J. Appl. Numer. Optim. 6 (2024), No. 2, pp. 249-269
Available online at http://jano.biemdas.com
https://doi.org/10.23952/jano.6.2024.2.06

ON THE STRONG CONVERGENCE OF AN INERTIAL PROXIMAL ALGORITHM WITH A TIME SCALE, HESSIAN-DRIVEN DAMPING, AND A TIKHONOV REGULARIZATION TERM

AKRAM CHAHID BAGY, ZAKI CHBANI*, HASSAN RIAHI

Department of Mathematics, Faculty of Sciences Semlalia, Cadi Ayyad University, Marrakech, Morocco

Abstract. This paper concerns with convergence properties of an inertial proximal algorithm that contains a Tikhonov term regularization, time scale parameter, and a Hessian-driven damping in a Hilbert space. More precisely, we prove the strong convergence of the proximal algorithm obtained by temporal discretization of a continuous dynamic that we treated earlier in a previous work. We also obtain the convergence of the values to the global minimum of the objective function, and a strong convergence of the gradient and the velocity towards zero. Finally, we present a numerical example to illustrate our results.

Keywords. Convex optimization; Heavy-ball method; Hessian-driven damping; Proximal point algorithm; Tikhonov approximation.

2020 Mathematics Subject Classification. 65K15, 65K15, 90C30.

1. INTRODUCTION

In this paper, $\mathscr{H}$ is a real Hilbert space, with inner product and norm denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ respectively, while $f : \mathscr{H} \to \mathbb{R}$ is a convex and continuously differentiable function, whose gradient denoted by ∇f is Lipschitz continuous on bounded sets. We are interested in the following minimization problem:

$$\min\{f(x) : x \in \mathscr{H}\} \tag{1.1}$$

whose solution set $S := \operatorname{argmin}_{\mathscr{H}} f$ is nonempty.

One of the best-known methods for solving problem (1.1) is the classical proximal point algorithm initiated by Martinet [1] which is defined, for a general set-valued monotone maximal operator $T : \mathscr{H} \to 2^{\mathscr{H}}$, as follows

$$x_{k+1} = J^T_{\lambda_k}(x_k), \tag{\textbf{PPA}}$$

where $(\lambda_k)_{k\in\mathbb{N}}$ represents the step sequence of the algorithm, and $J^T_{\lambda_k} : \mathscr{H} \to \mathscr{H}$ is the resolvent of the monotone operator T, defined by $J^T_{\lambda_k}(x) := (I + \lambda_k T)^{-1}(x)$.

*Corresponding author.

E-mail address: akram.bagy@gmail.com (A. C. Bagy), chbaniz@uca.ac.ma (Z. Chbani), h-riahi@uca.ac.ma (H. Riahi).

Received 5 November 2023; Accepted 11 March 2024; Published online 1 April 2024.

Given that $T^{-1}(0) \neq \emptyset$ and λ_k bounded away from zero, Rockafellar [2] demonstrated that (**PPA**) weakly converges to a solution of problem (1.1), and then asked whether (**PPA**) converges strongly or not? Güler [3] answered this question by creating an example where the sequence converges weakly but not strongly. Subsequently, a number of authors suggested modifications to (**PPA**) to improve weak convergence to strong convergence; see, e.g., [4, 5, 6, 7, 8, 9].

We stress that when $T = \nabla f$, the (**PPA**) described here applies to problem (1.1), and it is clear that (**PPA**) is equivalent to

$$\frac{x_{k+1} - x_k}{\lambda_k} + \nabla f(x_{k+1}) = 0,$$

which can be considered as an implicit discretisation of the following continuous dynamical system $\dot{x}(t) + \nabla f(x(t)) = 0$. This was studied by many authors recently; see, e.g., [10, 11, 12, 13]. More generally, Polyak [14] considered the Heavy ball system

$$\ddot{x}(t) + \alpha \dot{x}(t) + \nabla f(x(t)) = 0, \tag{HBF}$$

where the damping coefficient $\alpha > 0$ is fixed. In the case that f is strongly convex, he proved convergence at an exponential rate of $f(x(t))$ to $\min_{\mathscr{H}} f$; see [15]. The weak convergence of the trajectories of (**HBF**) was obtained by Alvarez in [16]. In the quest of strong convergence of trajectories, Attouch and Czarnecki [17] added the Tikhonov term $\varepsilon(t)x(t)$ to the (**HBF**) system

$$\ddot{x}(t) + \alpha \dot{x}(t) + \nabla f(x(t)) + \varepsilon(t)x(t) = 0. \tag{HBFC}$$

