

Convergence rate of inertial Forward–Backward algorithm beyond Nesterov’s rule

Vassilis Apidopoulos¹  · Jean-François Aujol¹ · Charles Dossal¹

Received: 30 June 2017 / Accepted: 1 November 2018 / Published online: 12 November 2018

© Springer-Verlag GmbH Germany, part of Springer Nature and Mathematical Optimization Society 2018

Abstract

In this paper we study the convergence of an Inertial Forward–Backward algorithm, with a particular choice of an over-relaxation term. In particular we show that for a sequence of over-relaxation parameters, that do not satisfy Nesterov’s rule, one can still expect some relatively fast convergence properties for the objective function. In addition we complement this work by studying the convergence of the algorithm in the case where the proximal operator is inexactly computed with the presence of some errors and we give sufficient conditions over these errors in order to obtain some convergence properties for the objective function.

Keywords Convex optimization · Proximal operator · Inertial FB algorithm · Nesterov’s rule · Rate of convergence

Mathematics Subject Classification 49M20 · 46N10 · 90C25 · 65K10

This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR GOTMI) (ANR-16-CE33-0010-01). Jean-François Aujol is a member of Institut Universitaire de France (IUF).

✉ Vassilis Apidopoulos
vasileios.apidopoulos@math.u-bordeaux.fr

Jean-François Aujol
jean-francois.aujol@math.u-bordeaux.fr

Charles Dossal
charles.dossal@math.u-bordeaux.fr

¹ IMB, UMR 5251, Université de Bordeaux, 33400 Talence, France

1 Introduction

In a Hilbert space \mathcal{H} (possibly infinite dimensional). We are interested in solving the following minimization problem:

$$\min_{x \in \mathcal{H}} F(x) \quad (\text{M})$$

where $F = f + g : \mathcal{H} \longrightarrow \bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$, with:

- H.1 f a convex function in $\mathscr{C}^{1,1}(\mathcal{H})$ with L -Lipschitz gradient.
- H.2 g a convex, lower semi-continuous and proper function (possibly non-smooth).
- H.3 The set of minimizers $X^* = \operatorname{argmin}\{F\}$ is non-empty.

In what follows we will consider these conditions as guaranteed and we will denote $x^* \in \mathcal{H}$ a minimizer of F .

In order to solve the problem (M), several algorithms have been proposed based on the use of the proximal operator due to the non differentiable part. One of the basics is the Forward–Backward algorithm (FB), which consists in applying iteratively at every point the non-expansive operator $T_\gamma : \mathcal{H} \longrightarrow \mathcal{H}$, defined as:

$$T_\gamma(x) = \operatorname{prox}_{\gamma g}(x - \gamma \nabla f(x)) \quad \forall x \in \mathcal{H}$$

where $\operatorname{prox}_{\gamma g}$ designs the proximal operator of g .¹

Formally, for $x_0 \in \mathcal{H}$ and $0 < \gamma \leq \frac{2}{L}$, for all $n \geq 1$ the FB-algorithm reads:

$$\begin{aligned} x_{n+1} &= T_\gamma(x_n) \\ &= \operatorname{prox}_{\gamma g}(x_n - \gamma \nabla f(x_n)) \end{aligned} \quad (\text{FB})$$

The FB turns out to be a descent algorithm in terms of the values of the function (i.e. $F(x_{n+1}) \leq F(x_n)$), with a sublinear rate of convergence of order 1 for the objective function, i.e. $\forall n \geq 1, F(x_n) - F(x^*) \leq \frac{C}{\gamma^n}$, where $C > 0$ is a positive constant. It was also proven that the generated iterates weakly converge to a minimizer x^* .

In the seminal work of Nesterov in [16] (see also [17]), it was shown that considering an inertial version of the Forward–Backward algorithm can lead to some significant fast convergence properties for the trajectories generated. These ideas are further developed in the semi-differential case (where g is not necessarily differentiable) in [12] and notably in [9]. The basic scheme of these inertial Forward–Backward algorithms (i-FB) is the following:

Given $x_0 = y_0 \in \mathcal{H}$, $\{a_n\}_{n \in \mathbb{N}}$ a positive sequence, such that $a_n \nearrow 1$ and $0 < \gamma < \frac{2}{L}$, for all $n \geq 1$, define

$$\begin{aligned} x_n &= T_\gamma(y_{n-1}) \\ &= \operatorname{prox}_{\gamma g}(y_{n-1} - \gamma \nabla f(y_{n-1})) \\ y_n &= x_n + a_n(x_n - x_{n-1}) \end{aligned} \quad (1.1)$$

¹ For a definition of the prox operator, see (2.1) in the next section.

In particular i-FB algorithm consists in applying the non-expansive operator T_γ in an extrapolation of the previous iterate with a momentum term ($a_n(x_n - x_{n-1})$), that can boost up the velocity for a suitable choice of $\{a_n\}_{n \in \mathbb{N}}$.

There is a vast literature concerning the study of this type of inertial FB algorithms sometimes called as FISTA (to name but a few, we address the reader to [2,3,9,10,14–17,20]). The basic idea behind the sequence $\{a_n\}_{n \in \mathbb{N}}$ is that it can be written as $a_n = \frac{s_n - 1}{s_{n+1}}$, where $\{s_n\}_{n \in \mathbb{N}}$, is a sequence that verifies Nesterov’s rule, i.e.:

$$s_n^2 + s_{n+1} - s_{n+1}^2 \geq 0 \quad \forall n \in \mathbb{N} \quad (\text{NR})$$

It has been shown (see for example [9,10,16,17]) that if the relation (NR) hold true, then one can obtain a better convergence rate towards the minimum, than the classical FB algorithm. More precisely $F(x_n) - F(x^*)$ has a sublinear rate of convergence of order 2, i.e. $F(x_n) - F(x^*) \leq \frac{C}{\gamma n^2}$, $\forall n \geq 1$ where $C > 0$ is a positive constant.

A choice of a particular interest for the sequence $\{a_n\}_{n \in \mathbb{N}}$, is when $a_n = \frac{n}{n+b}$ for all $n \geq 1$ (this corresponds to $s_n = \frac{n+b-1}{b-1}$), where $b > 1$. With this choice, Nesterov’s rule (NR) is equivalent to considering $b \geq 3$. Nevertheless, in [10] (see also [3]) the authors show that by assuming that $b > 3$, one can additionally expect the weak convergence of the iterates $\{x_n\}_{n \in \mathbb{N}}$ generated by the i-FB algorithm (1.1), to a minimizer x^* of F . In addition in [5] the authors show that by taking $b > 3$ can asymptotically improve the rate of convergence of $F(x_n) - F(x^*)$ which is actually $o(n^{-2})$.

Of course, other choices for the over-relaxation sequence $\{a_n\}_{n \in \mathbb{N}}$ are possible, depending also on the geometry of the minimizing function F . Unfortunately a detailed study of the i-FB algorithm with a general sequence of parameters $\{a_n\}_{n \in \mathbb{N}}$ goes beyond the scope of this paper. For this issue we address the reader to [2], where the authors study different conditions over the sequence $\{a_n\}_{n \in \mathbb{N}}$, in order to obtain several convergence properties for the i-FB algorithm.

In this paper we study the case of i-FB algorithm where $a_n = \frac{n}{n+b}$ for all $n \geq 1$, with $b \in (0, 3)$, which means that the Nesterov’s rule (NR) is not satisfied.² In particular we deduce some relatively fast convergence rate for the objective function $F(x_n) - F(x^*)$, as also for the local variation of the iterates $\|x_n - x_{n-1}\|$. The exact estimate bounds that we find for these quantities, for $b \in (0, 3)$, are the following (see Corollary 2):

$$F(x_n) - F(x^*) \leq \frac{C}{(n+b-1)^{\frac{2b}{3}}} \quad \text{and} \quad \|x_n - x_{n-1}\| \leq \frac{C}{(n+b-1)^{\frac{b}{3}}} \quad (1.2)$$

for all $n \geq 1$, where $C > 0$ is a positive constant.

