proof of (50). Recall that $\lim(\tau_n/n) = 1/2$. In turn, because $(\forall n\in\mathbb{N}^*)\ \tau_n^2 = \tau_{n+1}^2 - \tau_{n+1}$, it follows that

$$\frac{n(\tau_n - \tau_{n+1})}{\tau_{n+1}} = \frac{n(\tau_n^2 - \tau_{n+1}^2)}{\tau_{n+1}(\tau_n + \tau_{n+1})} = \frac{-n\tau_{n+1}}{\tau_{n+1}(\tau_n+\tau_{n+1})} = \frac{-1}{\dfrac{\tau_n}{n} + \dfrac{\tau_{n+1}}{n+1}\dfrac{n+1}{n}} \to \frac{-1}{\frac{1}{2}+\frac{1}{2}} = -1 \tag{135}$$

and therefore that

$$n\left(\frac{\tau_n - 1}{\tau_{n+1}} - 1 + \frac{3}{n}\right) = \frac{n(\tau_n - \tau_{n+1})}{\tau_{n+1}} - \frac{n+1}{\tau_{n+1}}\frac{n}{n+1} + 3 \to -1 - 2 + 3 = 0. \tag{136}$$

Hence, (50) holds. ∎