They proved, when $\varepsilon(\cdot)$ tends slowly to zero, i.e., $\int_0^{+\infty} \varepsilon(t)\, dt = +\infty$, that any solution $x(\cdot)$ of (**HBFC**) converges strongly to the minimum norm element of $\operatorname{argmin}_{\mathscr{H}} f$. (**HBFC**) system is a particular case of the general dynamic model

$$\ddot{x}(t) + \alpha \dot{x}(t) + \nabla f(x(t)) + \varepsilon(t)\nabla g(t) = 0,$$

studied by Attouch and Czarnecki in [18] which involves two potential functions f and g intervening with different time scales. When $\varepsilon(\cdot)$ tends to zero moderately slowly, they showed that the trajectories converge asymptotically to equilibria that are solutions to the following hierarchical minimization problem: they minimized the potential g on the set of minimizers of f.

In the quest for faster convergence, Attouch, Chbani and Riahi carried out another development in [19] which consists in the study of the dynamical system with asymptotically vanishing damping

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) + \varepsilon(t)x(t) = 0. \tag{AVD$_{\alpha,\varepsilon}$}$$

It is a Tikhonov regularization of the dynamic system

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0, \tag{AVD$_{\alpha}$}$$

which was introduced by Su, Boyd and Candès in [20], where they showed that, for $\alpha = 3$, the above system can be considered as a continuous version of the Nesterov accelerated gradient method [21, 22, 23, 24] with $f(x(t)) - \min_{\mathscr{H}} f = \mathscr{O}\left(\frac{1}{t^2}\right)$ as $t \to +\infty$. For the other two cases $\alpha > 3$ and $\alpha < 3$, we refer to [25, 26, 27, 28, 29, 30].

According to the structure of the heavy ball method for strongly convex functions, Attouch and László in [31] considered the following system

$$\ddot{x}(t)+\alpha\sqrt{\varepsilon(t)}\dot{x}(t)+\nabla f(x(t))+\varepsilon(t)x(t)=0,$$

where the viscous damping coefficient is proportional to the square root of the Tikhonov regularization parameter. Given that $\alpha>0$ and $\varepsilon(t)=\frac{1}{t^r}$ with $\frac{2}{3}<r<2$, they proved that $f(x(t))-\min_{\mathscr{H}} f=\mathscr{O}\left(\frac{1}{t^{\frac{3r}{2}-1}}\right)$, as $t\to+\infty$. In addition, $\liminf_{t\to+\infty}\|x(t)-x^*\|=0$ as soon as $\alpha>3$, where x^* the element of minimum norm of $\arg\min_{\mathscr{H}} f$.

In [32], Attouch et al. succeeded in obtaining $\lim_{t\to\infty}\|x(t)-x^*\|=0$ in the case that $\alpha>0$ and $\varepsilon(t)=\frac{1}{t^r}$ with $0<r<2$. They obtained, for t large enough, the following convergence rates

$$f(x(t))-\min_{\mathscr{H}} f=\mathscr{O}\left(\frac{1}{t^r}\right),\quad \left\|x(t)-x_{\varepsilon(t)}\right\|^2=\mathscr{O}\left(\frac{1}{t^{\frac{2-r}{2}}}\right)$$
$$\text{and}\quad \|\dot{x}(t)\|^2=\mathscr{O}\left(\frac{1}{t^{\frac{2+r}{2}}}\right)$$

where $x_{\varepsilon(t)}=\underset{\mathscr{H}}{\operatorname{argmin}}\{f(x)+\frac{\varepsilon(t)}{2}\|x\|^2\}$, which ensures the strong convergence of $x(t)$ to x^*. Later, in [33], the authors showed that the trajectories generated by the system

$$\ddot{x}(t)+\alpha\sqrt{\varepsilon(t)}\dot{x}(t)+\delta\nabla^2 f(x(t))\dot{x}(t)+\nabla f(x(t))+\varepsilon(t)x(t)=0, \tag{1.2}$$

which is driven by the Hessian of the function f, retain the same convergence rates for $\varepsilon(t)=\frac{1}{t^r}$ with $1<r<2$, and induce a significant attenuation of oscillations.