In addition we deduce “almost” the same convergence properties, in the perturbed case, where every iterate is inexactly calculated by some error parameters, under some control hypotheses over these errors (see Corollary 3 and Remark 2).

² To this issue, we address the reader to Remark 3 at the end of this document.

This work consists a discrete counterpart of the analysis made for the study of the following differential equation

$$\ddot{x}(t) + \frac{b}{t}\dot{x}(t) + \nabla F(x(t)) = 0 \quad (1.3)$$

made recently in [4, 7] in the differential setting (where F is differentiable) and in [1] in the non-differential setting. For a general presentation of the connexion between the differential equation (1.3) and the i-FB algorithm (1.1) we address the reader to [3] and [20]. These works will provide us a useful guide for our study.

The paper is organized as follows. In Sect. 2, we give the basic definitions and tools necessary for our analysis. In the third section we study the convergence rates for a special choice of over-relaxation terms for the i-FB algorithm. Finally in Sect. 4 we present the same type of analysis for the inexact i-FB algorithm where every new iterate of the algorithm is inexactly calculated with the presence of some errors.

2 Definitions and basic notions

Given a function $G : \mathcal{H} \rightarrow \mathbb{R}$, we define its subdifferential, as the multi-valued operator $\partial G : \mathcal{H} \rightarrow 2^{\mathcal{H}}$, such that for all $x \in \mathcal{H}$:

$$\partial G(x) = \{z \in \mathcal{H} : \forall y \in \mathcal{H}, G(x) \leq G(y) + \langle z, x - y \rangle\}$$

We also recall the definition of the proximal operator which is the basic tool for i-FB algorithm. If $F : \mathcal{H} \rightarrow \bar{\mathbb{R}}$, is a lower semi-continuous, proper and convex function, the proximal operator of F is the operator $\text{prox}_F : \mathcal{H} \rightarrow \mathcal{H}$, such that:

$$\text{prox}_F(x) = \underset{y \in \mathcal{H}}{\operatorname{argmin}} \left\{ F(y) + \frac{\|x - y\|^2}{2} \right\}, \quad \forall x \in \mathcal{H} \quad (2.1)$$

Here we must point out that the proximal operator is well-defined, since by the hypothesis made on F , for every $x \in \mathcal{H}$, the strongly convex function $y \mapsto F(y) + \frac{\|x-y\|^2}{2}$, admits a unique minimizer.

Equivalently the proximal operator can be also seen as the resolvent of the maximal monotone operator ∂F , i.e. for all $x \in \mathcal{H}$ and γ a positive parameter we have that:

$$\text{prox}_{\gamma F}(x) = (Id + \gamma \partial F)^{-1}(x) \quad (2.2)$$

For a detailed study concerning the subdifferential and the proximal operator and their properties, we address the reader to [8].

3 Convergence analysis for the i-FB algorithm

In this section we present the results concerning the convergence analysis of the i-FB algorithm with a special choice of the over-relaxation terms.

Firstly we recall the i-FB algorithm as the one considered in [10]:

Algorithm 1 i-FB

Let $0 < \gamma < \frac{1}{L}$ and $b \in (0, 3)$. We consider the sequences $\{a_n\}_{n \in \mathbb{N}^*}$, $\{x_n\}_{n \in \mathbb{N}}$, $\{y_n\}_{n \in \mathbb{N}}$, such that $x_0 = y_0 \in \mathcal{H}$, and for every $n \in \mathbb{N}^*$ we set:

$$x_n = T(y_{n-1}) \quad (3.1)$$

$$y_n = x_n + a_n(x_n - x_{n-1}) \quad \text{with} \quad a_n = \frac{n}{n+b} \quad (3.2)$$

$$\text{where } T(x) = \text{Prox}_{\gamma g}(x - \gamma \nabla f(x)) \quad (3.3)$$

We also consider the following sequences: $\{t_n\}_{n \in \mathbb{N}^*}$, $\{\delta_n\}_{n \in \mathbb{N}^*}$, $\{w_n\}_{n \in \mathbb{N}}$ and $\{E_n\}_{n \in \mathbb{N}}$ such that:

$$t_n = n + b - 1 \quad (3.4)$$

$$\delta_n = \|x_n - x_{n-1}\|^2 \quad (3.5)$$

$$w_n = F(x_n) - F(x^*) \quad (3.6)$$

We remark that by definition of t_n it follows that $a_n = \frac{t_n - (b-1)}{t_{n+1}}$.

In addition for $\lambda > 0$ and $\xi > 0$, we define the sequences $\{v_n\}_{n \in \mathbb{N}}$ and $\{E_n\}_{n \in \mathbb{N}^*}$, such that for all $n \geq 1$:

$$v_n = \|\lambda(x_{n-1} - x^*) + t_n(x_n - x_{n-1})\|^2 \quad (3.7)$$

$$E_n = t_n^2 w_n + \frac{1}{2\gamma} \underbrace{\|\lambda(x_{n-1} - x^*) + t_n(x_n - x_{n-1})\|^2}_{=v_n} + \frac{\xi}{2\gamma} \|x_{n-1} - x^*\|^2 \quad (3.8)$$

The sequence $\{E_n\}_{n \in \mathbb{N}}$ was used implicitly for the study of the i-FB algorithm in numerous articles (see for example, [2,3,9,10,20]). Although it was introduced explicitly in [3,20], as an energy function associated to the dynamical system (1.3), corresponding to i-FB algorithm.

In these works it has been shown that for higher values of parameter b which are greater than 3, the sequence $\{E_n\}_{n \in \mathbb{N}}$ is non-increasing. This leads to some fast convergence properties of the sequence $\{w_n\}_{n \in \mathbb{N}}$ (which is of inverse quadratic order), as also to convergence of the generated sequence $\{x_n\}_{n \in \mathbb{N}}$ to a minimizer when $b > 3$ (see for example [3,10,20]). As also pointed out, the choice of $b = 3$ (which corresponds to the Nesterov's accelerated algorithm) seems critical for this non-increasing property of $\{E_n\}_{n \in \mathbb{N}}$. More precisely, concerning the i-FB algorithm, the following Theorem holds:

Theorem 1 (Su et al. [20], Attouch et al. [3]) Let $0 < \gamma \leq \frac{1}{L}$, $b \geq 3$ and $\{x_n\}_{n \in \mathbb{N}}$ the sequence generated by i-FB. Then for $\lambda = b - 1$ and $\xi = 0$ the sequence $\{E_n\}_{n \in \mathbb{N}}$ is non-increasing.

Corollary 1 Let $0 < \gamma \leq \frac{1}{L}$, $b \geq 3$ and $\{x_n\}_{n \in \mathbb{N}}$ the sequence generated by i-FB. Then there exists a constant $C > 0$ such that for all $n \geq 1$, it holds:

$$w_n \leq \frac{C}{(n + b - 1)^2} \quad (3.9)$$

Our study focus on the convergence rates of the objective sequence $\{w_n\}_{n \in \mathbb{N}}$ for small values of b . In particular we show that for $b \in (0, 3)$, one can still obtain some relatively fast convergence rate for $\{w_n\}_{n \in \mathbb{N}}$ despite the fact that the energy-sequence $\{E_n\}_{n \in \mathbb{N}}$ is not necessarily non-increasing. We now present the main result of this paper.

Theorem 2 Let $0 < \gamma \leq \frac{1}{L}$, $b \in (0, 3)$ and $\{x_n\}_{n \in \mathbb{N}}$ the sequence generated by i-FB. Then for $\lambda = \frac{2b}{3}$ and $\xi = \frac{4b^2}{9}$, there exists a constant $C > 0$, such that for all $n \geq 1$, it holds:

$$E_n \leq C(n + b - 1)^{\frac{2(3-b)}{3}} \quad (3.10)$$

Corollary 2 Under the hypotheses of Theorem 2, there exists a positive constant $C > 0$, such that for all $n \geq 1$, we have:

$$F(x_n) - F(x^*) \leq \frac{C}{(n + b - 1)^{\frac{2b}{3}}} \quad (3.11)$$

If in addition the sequence $\{x_n\}_{n \in \mathbb{N}}$ is bounded, then there exists a positive constant $C' > 0$, such that for all $n \geq 1$, we have:

$$\|x_n - x_{n-1}\| \leq \frac{C'}{(n + b - 1)^{\frac{b}{3}}} \quad (3.12)$$

Remark 1 If we suppose that F is also coercive, then the boundedness of the sequence $\{F(x_n)\}_{n \in \mathbb{N}}$, implies the boundedness of $\{x_n\}_{n \in \mathbb{N}}$. In this case the estimate (3.12) comes directly from (3.11).

Proof (Corollary 2) By using the estimation (3.10) of Theorem 2 and the fact that the sequence $\{E_n\}_{n \geq 1}$ is a sum of positive terms we deduce that:

$$(n + b - 1)^2 (F(x_n) - F(x^*)) \leq C(n + b - 1)^{2 - \frac{2b}{3}} \quad (3.13)$$

which gives the first estimate (3.11) of Corollary 2.

For the second point using (3.10) and the fact that the sequence $\{E_n\}_{n \geq 1}$ is a sum of positive terms we find:

$$\left\| \frac{2b}{3}(x_{n-1} - x^*) + (n + b - 1)(x_n - x_{n-1}) \right\|^2 \leq C(n + b - 1)^{2 - \frac{2b}{3}} \quad (3.14)$$

By taking the square roots in both sides and using the triangle inequality we deduce:

$$(n+b-1)\|x_n - x_{n-1}\| \leq \sqrt{C}(n+b-1)^{1-\frac{b}{3}} + \frac{2b}{3} \sup_{n \geq 1} \{\|x_{n-1} - x^*\|\} \quad (3.15)$$

Hence if the trajectory $\{x_n\}_{n \in \mathbb{N}}$ is bounded, we conclude the second point of Corollary 2. \square

The strategy of the proof of Theorem (2) is the following. Firstly we study the local variation of the sequence $\{E_n\}_{n \in \mathbb{N}}$ (i.e. the difference $E_{n+1} - E_n$). Using some Lyapunov-type analysis, for some suitable choices of parameters $\lambda > 0$ and $\xi > 0$, we are able to control the growth of $\{E_n\}_{n \in \mathbb{N}}$ up to a suitable order. Once this control-estimate is proven, an application of a discrete version of Gronwall's lemma (see Lemma 5 in Appendix) will provide the bound for the sequence $\{E_n\}_{n \in \mathbb{N}}$ as given in Theorem 2.

For the proof of Theorem 2 we also make use of the following lemma (see Lemma 2.2 in [9] or Lemma 1, in [10]):

Lemma 1 *For any $y \in \mathcal{H}$ and $0 < \gamma \leq \frac{1}{L}$ we have that for every $x \in \mathcal{H}$:*

$$2\gamma(F(x) - F(T(y))) \geq \|T(y) - x\|^2 - \|y - x\|^2 \quad (3.16)$$

In order to prove the assertion of Theorem 2, we will use the next Lemma which shows the control-order of the growth of $\{E_n\}_{n \in \mathbb{N}}$.

Lemma 2 *Let $0 < \gamma \leq \frac{1}{L}$, $b \in (0, 3)$ $\lambda = \frac{2b}{3}$, $\xi = \frac{2b(3-b)}{9}$ and $\{x_n\}_{n \in \mathbb{N}}$ the sequence generated by i-FB. Then the following recursive formula holds for all $n \geq 1$:*

$$E_{n+1} - E_n \leq \left(\frac{a}{(n+b-1)^2} + \frac{c}{(n+b-1)} \right) E_n \quad (3.17)$$

where $a = \frac{(3-b)(3+b)}{9}$ and $c = \frac{2(3-b)}{3}$.

Due to the technical details of the proof of Lemma 2, we will first present a sketch of it in order to give a better insight.

1. We start by investigating the local variation of the sequence $\{E_n\}_{n \geq 1}$. By using Lemma 1 and performing some algebraic computations we obtain a relation of the following form:

$$2\gamma(E_{n+1} - E_n) \leq 2\gamma\alpha_{n,\lambda,\xi} w_n + \beta_{n,\lambda,\xi} \delta_n + \gamma_{n,\lambda,\xi} \langle x_n - x_{n-1}, x_{n-1} - x^* \rangle \quad (3.18)$$

At this point, in order to prove Theorem 1 it is sufficient to choose suitable values for λ and ξ in order to show that $\gamma_{n,\lambda,\xi} = 0$ and $\alpha_{n,\lambda,\xi}, \beta_{n,\lambda,\xi} \leq 0$ for all $n \geq 1$, under the supplementary hypothesis that $b \geq 3$.

2. Here instead we are interested in the case where $b \in (0, 3)$ and $\alpha_{n,\lambda,\xi}, \beta_{n,\lambda,\xi}$ are not necessarily non-positive for all $n \geq 1$. At this point we express E_n in function of w_n and δ_n and we find a relation of the form:

$$2\gamma(E_{n+1} - E_n) \leq 2\gamma \frac{c}{t_n} E_n + R_{n,\lambda,\xi} \quad (3.19)$$

3. Finally by some suitable values for λ and ξ we show that: $R_{n,\lambda,\xi} \leq 2\gamma \frac{a}{t_n^2} E_n$

We now pass to a detailed presentation of this proof.

Proof By applying Lemma (1) to $y = y_n$ and $x = \left(1 - \frac{\lambda}{t_{n+1}}\right)x_n + \frac{\lambda}{t_{n+1}}x^*$ we obtain (here $\lambda \in (0, 1+b)$):

$$\begin{aligned} 2\gamma & \left(F\left(\left(1 - \frac{\lambda}{t_{n+1}}\right)x_n + \frac{\lambda}{t_{n+1}}x^*\right) - F(x_{n+1}) \right) \\ & \geq \left\| x_{n+1} - x_n + \frac{\lambda}{t_{n+1}}(x_n - x^*) \right\|^2 - \left\| a_n(x_n - x_{n-1}) + \frac{\lambda}{t_{n+1}}(x_n - x^*) \right\|^2 \end{aligned} \quad (3.20)$$

By using the convexity of F we obtain:

$$\begin{aligned} 2\gamma & \left[\left(1 - \frac{\lambda}{t_{n+1}}\right)F(x_n) + \frac{\lambda}{t_{n+1}}F(x^*) - F(x_{n+1}) \right] \\ & \geq \left\| x_{n+1} - x_n + \frac{\lambda}{t_{n+1}}(x_n - x^*) \right\|^2 - \left\| a_n(x_n - x_{n-1}) + \frac{\lambda}{t_{n+1}}(x_n - x^*) \right\|^2 \end{aligned} \quad (3.21)$$