Very recently, we studied in [34] the following dynamic system involving a Tikhonov term $\frac{c}{\beta(t)}x(t)$ and a Hessian-driven damping:

$$\ddot{x}(t)+\alpha\dot{x}(t)+\delta\frac{d}{dt}\left(\nabla f(x(t))\right)+\beta(t)\left(\nabla f(x(t))+\frac{c}{\beta(t)}x(t)\right)=0, \tag{1.3}$$

where α, $c,\delta>0$, and the time scale parameter $\beta(\cdot)$ is a positive nondecreasing function with $\lim_{t\to+\infty}\beta(t)=+\infty$. By imposing adequate hypothesis on first and second order derivatives of β, which may include the special cases $t^p(\ln(t))^q$ with $p>1$, $q\geq 0$ and $e^{\gamma t^p}$ with $p\in]0,1[$, $\gamma>0$, we obtained, when $t\to+\infty$:

- Convergence rate of values: $f(x(t))-\min_{\mathscr{H}} f=\mathscr{O}\left(\frac{1}{\beta(t)}\right)$.
- Strong convergence of the trajectories with: $\|x(t)-x_t\|^2=\mathscr{O}\left(\frac{\dot{\beta}(t)}{\beta(t)}+e^{-\mu t}\right)$;

  where $x_t:=\underset{\mathscr{H}}{\operatorname{argmin}}\{f(x)+\frac{c}{2\beta(t)}\|x\|^2\}$ and $\mu<\frac{\alpha}{2}$.
- Strong convergence rate of the gradients:
$$\|\nabla f(x(t))\|^2=\mathscr{O}\left(\frac{\dot{\beta}(t)}{\beta(t)}+e^{-\mu t}+\frac{1}{\beta^2(t)}\right).$$
- Strong convergence rate of velocity: $\|\dot{x}(t)\|^2=\mathscr{O}\left(\frac{\dot{\beta}(t)}{\beta(t)}+e^{-\mu t}+\frac{1}{\beta^2(t)}\right)$.

The case $\delta = 0$, i.e., without Hessian-driven damping, was treated in [35].

We see that the Tikhonov regularization coefficient $\frac{c}{\beta(t)}$ must vanish since $\beta(t)$ goes to $+\infty$. Thus dynamical system (1.3) can be seen as a combination of two techniques: a time scaling of the damped inertial gradient system (see [36, 37]) and the Tikhonov regularization (see [32]). The technique of time scaling was employed for dynamical system (1.3) in order to expedite the convergence of function values along its trajectory, and the presence of the Tikhonov term enhances the convergence of trajectories from weak convergence to strong convergence. Furthermore, it ensures convergence towards the minimizer with the minimal norm rather than any arbitrary element in the set of all minimizers.

Our contribution: In this work, we study the associated proximal algorithm obtained by temporal discretization of (1.3). More precisely, we show that the corresponding proximal algorithm has convergence properties similar to those of the continuous dynamics algorithm. In this context, we consider the following implicit discretization case of (1.3) in time with step $h > 0$. Setting $s = h^2$, we have, for all $k \geq 1$,

$$\begin{aligned}&(x_{k+1}-x_k)-(x_k-x_{k-1})+\alpha\sqrt{s}(x_{k+1}-x_k)+\delta\sqrt{s}\left[\nabla f(x_{k+1})-\nabla f(x_k)\right]\\&+s\beta_k\nabla f(x_{k+1})+csx_{k+1}=0,\end{aligned} \tag{1.4}$$

which gives

$$\left(1+\alpha\sqrt{s}+sc\right)(x_{k+1}-x_k)+(\delta\sqrt{s}+s\beta_k)\nabla f(x_{k+1}) = x_k - x_{k-1}+\delta\sqrt{s}\nabla f(x_k)-sc)x_k. \tag{1.5}$$

Let us denote briefly $d = \dfrac{1}{1+\alpha\sqrt{s}+sc}$ and $\lambda_k = d\,(\delta\sqrt{s}+s\beta_k)$. By multiplying (1.5) by d, we obtain

$$(I+\lambda_k\nabla f)(x_{k+1}) = x_k + d\left[x_k - x_{k-1}+\delta\sqrt{s}\nabla f(x_k)\right]-dscx_k.$$

We can then formulate the Inertial Proximal Algorithm as

(**IPATTH**): **I**nertial **P**roximal **A**lgorithm with **T**ikhonov regularization **T**ime scale, and **H**essian damping.
Let $\alpha,\ c,\ \delta > 0$. Step k: set $d = \dfrac{1}{1+\alpha\sqrt{s}+sc}$ and $\lambda_k = d\,(\delta\sqrt{s}+s\beta_k)$ $\begin{cases} y_k = x_k + d(x_k - x_{k-1}+\delta\sqrt{s}\nabla f(x_k)) \\ x_{k+1} = \operatorname{prox}_{\lambda_k f}(y_k - dscx_k). \end{cases}$ (**IPATTH**)

The objective of this work is to obtain simultaneously strong convergence of algorithm **(IPATTH)**, as well as convergence rates similar to those obtained in the continuous cases mentioned above.

The remainder of the paper is organized as follows. In Section 2, we recall basic facts concerning Thikonov approximation, and formulate the proposed Lyapunov energy sequence, and we then present the main estimate of this sequence. In Section 3, we give the main results of the paper concerning the asymptotic convergence properties of algorithm **(IPATTH)**. Section 4 focuses on two specific cases of the sequence β_k. Finally, the last section, Section 5, contains numerical illustrations to end this paper.

2. ESTIMATION OF THE LYAPUNOV ENERGY SEQUENCE

Throughout this section, we assume that

$$\begin{cases} f : \mathcal{H} \longrightarrow \mathbb{R} \textit{ is convex, of class } C^1, \nabla f \textit{ is Lipschitz continuous on bounded set}, \\ \operatorname{argmin} f \neq \emptyset \textit{ and denote } x^* \textit{ as its minimum norm element}; \end{cases} \tag{$\mathbf{H}_0'$}$$

$$\begin{cases} (\beta_k)_{k\geq k_0} \textit{ is a nondecreasing sequence, such that} \\ \lim\limits_{k\to+\infty} \beta_k = +\infty \text{ and } \lim\limits_{k\to+\infty} \beta_{k+1} - \beta_k = +\infty. \end{cases} \tag{$\mathbf{H}_0$}$$

Let us define the real function $\phi_k : \mathcal{H} \to \mathbb{R}$ by $\phi_k(x) := f(x) + \frac{c}{2\beta_k}\|x\|^2$. It is clear that ϕ_k is $\frac{c}{\beta_k}$-strongly convex. Then, for each $k \geq k_0$, there exists a unique minimizer $x_{\beta_k} \in \mathcal{H}$ of the strongly convex function ϕ_k. The first order optimality condition gives

$$\nabla f(x_{\beta_k}) + \frac{c}{\beta_k} x_{\beta_k} = 0. \tag{2.1}$$

This means that $x_{\beta_k} = \operatorname{prox}_{\frac{\beta_k}{c} f}(0) = (I + \frac{\beta_k}{c}\nabla f)^{-1}(0)$, and then the following properties are satisfied

$$\forall k \geq k_0 : \ \left\|x_{\beta_k}\right\| \leq \|x^*\| \quad \text{and} \quad \lim_{k\longrightarrow+\infty} \left\|x_{\beta_k} - x^*\right\| = 0. \tag{2.2}$$

Let us introduce the discrete energy sequence $(E_{p,k})_{k\geq k_0} \subset \mathbb{R}^+$ that plays a key role in our Lyapunov analysis. It is defined by

$$E_k = s\beta_k \left(\phi_k(x_k) - \phi_k(x_{\beta_k})\right) + \frac{1}{2}\|v_k\|^2, \tag{2.3}$$
$$\text{with } v_k = \tau(x_k - x_{\beta_{k-1}}) + (x_k - x_{k-1}) + \delta\sqrt{s}\nabla f(x_k),$$

where τ is a positive constant.

We now give, in the following Lemma, the relation between the asymptotic behavior of sequence (E_k) and the convergence rate of the values and iterates.

Lemma 2.1. *Let (x_k) be the sequence generated by the algorithm* (**IPATTH**)*, and (E_k) be the energy sequence defined in* (2.3)*. Then, for any $k \geq k_0$,*

$$f(x_k) - \min_{\mathcal{H}} f \leq \frac{1}{\beta_k}\left(\frac{E_k}{s} + \frac{c}{2}\left(\|x^*\|^2 - \|x_k\|^2\right)\right) \tag{2.4}$$

and

$$\left\|x_k - x_{\beta_k}\right\|^2 \leq \frac{2}{sc} E_k. \tag{2.5}$$

If $\lim\limits_{k\to+\infty} E_k = 0$, then x_k converges strongly to x^*.

Proof. Based on the definition of ϕ_k and E_k, we have

$$\begin{aligned} f(x_k)-\min_{\mathscr{H}} f &= \phi_k(x_k)-\frac{c}{2\beta_k}\|x_k\|^2-\phi_k(x^*)+\frac{c}{2\beta_k}\|x^*\|^2 \\ &= [\phi_k(x_k)-\phi_k(x_{\beta_k}))]+\left[\underbrace{\phi_k(x_{\beta_k})-\phi_k(x^*)}_{\leq 0}\right]+\frac{c}{2\beta_k}\|x^*\|^2-\frac{c}{2\beta_k}\|x_k\|^2 \\ &\leq \phi_k(x_k)-\phi_k(x_{\beta_k})+\frac{c}{2\beta_k}\left(\|x^*\|^2-\|x_k\|^2\right) \\ &\leq \frac{1}{\beta_k}\left(\frac{E_k}{s}+\frac{c}{2}\left(\|x^*\|^2-\|x_k\|^2\right)\right). \end{aligned}$$