By adding and subtracting $2\gamma F(x^*)$ to the left-hand side and definition of w_n , we have:

$$\begin{aligned} 2\gamma & \left[\left(1 - \frac{\lambda}{t_{n+1}}\right)w_n - w_{n+1} \right] \geq \left\| x_{n+1} - x_n + \frac{\lambda}{t_{n+1}}(x_n - x^*) \right\|^2 \\ & \quad - \left\| a_n(x_n - x_{n-1}) + \frac{\lambda}{t_{n+1}}(x_n - x^*) \right\|^2 \end{aligned} \quad (3.22)$$

By multiplying both sides by t_{n+1}^2 , we obtain:

$$\begin{aligned} 2\gamma & \left((t_{n+1}^2 - \lambda t_{n+1})w_n - t_{n+1}^2 w_{n+1} \right) \geq \|t_{n+1}(x_{n+1} - x_n) + \lambda(x_n - x^*)\|^2 \\ & \quad - \|n(x_n - x_{n-1}) + \lambda(x_n - x^*)\|^2 \end{aligned} \quad (3.23)$$

By adding and subtracting $2\gamma t_n^2 w_n$ to the left-hand side we obtain:

$$\begin{aligned} 2\gamma \left(k_{n+1} w_n + t_n^2 w_n - t_{n+1}^2 w_{n+1} \right) &\geq \underbrace{\|t_{n+1}(x_{n+1} - x_n) + \lambda(x_n - x^*)\|^2}_{=v_{n+1}} \\ &\quad - \|n(x_n - x_{n-1}) + \lambda(x_n - x^*)\|^2 \end{aligned} \quad (3.24)$$

where

$$\begin{aligned} k_{n+1} &= t_{n+1}^2 - \lambda t_{n+1} - t_n^2 = (n+b)^2 - \lambda(n+b) - (n+b-1)^2 \\ &= n^2 + 2bn + b^2 - \lambda n - \lambda b - n^2 - 2(b-1)n - b^2 + 2b - 1 \\ &= (2-\lambda)(n+b) - 1 \end{aligned} \quad (3.25)$$

So that:

$$2\gamma(t_{n+1}^2 w_{n+1} - t_n^2 w_n) \leq 2\gamma k_{n+1} w_n + \|n(x_n - x_{n-1}) + \lambda(x_n - x^*)\|^2 - v_{n+1} \quad (3.26)$$

Hence by using the last inequality and the identity

$$\|u - z\|^2 - \|v - z\|^2 = \|u - v\|^2 + 2\langle u - v, v - z \rangle \quad \forall u, v, z \in \mathcal{H} \quad (\text{PI})$$

and the definition of E_n we have that:

$$\begin{aligned} 2\gamma(E_{n+1} - E_n) &= 2\gamma(t_{n+1}^2 w_{n+1} - t_n^2 w_n) + v_{n+1} - v_n + \xi(\|x_n - x^*\|^2 - \|x_{n-1} - x^*\|^2) \\ (3.26) \quad &\leq 2\gamma k_{n+1} w_n + \|n(x_n - x_{n-1}) + \lambda(x_n - x^*)\|^2 \\ &\quad - \|t_n(x_n - x_{n-1}) + \lambda(x_{n-1} - x^*)\|^2 + \xi(\|x_n - x^*\|^2 - \|x_{n-1} - x^*\|^2) \\ \left(\text{PI} \right) z = \lambda x^* &= 2\gamma k_{n+1} w_n + (\lambda + 1 - b)^2 \|x_n - x_{n-1}\|^2 \\ &\quad + 2(\lambda + 1 - b) \langle x_n - x_{n-1}, t_n(x_n - x_{n-1}) + \lambda(x_{n-1} - x^*) \rangle \\ \left(\text{PI} \right) z = x^* &+ \xi \|x_n - x_{n-1}\|^2 + 2\xi \langle x_n - x_{n-1}, x_{n-1} - x^* \rangle \\ &= 2\gamma k_{n+1} w_n + \left((\lambda + 1 - b)^2 + \xi + 2(\lambda + 1 - b)t_n \right) \|x_n - x_{n-1}\|^2 \\ &\quad + 2(\lambda(\lambda + 1 - b) + \xi) \langle x_n - x_{n-1}, x_{n-1} - x^* \rangle \end{aligned} \quad (3.27)$$

By definition of E_n we also have

$$2\gamma E_n = 2\gamma t_n^2 w_n + (\lambda^2 + \xi) \|x_{n-1} - x^*\|^2 + t_n^2 \|x_n - x_{n-1}\|^2 + 2\lambda t_n \langle x_n - x_{n-1}, x_{n-1} - x^* \rangle \quad (3.28)$$

so that

$$t_n \|x_n - x_{n-1}\|^2 = \frac{2\gamma}{t_n} E_n - 2\gamma t_n w_n - \frac{(\lambda^2 + \xi)}{t_n} \|x_{n-1} - x^*\|^2 - 2\lambda \langle x_n - x_{n-1}, x_{n-1} - x^* \rangle \quad (3.29)$$

By injecting this last equality into (3.27), we find:

$$\begin{aligned} 2\gamma(E_{n+1} - E_n) &\leq 2\gamma(k_{n+1} - 2(\lambda + 1 - b)t_n)w_n + ((\lambda + 1 - b)^2 + \xi)\|x_n - x_{n-1}\|^2 \\ &\quad - \frac{2(\lambda + 1 - b)(\lambda^2 + \xi)}{t_n}\|x_{n-1} - x^*\|^2 + 2\gamma\frac{2(\lambda + 1 - b)}{t_n}E_n \\ &\quad + 2(\xi - \lambda(\lambda + 1 - b))\langle x_n - x_{n-1}, x_{n-1} - x^* \rangle \end{aligned} \quad (3.30)$$

By choosing $\xi = \lambda(\lambda + 1 - b)$ (here $\lambda \geq b - 1$), we obtain:

$$\begin{aligned} 2\gamma(E_{n+1} - E_n) &\leq 2\gamma(k_{n+1} - 2(\lambda + 1 - b)t_n)w_n \\ &\quad + (\lambda + 1 - b)(2\lambda + 1 - b)\|x_n - x_{n-1}\|^2 + 2\gamma\frac{2(\lambda + 1 - b)}{t_n}E_n \\ &\quad - 2\frac{\lambda(\lambda + 1 - b)(2\lambda + 1 - b)}{t_n}\|x_{n-1} - x^*\|^2 \end{aligned} \quad (3.31)$$

By definition of k_{n+1} (3.25), we obtain:

$$\begin{aligned} 2\gamma(E_{n+1} - E_n) &\leq 2\gamma((2 - \lambda)(n + b) - 1 - 2(\lambda + 1 - b)t_n)w_n \\ &\quad + (\lambda + 1 - b)(2\lambda + 1 - b)\|x_n - x_{n-1}\|^2 + 2\gamma\frac{2(\lambda + 1 - b)}{t_n}E_n \\ &\quad - 2\frac{\lambda(\lambda + 1 - b)(2\lambda + 1 - b)}{t_n}\|x_{n-1} - x^*\|^2 \\ &= 2\gamma(2b - 3\lambda)(n + b)w_n + 2\gamma(2(\lambda - b) + 1)w_n \\ &\quad + (\lambda + 1 - b)(2\lambda + 1 - b)\|x_n - x_{n-1}\|^2 + 2\gamma\frac{2(\lambda + 1 - b)}{t_n}E_n \\ &\quad - 2\frac{\lambda(\lambda + 1 - b)(2\lambda + 1 - b)}{t_n}\|x_{n-1} - x^*\|^2 \end{aligned} \quad (3.32)$$

By choosing $\lambda = \frac{2b}{3}$, we find:

$$\begin{aligned} 2\gamma(E_{n+1} - E_n) &\leq 2\gamma\frac{(3 - 2b)}{3}w_n + \frac{(3 - b)(3 + b)}{9}\|x_n - x_{n-1}\|^2 \\ &\quad - 2\frac{2b(3 - b)(3 + b)}{27t_n}\|x_{n-1} - x^*\|^2 + 2\gamma\frac{2(3 - b)}{3(n + b - 1)}E_n \end{aligned} \quad (3.33)$$

In this point, firstly we express the term $\|x_n - x_{n-1}\|^2$ with the aid of E_n and w_n and then we regroup the different terms.