On the other hand, from the strong convexity of ϕ_k and $x_{\beta_k} := \operatorname{argmin}_{\mathscr{H}} \phi_k$, we obtain

$$\phi_k(x_k)-\phi_k(x_{\beta_k}) \geq \frac{c}{2\beta_k}\left\|x_k-x_{\beta_k}\right\|^2,$$

which, combined with (2.3), gives $\dfrac{E_k}{s\beta_k} \geq \dfrac{c}{2\beta_k}\left\|x_k-x_{\beta_k}\right\|^2$, which yields (2.5). □

In the following Lemma, we give some properties of the viscosity curve $k \to x_{\beta_k}$, in relation to the Moreau envelope. Firstly, let us recall that the Moreau envelope of f is the function $f_\lambda : \mathscr{H} \longrightarrow \mathbb{R}$ $(\lambda > 0)$, defined by

$$f_\lambda(x) = \min_{y\in\mathscr{H}}\left\{f(y)+\frac{1}{2\lambda}\|x-y\|^2\right\}, \text{ for any } x \in \mathscr{H}, \tag{2.6}$$

and $\operatorname{prox}_{\lambda f}(x)$ is the unique point where the minimum value is achieved in (2.6), *i.e.*,

$$f_\lambda(x) = f(\operatorname{prox}_{\lambda f}(x))+\frac{1}{2\lambda}\left\|x-\operatorname{prox}_{\lambda f}(x)\right\|^2.$$

Lemma 2.2. *For all $k \geq k_0$, the following properties are satisfied:*

i) $\phi_k(x_{\beta_k})-\phi_{k+1}(x_{\beta_{k+1}}) \leq \dfrac{c}{2}\left(\dfrac{1}{\beta_k}-\dfrac{1}{\beta_{k+1}}\right)\left\|x_{\beta_{k+1}}\right\|^2,$

ii) $\left\|x_{\beta_{k+1}}-x_{\beta_k}\right\|^2 \leq \dfrac{\beta_{k+1}-\beta_k}{\beta_{k+1}}\langle x_{\beta_{k+1}}, x_{\beta_{k+1}}-x_{\beta_k}\rangle$, *and thus*

$$\left\|x_{\beta_{k+1}}-x_{\beta_k}\right\| \leq \frac{\beta_{k+1}-\beta_k}{\beta_{k+1}}\left\|x_{\beta_{k+1}}\right\|.$$

Proof. i) We have

$$\phi_k(x_{\beta_k}) = f_{\frac{\beta_k}{c}}(0) = \min_{y\in\mathscr{H}}\left\{f(y)+\frac{c}{2\beta_k}\|y\|^2\right\} \leq f(x_{\beta_{k+1}})+\frac{c}{2\beta_k}\left\|x_{\beta_{k+1}}\right\|^2,$$

and $\phi_{k+1}(x_{\beta_{k+1}}) = f_{\frac{\beta_{k+1}}{c}}(0) = f(x_{\beta_{k+1}})+\frac{c}{2\beta_{k+1}}\left\|x_{\beta_{k+1}}\right\|^2$, which implies that

$$\phi_k(x_{\beta_k})-\phi_{k+1}(x_{\beta_{k+1}}) \leq \frac{c}{2}\left(\frac{1}{\beta_k}-\frac{1}{\beta_{k+1}}\right)\left\|x_{\beta_{k+1}}\right\|^2.$$

ii) One has $\frac{-c}{\beta_k}x_{\beta_k} = \nabla f(x_{\beta_k})$ and $\frac{-c}{\beta_{k+1}}x_{\beta_{k+1}} = \nabla f(x_{\beta_{k+1}})$, so the monotonicity of ∇f gives $\left\langle \frac{c}{\beta_k}x_{\beta_k} - \frac{c}{\beta_{k+1}}x_{\beta_{k+1}}, x_{\beta_{k+1}} - x_{\beta_k} \right\rangle \geq 0$. It follows that

$$\frac{c}{\beta_k}\left\|x_{\beta_{k+1}} - x_{\beta_k}\right\|^2 \leq \left(\frac{c}{\beta_k} - \frac{c}{\beta_{k+1}}\right)\left\langle x_{\beta_{k+1}}, x_{\beta_{k+1}} - x_{\beta_k}\right\rangle,$$

which gives the claim. The last statement follows directly from the Cauchy-Schwarz inequality.