From the inequality

$$\|\alpha\|^2 \leq 2\|\alpha + \beta\|^2 + 2\|\beta\|^2, \quad \forall \alpha, \beta \in \mathcal{H}$$

and the definition of E_n , we have (for $\alpha = t_n(x_n - x_{n-1})$ and $\beta = \lambda(x_{n-1} - x^*)$) we find:

$$\begin{aligned} 2\gamma E_n &\geq 2\gamma t_n^2 w_n + \frac{t_n^2}{2} \|x_n - x_{n-1}\|^2 + (\xi - \lambda^2) \|x_{n-1} - x^*\|^2 \\ (\xi = \lambda(\lambda + 1 - b)) &= 2\gamma t_n^2 w_n + \frac{t_n^2}{2} \|x_n - x_{n-1}\|^2 - \lambda(b-1) \|x_{n-1} - x^*\|^2 \end{aligned} \quad (3.34)$$

Therefore, we obtain:

$$\|x_n - x_{n-1}\|^2 \leq 2\gamma \frac{2}{t_n^2} E_n - 4\gamma w_n + \frac{2\lambda(b-1)}{t_n^2} \|x_{n-1} - x^*\|^2 \quad (3.35)$$

By injecting the inequality (3.35) into (3.30), we obtain:

$$\begin{aligned} 2\gamma(E_{n+1} - E_n) &\leq 2\gamma \left(\frac{3-2b}{3} - 2 \frac{(3-b)(3+b)}{9} \right) w_n \\ &\quad + 2 \frac{2b(3-b)(3+b)}{27} \left(\frac{b-1}{t_n^2} - \frac{1}{t_n} \right) \|x_{n-1} - x^*\|^2 \\ &\quad + 2\gamma \frac{2(3-b)(3+b)}{9t_n^2} E_n + 2\gamma \frac{2(3-b)}{3(n+b-1)} E_n \end{aligned} \quad (3.36)$$

Therefore we have:

$$\begin{aligned} 2\gamma(E_{n+1} - E_n) &\leq 2\gamma \frac{(2b^2 - 6b - 9)}{9} w_n - \frac{2b(3-b)(b+3)n}{27(n+b-1)^2} \|x_{n-1} - x^*\|^2 \\ &\quad + 2\gamma \frac{2(3-b)(b+3)}{9(n+b-1)^2} E_n + 2\gamma \frac{2(3-b)}{3(n+b-1)} E_n \\ &= 2\gamma B_1 w_n - B_2 n \frac{\|x_{n-1} - x^*\|^2}{(n+b-1)^2} + \frac{2\gamma a}{(n+b-1)^2} E_n \\ &\quad + \frac{2\gamma c}{n+b-1} E_n \end{aligned} \quad (3.37)$$

where

$$B_1 = \frac{2b^2 - 6b - 9}{9} < 0, \quad \forall b \in (0, 3)$$

$$B_2 = \frac{2b(3-b)(b+3)}{27} > 0, \quad \forall b \in (0, 3)$$

$$a = \frac{2(3-b)(b+3)}{9} > 0, \quad \forall b \in (0, 3)$$

$$c = \frac{2(3-b)}{3} > 0, \quad \forall b \in (0, 3)$$

Hence it follows that for all $n \geq 1$:

$$E_{n+1} - E_n \leq \frac{a}{(n+b-1)^2} E_n + \frac{c}{(n+b-1)} E_n \quad (3.38)$$

which concludes the proof of Lemma 2, with $a = \frac{2(3-b)(b+3)}{9}$ and $c = \frac{2(3-b)}{3}$. \square

We are now ready to give the proof of Theorem 2, by using the estimation (3.17) of Lemma 2 and a discretized version of Gronwall's Lemma (see Lemma 5).

Proof (Theorem 2)

From Lemma 2 for all $n \geq 1$, we have:

$$E_{n+1} - E_n \leq \frac{a}{(n+b-1)^2} E_n + \frac{c}{(n+b-1)} E_n \quad (3.39)$$

with $a = \frac{2(3-b)(b+3)}{9}$ and $c = \frac{2(3-b)}{3}$.

For ease of notation for all $i \in \mathbb{N}$ and $b \in (0, 3)$ we denote as i_b the quantity $i + b - 1$.

By summing (3.39) over n , we find that for all $n \geq 1$ it holds:

$$E_{n+1} \leq E_1 + \sum_{i=1}^n \left(\frac{a}{i_b} + c \right) \frac{E_i}{i_b} \quad (3.40)$$

By applying Lemma 5, for all $n \geq 1$ we find:

$$E_{n+1} \leq E_1 \prod_{i=1}^n \left(1 + \frac{c}{i_b} + \frac{a}{i_b^2} \right) = E_1 e^{\left(\sum_{i=1}^n \log \left(1 + \frac{c}{i_b} + \frac{a}{i_b^2} \right) \right)} \quad (3.41)$$

The function $G(x) = \log \left(1 + \frac{c}{x} + \frac{a}{x^2} \right)$ is positive and non-increasing in $[1, +\infty)$, therefore by summation-integral comparison test for all $n \geq 1$, we have:

$$\sum_{i=1}^n G(i) \leq G(1) + \int_2^n G(x) dx = \log(1+c+a) + \int_2^n \left(\log \left(1 + \frac{c}{x} + \frac{a}{x^2} \right) \right) dx \quad (3.42)$$

By standard integration techniques, for all $n \geq 1$ we find:

$$\begin{aligned}
\int_2^n \log\left(1 + \frac{c}{x} + \frac{a}{x^2}\right) dx &= n \log\left(1 + \frac{c}{n} + \frac{a}{n^2}\right) + \int_2^n \frac{cx + 2a}{x^2 + cx + a} dx + A \\
&= n \log\left(1 + \frac{c}{n} + \frac{a}{n^2}\right) + \frac{c}{2} \int_2^n \frac{2x + c}{x^2 + cx + a} dx + A \\
&\quad + \left(2a - \frac{c^2}{2}\right) \int_2^n \frac{1}{x^2 + cx + a} dx \\
&= n \log\left(1 + \frac{c}{n} + \frac{a}{n^2}\right) + \frac{c}{2} \log(n^2 + cn + a) + A \\
&\quad + \frac{1}{2} \int_2^n \frac{dx}{\left(\frac{2x+c}{\sqrt{4a-c^2}}\right)^2 + 1} \\
&= \frac{c}{2} \log(n^2 + cn + a) + n \log\left(1 + \frac{c}{n} + \frac{a}{n^2}\right) + A \\
&\quad + \left(\sqrt{4a - c^2}\right) \arctan\left(\frac{2n + c}{\sqrt{4a - c^2}}\right)
\end{aligned} \tag{3.43}$$

where $A > 0$ is a renamed constant at each step. Here we stress out that every step is justified, since $4a > c^2$. As the function arctan is bounded for all $n \geq 1$, we obtain:

$$\int_2^n \left(\log\left(1 + \frac{c}{x} + \frac{a}{x^2}\right) \right) dx \leq \frac{c}{2} \log(n^2 + cn + a) + n \log\left(1 + \frac{c}{n} + \frac{a}{n^2}\right) + A \tag{3.44}$$

where $A > 0$ is a (renamed) constant which can be chosen positive. By injecting this last inequality into (3.42), for all $n \geq 1$ and $n_b = n + b - 1$, we have

$$\begin{aligned}
\sum_{i=1}^n G(i_b) &\leq \log(1 + c + a) + \frac{c}{2} \log(n_b^2 + cn_b + a) + n_b \log\left(1 + \frac{c}{n_b} + \frac{a}{n_b^2}\right) + A \\
&\leq \log(1 + c + a) + \frac{c}{2} \log((1 + a + c)n_b^2) + n_b \log\left(1 + \frac{c}{n_b} + \frac{a}{n_b^2}\right) + A \\
&= 2 \log(1 + a + c) + \log(n_b^c) + \log\left(1 + \frac{c + \frac{a}{n_b}}{n_b}\right)^{n_b} + A
\end{aligned} \tag{3.45}$$