□

Under general assumptions, we give in the following Theorem a very important estimate of the Lyapunov energy sequence (E_k) defined in (2.3).

Theorem 2.1. *Let (x_k) be a sequence generated via the algorithm* (**IPATTH**) *with $c \geq \alpha^2$. Suppose* ($\mathbf{H}_0$) *and*

$$\lim_{k \longrightarrow +\infty} \frac{\beta_{k+1}}{\beta_k} < 1 + \alpha\sqrt{s}. \qquad (\mathbf{H}_1)$$

Then there exist $a > 1$ and $k_1 \geq k_0$ such that, for all $k \geq k_1$,

$$E_{k+1} \leq e^{\rho(k_1-k-1)}E_{k_1} + \frac{\|x^*\|^2}{e^{\rho(k+1)}}\left(\sum_{j=k_1}^{k} e^{\rho(j+1)}\theta_j\right); \qquad (2.7)$$

where $\rho = \dfrac{\alpha\sqrt{s}}{1+a+\alpha\sqrt{s}}$ and

$$\theta_k = \frac{1}{2}\left(a\rho(2a+1)\frac{(\beta_k - \beta_{k-1})^2}{\beta_k^2} + sc\left(\frac{\beta_{k+1}}{\beta_k} - 1\right) + \frac{a\rho^2\delta c\sqrt{s}}{(1-\rho)\beta_k}\right).$$

Proof. To simplify the writing of the proof, we use the following notations: for any sequence (u_k) in $\mathscr{H}$, we write, for $k \geq 0$,

$$\dot{u}_k := u_{k+1} - u_k \quad \text{and} \quad \ddot{u}_{k-1} := \dot{u}_k - \dot{u}_{k-1} = u_{k+1} - 2u_k + u_{k-1}.$$

We have

$$\begin{aligned}
&E_{k+1} - E_k \\
&= s\beta_{k+1}\left(\phi_{k+1}(x_{k+1}) - \phi_{k+1}(x_{\beta_{k+1}})\right) - s\beta_k\left(\phi_k(x_k) - \phi_k(x_{\beta_k})\right) + \frac{1}{2}\left(\|v_{k+1}\|^2 - \|v_k\|^2\right).
\end{aligned}$$

Combining this last equality with this elementary algebraic inequality

$$\frac{1}{2}\left(\|v_{k+1}\|^2 - \|v_k\|^2\right) = \langle v_{k+1} - v_k, v_{k+1}\rangle - \frac{1}{2}\|v_{k+1} - v_k\|^2 \leq \langle \dot{v}_k, v_{k+1}\rangle,$$

we obtain

$$\begin{aligned}
E_{k+1} - E_k &\leq s\beta_{k+1}\left(\phi_{k+1}(x_{k+1}) - \phi_{k+1}(x_{\beta_{k+1}})\right) - s\beta_k\left(\phi_k(x_k) - \phi_k(x_{\beta_k})\right) + \langle \dot{v}_k, v_{k+1}\rangle \\
&= s\beta_k\left[\phi_k(x_{k+1}) - \phi_k(x_k)\right] + s(\beta_{k+1} - \beta_k)\left[\phi_k(x_{k+1}) - \phi_k(x_{\beta_k})\right] \\
&\quad + s\beta_{k+1}\left[\phi_{k+1}(x_{k+1}) - \phi_k(x_{k+1}) + \phi_k(x_{\beta_k}) - \phi_{k+1}(x_{\beta_{k+1}})\right] + \langle \dot{v}_k, v_{k+1}\rangle.
\end{aligned}$$

Obviously, from the definition of ϕ_k we have

$$\phi_{k+1}(x_{k+1})-\phi_k(x_{k+1})=\frac{c}{2}\left(\frac{1}{\beta_{k+1}}-\frac{1}{\beta_k}\right)\|x_{k+1}\|^2.$$

According to Lemma 2.2 *i*), we obtain

$$\begin{aligned}E_{k+1}-E_k\leq\ & s\beta_k\left[\phi_k(x_{k+1})-\phi_k(x_k)\right]+s(\beta_{k+1}-\beta_k)\left[\phi_k(x_{k+1})-\phi_k(x_{\beta_k})\right]\\&+\frac{sc}{2}\left(\frac{\beta_{k+1}}{\beta_k}-1\right)\left[\left\|x_{\beta_{k+1}}\right\|^2-\|x_{k+1}\|^2\right]+\langle\dot{v}_k,v_{k+1}\rangle.\end{aligned}\tag{2.8}$$

On the other hand, we have, for every $k\geq k_0$,

$$\begin{aligned}\dot{v}_k=v_{k+1}-v_k&=\tau(x_{k+1}-x_{\beta_k})+(x_{k+1}-x_k)+\delta\sqrt{s}\nabla f(x_{k+1})\\&\quad-\tau(x_k-x_{\beta_{k-1}})-(x_k-x_{k-1})-\delta\sqrt{s}\nabla f(x_k)\\&=\tau(\dot{x}_k-\dot{x}_{\beta_{k-1}})+\ddot{x}_{k-1}+\delta\sqrt{s}\left[\nabla f(x_{k+1})-\nabla f(x_k)\right].\end{aligned}$$

Returning to (1.4), we also obtain

$$\ddot{x}_{k-1}+\delta\sqrt{s}\left[\nabla f(x_{k+1})-\nabla f(x_k))\right]=-\alpha\sqrt{s}\dot{x}_k-s\beta_k\nabla f(x_{k+1})-csx_{k+1}.$$

The combination of the last two equalities gives

$$\begin{aligned}\dot{v}_k&=\tau(\dot{x}_k-\dot{x}_{\beta_{k-1}})-\alpha\sqrt{s}\dot{x}_k-s\beta_k\nabla f(x_{k+1})-csx_{k+1}\\&=(\tau-\alpha\sqrt{s})\dot{x}_k-\tau\dot{x}_{\beta_{k-1}}-s\beta_k\left(\nabla f(x_{k+1})+\frac{c}{\beta_k}x_{k+1}\right)\\&=(\tau-\alpha\sqrt{s})\dot{x}_k-\tau\dot{x}_{\beta_{k-1}}-s\beta_k\nabla\phi_k(x_{k+1}).\end{aligned}$$

Thus

$$\begin{aligned}\langle\dot{v}_k,v_{k+1}\rangle&=\left\langle(\tau-\alpha\sqrt{s})\dot{x}_k-\tau\dot{x}_{\beta_{k-1}}-s\beta_k\nabla\phi_k(x_{k+1}),\tau(x_{k+1}-x_{\beta_k})+\dot{x}_k+\delta\sqrt{s}\nabla f(x_{k+1})\right\rangle\\&=\tau(\tau-\alpha\sqrt{s})\left\langle\dot{x}_k,x_{k+1}-x_{\beta_k}\right\rangle+(\tau-\alpha\sqrt{s})\|\dot{x}_k\|^2+\delta\sqrt{s}(\tau-\alpha\sqrt{s})\langle\dot{x}_k,\nabla f(x_{k+1})\rangle\\&\quad-\tau^2\left\langle\dot{x}_{\beta_{k-1}},x_{k+1}-x_{\beta_k}\right\rangle-\tau\left\langle\dot{x}_{\beta_{k-1}},\dot{x}_k\right\rangle-\delta\tau\sqrt{s}\left\langle\dot{x}_{\beta_{k-1}},\nabla f(x_{k+1})\right\rangle\\&\quad-s\tau\beta_k\left\langle\nabla\phi_k(x_{k+1}),x_{k+1}-x_{\beta_k}\right\rangle-s\beta_k\langle\nabla\phi_k(x_{k+1}),\dot{x}_k\rangle\\&\quad-\delta s\sqrt{s}\beta_k\langle\nabla\phi_k(x_{k+1}),\nabla f(x_{k+1})\rangle.\end{aligned}\tag{2.9}$$

According to condition $(\mathbf{H}_1)$ and $\lim_{a\to+\infty}\frac{(a-1)\alpha\sqrt{s}}{a+1+\alpha\sqrt{s}}=\alpha\sqrt{s}$, one can choose $a>1$ such that, for all k large enough, $\frac{\beta_{k+1}}{\beta_k}-1\leq\frac{(a-1)\alpha\sqrt{s}}{a+1+\alpha\sqrt{s}}$; equivalently

$$\frac{\beta_{k+1}}{\beta_k}\leq\frac{a(\alpha\sqrt{s}+1)+1}{a+1+\alpha\sqrt{s}}.\tag{2.10}$$

From Lemma 2.2 *ii*) and inequality (2.2), we have

$$-\tau\left\langle\dot{x}_{\beta_{k-1}},\dot{x}_k\right\rangle\leq\frac{a\tau}{2}\left\|\dot{x}_{\beta_{k-1}}\right\|^2+\frac{\tau}{2a}\|\dot{x}_k\|^2\leq\frac{a\tau}{2}\frac{\dot{\beta}_{k-1}^2}{\beta_k^2}\|x^*\|^2+\frac{\tau}{2a}\|\dot{x}_k\|^2,\tag{2.11}$$

and
$$-\delta\tau\sqrt{s}\left\langle \dot{x}_{\beta_{k-1}}, \nabla f(x_{k+1})\right\rangle \leq \frac{a\tau}{2}\left\|\dot{x}_{\beta_{k-1}}\right\|^2 + \frac{s\delta^2\tau}{2a}\|\nabla f(x_{k+1})\|^2$$
$$\leq \frac{a\tau}{2}\frac{\dot{\beta}_{k-1}^2}{\beta_k^2}\|x^*\|^2 + \frac{s\delta^2\tau}{2a}\|\nabla f(x_{k+1})\|^2. \tag{2.12}$$