By injecting the last inequality into (3.41), for all $n \geq 1$, we obtain:

$$E_{n+1} \leq E_1 \left[2(1 + a + c)A \left(1 + \frac{c + \frac{a}{n_b}}{n_b}\right)^{n_b} n_b^c \right] \leq C n_b^c \tag{3.46}$$

for a positive constant $C > 0$, since $\left(1 + \frac{c + \frac{a}{n_b}}{n_b}\right)^{n_b}$ is bounded, as a convergent sequence (such a bound is for example e^{c+a}). This concludes the proof of Theorem 2 by substituting $n_b = n + b - 1$. \square

4 The perturbed case

In many cases the computation of the proximal operator of a function is not exact. In this section we present i-FB algorithm in presence of some error parameters on the calculation of the proximal operator of the function g and the gradient of the function f , as the ones considered in [6,11,18,21]. In what follows we keep the same notations as in the unperturbed case for the different sequences.

4.1 Inexact computations of the proximal point

In this section, we introduce the different notions used to approximate a proximal operator in this work. Our presentation follows the one of [6]. As recalled in the first section, if F is a proper, convex and l.s.c function and $\gamma > 0$, we can define the proximal map $\text{prox}_{\gamma F}$ by

$$\text{prox}_{\gamma F}(y) = \underset{x \in \mathcal{H}}{\operatorname{argmin}} \left\{ F(x) + \frac{1}{2\gamma} \|x - y\|^2 \right\}$$

Let us denote by

$$G_\gamma(x) = F(x) + \frac{1}{2\gamma} \|x - y\|^2. \quad (4.1)$$

The first order optimality condition for a convex minimum problem yields

$$z = \text{prox}_{\gamma F}(y) \iff 0 \in \partial G_\gamma(z) \iff \frac{y - z}{\gamma} \in \partial F(z) \quad (4.2)$$

We now introduce the notion of ε -subdifferential of F at the point $z \in \text{dom}F$ as:

$$\partial_\varepsilon F(z) = \{y \in \mathcal{H} \mid F(x) \geq F(z) + \langle x - z, y \rangle - \varepsilon, \forall x \in \mathcal{H}\} \quad (4.3)$$

It is worth noticing that it holds:

$$0 \in \partial_\varepsilon F(z) \iff F(z) \leq \inf F + \varepsilon \quad (4.4)$$

The ε -subdifferential is a generalization of the subdifferential as given in section 2. Note that if $\varepsilon > 0$, then $\partial f(x) \subset \partial_\varepsilon f(x)$.

We start by giving some definitions on the different types of approximations of the proximal operator that on can find in [6], following [18,21].

Definition 1 We say that $z \in \mathcal{H}$ is a type 1 approximation of $\text{prox}_{\gamma F}(y)$ with ε precision and we write $z \approx_1 \text{prox}_{\gamma F}(y)$ if and only if

$$0 \in \partial_\varepsilon G_\gamma(z) \quad (4.5)$$

Definition 2 We say that $z \in \mathcal{H}$ is a type 2 approximation of $\text{prox}_{\lambda F}(y)$ with ε precision and we write $z \approx_2 \text{prox}_{\gamma F}(y)$ if and only if

$$\frac{y - z}{\gamma} \in \partial_\varepsilon F(z) \quad (4.6)$$

Notice that if $z \approx_2 \text{prox}_{\gamma F}(y)$, then $z \approx_1 \text{prox}_{\gamma F}(y)$ (see Proposition 1 in [18]).

Finally we make call of a technical lemma taken from [19] (see Lemma 2), that enables to consider approximations of types $j = 1$ or $j = 2$ in the same setting, in the forthcoming analysis.

Lemma 3 If $x \in \mathcal{H}$ is a type 2 approximation of $\text{prox}_{\gamma F}(y)$ with ε precision, then there exists r such that $\|r\| \leq \sqrt{2\gamma\varepsilon}$ and

$$\frac{y - x - r}{\gamma} \in \partial_\varepsilon F(x) \quad (4.7)$$

Notice that when $r = 0$, then we obtain the definition of a type 2 approximation.

4.2 Convergence rate for inexact i-FB algorithm

In this framework we consider the inexact i-FB algorithm as follows:

Algorithm 2 Inexact i-FB

Let $0 < \gamma \leq \frac{1}{L}$ and $b \in (0, 3)$. We consider the sequences $\{t_n\}_{n \in \mathbb{N}^*}$, $\{x_n\}_{n \in \mathbb{N}}$, $\{y_n\}_{n \in \mathbb{N}}$, such that $x_0 = y_0 \in \mathcal{H}$ and for every $n \in \mathbb{N}^*$ we set:

$$x_n = T_{e_n}^{\varepsilon_n}(y_{n-1}) \quad (4.8)$$

$$y_n = x_n + a_n(x_n - x_{n-1}) \quad \text{where } a_n = \frac{n}{n+b} \quad (4.9)$$

$$\text{where } T_{e_n}^{\varepsilon_n}(x) \approx_j^{\varepsilon_n} \text{prox}_{sg}\left(x - \gamma(\nabla f(x) + e_n)\right) \quad \text{where } j \in \{1, 2\}$$

We present here the main results concerning the inexact i-FB algorithm:

Theorem 3 Let $0 < \gamma \leq \frac{1}{L}$, $b \in (0, 3)$ and $\{x_n\}_{n \in \mathbb{N}}$ the sequence generated by the inexact i-FB algorithm. Then for $\lambda = \frac{2b}{3}$ and $\xi = \frac{4b^2}{9}$, for every $\eta > 0$, there exists $C_\eta > 0$, such that for all $n \geq 1$, we have:

$$E_n \leq \frac{\left(2A_n + \sqrt{2(C_\eta + B_n)}\right)^2}{2\gamma} (n + b - 1)^{\frac{2(3-b)}{3} + \eta} \quad (4.10)$$

where

$$A_n = \sum_{i=1}^n t_i^{1-\frac{c+\eta}{2}} (\gamma \|e_i\| + \sqrt{2\gamma\varepsilon_i}) \quad \text{and} \quad B_n = \gamma \sum_{i=1}^n t_i^{2-(c+\eta)} \varepsilon_i \quad (4.11)$$

Corollary 3 Let $0 < \gamma \leq \frac{1}{L}$, $b \in (0, 3)$ and $\{x_n\}_{n \in \mathbb{N}}$ the sequence generated by the inexact i-FB algorithm. Then for every $\eta > 0$, there exists $C_\eta > 0$, such that for all $n \geq 1$, we have:

$$\begin{aligned} F(x_n) - F(x^*) &\leq \frac{\left(2A_n + \sqrt{2(C_\eta + B_n)}\right)^2}{2\gamma(n+b-1)^{\frac{2b}{3}-\eta}} \\ \text{and } \|x_n - x_{n-1}\|^2 &\leq \frac{\left(2A_n + \sqrt{2(C_\eta + B_n)}\right)^2}{(n+b-1)^{\frac{2b}{3}-\eta}} \end{aligned} \quad (4.12)$$

where

$$A_n = \sum_{i=1}^n t_i^{1-\frac{c+\eta}{2}} (\gamma \|e_i\| + \sqrt{2\gamma\varepsilon_i}) \quad \text{and} \quad B_n = \gamma \sum_{i=1}^n t_i^{2-(c+\eta)} \varepsilon_i \quad (4.13)$$

Remark 2 The last Corollary asserts that under the supplementary hypothesis over the perturbation terms A_n and B_n , the convergence rates for the inexact i-FB algorithm remain “almost” the same as in the unperturbed case (i-FB algorithm). Formally, let $0 < \gamma \leq \frac{1}{L}$, $b \in (0, 3)$ and $\{x_n\}_{n \in \mathbb{N}}$ the sequence generated by the inexact i-FB algorithm. If in addition, for every $\eta > 0$, we make the following assumptions:

$$\sum_{n=1}^{+\infty} n^{1-\frac{c+\eta}{2}} \|e_n\| \leq A < +\infty \quad \text{and} \quad \sum_{n=1}^{+\infty} n^{1-\frac{c+\eta}{2}} \sqrt{\varepsilon_n} \leq B < +\infty \quad (4.14)$$

Then there exists $C_\eta > 0$, such that for all $n \geq 1$, we have:

$$F(x_n) - F(x^*) \leq \frac{C_\eta}{2\gamma(n+b-1)^{\frac{2b}{3}-\eta}} \quad \text{and} \quad \|x_n - x_{n-1}\|^2 \leq \frac{C_\eta}{(n+b-1)^{\frac{2b}{3}-\eta}} \quad (4.15)$$

We begin by adapting the Lemma 1 of the previous section, for the perturbed version:

Lemma 4 Let $y \in \mathcal{H}$ and $\gamma \leq \frac{1}{L}$. For all $x \in \mathcal{H}$, we have:

$$F(x) - F(T_e^\varepsilon(y)) + \varepsilon + \langle e + \frac{r}{\gamma}, x - T_e^\varepsilon(y) \rangle \geq \frac{1}{2\gamma} (\|T_e^\varepsilon(y) - x\|^2 - \|y - x\|^2) \quad (4.16)$$

where $r \in \mathcal{H}$ such that $\|r\| \leq \sqrt{2\gamma\varepsilon}$

For a complete proof of Lemma 4, we address the reader to Lemma A.1. in [6].

We are now ready to present the proof of Theorem 3:

Proof (Proof of Theorem 3)

In the same way than the one in the unperturbed case, by applying Lemma 4 to $y = y_n$ and $x = \left(1 - \frac{\lambda}{t_{n+1}}\right)x_n + \frac{\lambda}{t_{n+1}}x^*$ we obtain (here $\lambda \in (0, 1+b)$):

$$\begin{aligned} 2\gamma(t_{n+1}^2 w_{n+1} - t_n^2 w_n) &\leq 2\gamma k_{n+1} w_n + \|(t_n - 1)(x_n - x_{n-1}) + \lambda(x_n - x^*)\|^2 \\ &\quad - v_{n+1} + 2\gamma t_{n+1}^2 \varepsilon_{n+1} \\ &\quad - 2\gamma t_{n+1} \langle e_{n+1} + \frac{r_{n+1}}{\gamma}, \lambda(x_n - x^*) + t_{n+1}(x_{n+1} - x_n) \rangle \end{aligned} \quad (4.17)$$

Therefore by using the last inequality, and performing the same computations as the ones made in proof of Theorem 2, we find that for all $n \geq 1$, it holds:

$$E_{n+1} - E_n \leq \frac{(c + \frac{a}{n+b-1})}{n+b-1} E_n - (n+b) \langle e_{n+1} + \frac{r_{n+1}}{\gamma}, v_{n+1} \rangle + (n+b)^2 \varepsilon_{n+1} \quad (4.18)$$

For ease of notation we denote $n_b = n + b - 1$, which we will replace at the end of the proof. Hence the previous inequality can be rewritten as:

$$E_{n+1} - E_n \leq \frac{(c + \frac{a}{n_b})}{n_b} E_n - (n_b + 1) \langle e_{n+1} + \frac{r_{n+1}}{\gamma}, v_{n+1} \rangle + (n_b + 1)^2 \varepsilon_{n+1} \quad (4.19)$$

Let $\eta > 0$. We define the sequence $\{H_n\}_{n \in \mathbb{N}^*}$, such that for all $n \geq 1$:

$$H_n = \frac{E_n}{n_b^{c+\eta}} + \sum_{i=1}^n i_b^{1-(c+\eta)} \langle e_i + \frac{r_i}{\gamma}, v_i \rangle - \sum_{i=1}^n i_b^{2-(c+\eta)} \varepsilon_i$$

For all $n \geq 1$, we have:

$$\begin{aligned} H_{n+1} - H_n &= \frac{E_{n+1}}{(n_b + 1)^{c+\eta}} - \frac{E_n}{n_b^{c+\eta}} + \frac{(n_b + 1) \langle e_{n+1} + \frac{r_{n+1}}{\gamma}, v_i \rangle}{(n_b + 1)^{c+\eta}} - \frac{(n_b + 1)^2 \varepsilon_{n+1}}{(n_b + 1)^{c+\eta}} \\ &= \frac{E_{n+1} - (1 + \frac{1}{n_b})^{c+\eta} E_n + (n_b + 1) \langle e_{n+1} + \frac{r_{n+1}}{\gamma}, v_i \rangle - (n_b + 1)^2 \varepsilon_{n+1}}{(n_b + 1)^c} \\ &= \frac{E_{n+1} - E_n - \frac{c}{n_b} E_n + (n_b + 1) \langle e_{n+1} + \frac{r_{n+1}}{\gamma}, v_i \rangle - (n_b + 1)^2 \varepsilon_{n+1}}{(n_b + 1)^c} \\ &\quad + \frac{O(n_b^{-2}) E_n - \frac{\eta}{n_b} E_n}{(n_b + 1)^c} \\ (4.19) \quad &\leq \left(O(n_b^{-1}) - \eta \right) \frac{E_n}{n_b^{1+c}} \end{aligned} \quad (4.20)$$

Since for any $\eta > 0$ there exists $N_\eta \in \mathbb{N}$, such that the right-hand side of the last inequality is non-positive, we deduce that the tail sequence $\{H_n\}_{n \geq N_\eta}$ is non-increasing. Therefore by setting $C_\eta = \max\{H_n : n \leq N_\eta\}$, using the definition of H_n , the Cauchy–Schwarz inequality and Lemma 3, for all $n \geq 1$ we find:

$$\begin{aligned} \frac{E_n}{n_b^{c+\eta}} &\leq C_\eta - \sum_{i=1}^n i_b^{1-(c+\eta)} \langle e_i + \frac{r_i}{\gamma}, v_i \rangle + \sum_{i=1}^n i_b^{2-(c+\eta)} \varepsilon_i \\ &\leq C_\eta + \sum_{i=1}^n i_b^{1-(c+\eta)} \|e_i + \frac{r_i}{\gamma}\| \|v_i\| + \sum_{i=1}^n i_b^{2-(c+\eta)} \varepsilon_i \\ &\leq C_\eta + \sum_{i=1}^n i_b^{1-(\frac{c+\eta}{2})} \left(\|e_i\| + \frac{\sqrt{2\gamma\varepsilon_i}}{\gamma} \right) i_b^{-\frac{c+\eta}{2}} \|v_i\| + \sum_{i=1}^n i_b^{2-(c+\eta)} \varepsilon_i \end{aligned} \quad (4.21)$$

Using the last inequality and the definition of $\{E_n\}_{n \geq 1}$, we find:

$$\left(n_b^{\frac{-c-\eta}{2}} \|v_n\| \right)^2 \leq 2C_\eta + 2B_n + 2 \sum_{i=1}^n i_b^{1-(\frac{c+\eta}{2})} (\gamma \|e_i\| + \sqrt{2\gamma\varepsilon_i}) i_b^{-\frac{c+\eta}{2}} \|v_i\| \quad (4.22)$$

By applying Lemma 6 with:

$$u_n = n_b^{\frac{-c-\eta}{2}} \|v_n\|, \quad a_n = 2n_b^{1-(\frac{c+\eta}{2})} (\gamma \|e_n\| + \sqrt{2\gamma\varepsilon_n}) \quad \text{and} \quad S_n = 2C_\eta + 2B_n \quad (4.23)$$

we find that for all $n \geq 1$, it holds:

$$\|v_n\| \leq \left(2A_n + \sqrt{2(C_\eta + B_n)} \right) n_b^{\frac{c+\eta}{2}} \quad (4.24)$$

By injecting this last inequality into (4.21) and multiplying both members by 2γ , we find:

$$\begin{aligned} \frac{2\gamma E_n}{n_b^{c+\eta}} &\leq 2A_n \left(2A_n + \sqrt{2(C_\eta + B_n)} \right) + 2(C_\eta + B_n) \\ &= \left(2A_n + \sqrt{2(C_\eta + B_n)} \right)^2 \end{aligned} \quad (4.25)$$

It follows that:

$$E_n \leq \frac{\left(2A_n + \sqrt{2(C_\eta + B_n)} \right)^2}{2\gamma} n_b^{c+\eta} \quad (4.26)$$

which by replacing n_b by $n + b - 1$ and $c = \frac{2(3-b)}{3}$, concludes the proof of Theorem 4. \square

Acknowledgements The authors would like to thank the anonymous reviewers for all their useful commentaries and advices and for pointing out some important references.

Appendix

The next two Lemmas are the discretized versions of Gronwall's Lemma and Gronwall's-Bellman's Lemma (see for example Theorem 4 in [13] and Lemma 1 in [19]).

Lemma 5 Let C_0 a positive real number and $\{u_n\}_{n \in \mathbb{N}}$, $\{a_n\}_{n \in \mathbb{N}}$ two non-negative sequences such that $u_1 \leq C_0$ and for all $n \geq 1$:

$$u_{n+1} \leq C_0 + \sum_{i=1}^n a_i u_i \quad (5.1)$$

Then for all $n \geq 1$ it holds:

$$u_{n+1} \leq C_0 \prod_{i=1}^n (1 + a_i) \quad (5.2)$$

Lemma 6 Let C_0 a positive real number and $\{u_n\}_{n \in \mathbb{N}}$, $\{a_n\}_{n \in \mathbb{N}}$ two non-negative sequences, such that for all $n \in \mathbb{N}^*$ it holds

$$u_n^2 \leq S_n + \sum_{i=1}^n a_i u_i$$

where $\{S_n\}_{n \in \mathbb{N}}$ is a non-decreasing sequence such that $u_1^2 \leq S_1$. Then for all $n \geq 1$, it holds:

$$u_n \leq \sum_{i=1}^n a_i + \sqrt{S_n}$$

Remark 3 While submitting this paper we were informed that in a parallel but independent way, Attouch and al., worked on the same problem and had just submitted the preprint “Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$ ” (see [4], <https://arxiv.org/abs/1706.05671>).

Nevertheless our method allow us to conclude with Corollary 2 concerning the estimates (3.11), i.e.

$$F(x_n) - F(x^*) \leq \frac{C}{(n+b-1)^{\frac{2b}{3}}} \quad \text{and} \quad \|x_n - x_{n-1}\| \leq \frac{C}{(n+b-1)^{\frac{b}{3}}} \quad (5.3)$$

which consists of an improved result of the one proven in [4], which corresponds to the following estimates (see Theorem 4.1 and Remark 4.1 of [4]):

$$F(x_n) - F(x^*) \leq \frac{C_p}{(n+b-1)^{2p}} \quad \text{and} \quad \|x_n - x_{n-1}\| \leq \frac{C_p}{(n+b-1)^p} \quad (5.4)$$

where $0 < p < \frac{b}{3}$. In addition we give a complete proof of Theorem 3, concerning the results for the inexact i-FB algorithm.

References

1. Apidopoulos, V., Aujol, J.F., Dossal, C.: The differential inclusion modeling FISTA algorithm and optimality of convergence rate in the case $b \leq 3$. *SIAM J. Optim.* **28**(1), 551–574 (2018)
2. Attouch, H., Cabot, A.: Convergence rates of inertial forward–backward algorithms. *SIAM J. Optim.* **28**(1), 849–874 (2018)
3. Attouch, H., Chbani, Z., Peypouquet, J., Redont, P.: Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Math. Program.* **168**, 1–53 (2016)
4. Attouch, H., Chbani, Z., Riahi, H.: Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$ (2017). arXiv preprint [arXiv:1706.05671](https://arxiv.org/abs/1706.05671)
5. Attouch, H., Peypouquet, J.: The rate of convergence of Nesterov’s accelerated forward–backward method is actually faster than $1/k^2$. *SIAM J. Optim.* **26**(3), 1824–1834 (2016)
6. Aujol, J.F., Dossal, C.: Stability of over-relaxations for the forward–backward algorithm, application to FISTA. *SIAM J. Optim.* **25**(4), 2408–2433 (2015)
7. Aujol, J.F., Dossal, C.: Optimal rate of convergence of an ode associated to the fast gradient descent schemes for $b > 0$. *J. Differ. Equ.* (preprint available hal-01547251) (2017)
8. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, Berlin (2011)
9. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**(1), 183–202 (2009)
10. Chambolle, A., Dossal, C.: On the convergence of the iterates of the fast iterative shrinkage/thresholding algorithm. *J. Optim. Theory Appl.* **166**(3), 968–982 (2015)
11. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward–backward splitting. *Multiscale Model. Simul.* **4**(4), 1168–1200 (2005)
12. Güler, O.: New proximal point algorithms for convex minimization. *SIAM J. Optim.* **2**(4), 649–664 (1992)
13. Holte, J.M.: Discrete Gronwall lemma and applications. In: MAA-NCS meeting at the University of North Dakota, vol. 24, pp. 1–7 (2009)
14. Johnstone, P.R., Moulin, P.: Local and global convergence of a general inertial proximal splitting scheme (2016). arXiv preprint [arXiv:1602.02726](https://arxiv.org/abs/1602.02726)
15. Kim, D., Fessler, J.A.: Optimized first-order methods for smooth convex minimization (2014). arXiv preprint [arXiv:1406.5468](https://arxiv.org/abs/1406.5468)
16. Nesterov, Y.: A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Sov. Math. Doklady* **27**, 372–376 (1983)
17. Nesterov, Y.: Introductory Lectures on Convex Optimization: A Basic Course. Springer, Berlin (2013)
18. Salzo, S., Villa, S.: Inexact and accelerated proximal point algorithms. *J. Convex Anal.* **19**(4), 1167–1192 (2012)
19. Schmidt, M., Le Roux, N., Bach, F.: Convergence rates of inexact proximal-gradient methods for convex optimization. In: NIPS (2011)
20. Su, W., Boyd, S., Candes, E.J.: A differential equation for modeling Nesterov’s accelerated gradient method: theory and insights. *J. Mach. Learn. Res.* **17**(153), 1–43 (2016)
21. Villa, S., Salzo, S., Baldassarre, L., Verri, A.: Accelerated and inexact forward–backward algorithms. *SIAM J. Optim.* **23**(3), 1607–1633 (2013)