In the same way,

$$\begin{aligned} -\tau^2\left\langle \dot{x}_{\beta_{k-1}}, x_{k+1} - x_{\beta_k}\right\rangle &\leq \frac{\tau}{2}\left\|\dot{x}_{\beta_{k-1}}\right\|^2 + \frac{\tau^3}{2}\left\|x_{k+1} - x_{\beta_k}\right\|^2 \\ &\leq \frac{\tau}{2}\frac{\dot{\beta}_{k-1}^2}{\beta_k^2}\|x^*\|^2 + \frac{\tau^3}{2}\left\|x_{k+1} - x_{\beta_k}\right\|^2. \end{aligned} \tag{2.13}$$

By the strong convexity of ϕ_k, we obtain

$$\phi_k(x_{\beta_k}) - \phi_k(x_{k+1}) \geq -\left\langle \nabla\phi_k(x_{k+1}), x_{k+1} - x_{\beta_k}\right\rangle + \frac{c}{2\beta_k}\left\|x_{k+1} - x_{\beta_k}\right\|^2,$$

which is equivalent to

$$-s\tau\beta_k\left\langle \nabla\phi_k(x_{k+1}), x_{k+1} - x_{\beta_k}\right\rangle \leq -s\tau\beta_k\left(\phi_k(x_{k+1}) - \phi_k(x_{\beta_k})\right) - \frac{sc\tau}{2}\left\|x_{k+1} - x_{\beta_k}\right\|^2. \tag{2.14}$$

Similarly, $\phi_k(x_k) - \phi_k(x_{k+1}) \geq -\langle \nabla\phi_k(x_{k+1}), x_{k+1} - x_k\rangle + \frac{c}{2\beta_k}\|x_{k+1} - x_k\|^2$, which is equivalent to

$$-s\beta_k\langle \nabla\phi_k(x_{k+1}), \dot{x}_k\rangle \leq -s\beta_k\left(\phi_k(x_{k+1}) - \phi_k(x_k)\right) - \frac{sc}{2}\|x_{k+1} - x_k\|^2. \tag{2.15}$$

By replacing (2.11), (2.12), (2.13), (2.14), and (2.15), in formulation (2.9), and after simplifications, we deduce that

$$\begin{aligned} \langle \dot{v}_k, v_{k+1}\rangle \leq{}& \left(\tau - \alpha\sqrt{s} + \frac{\tau}{2a}\right)\|\dot{x}_k\|^2 + \frac{\tau}{2}(\tau^2 - sc)\left\|x_{k+1} - x_{\beta_k}\right\|^2 + \frac{\tau}{2}(2a+1)\frac{\dot{\beta}_{k-1}^2}{\beta_k^2}\|x^*\|^2 \\ &+ \frac{s\tau\delta^2}{2a}\|\nabla f(x_{k+1})\|^2 - s\beta_k\left(\phi_k(x_{k+1}) - \phi_k(x_k)\right) - s\tau\beta_k\left(\phi_k(x_{k+1}) - \phi_k(x_{\beta_k})\right) \\ &+ \tau(\tau - \alpha\sqrt{s})\left\langle \dot{x}_k, x_{k+1} - x_{\beta_k}\right\rangle + \delta\sqrt{s}(\tau - \alpha\sqrt{s})\langle \nabla f(x_{k+1}), \dot{x}_k\rangle \\ &- \delta s\sqrt{s}\beta_k\langle \nabla\phi_k(x_{k+1}), \nabla f(x_{k+1})\rangle. \end{aligned}$$

By combining this last inequality with the following two equalities

$$\begin{aligned} \delta\sqrt{s}(\tau - \alpha\sqrt{s})\langle \nabla f(x_{k+1}), \dot{x}_k\rangle = \frac{1}{2}(\tau - \alpha\sqrt{s})\Big(&\left\|\dot{x}_k + \delta\sqrt{s}\nabla f(x_{k+1})\right\|^2 \\ &-s\delta^2\|\nabla f(x_{k+1})\|^2 - \|\dot{x}_k\|^2\Big), \end{aligned}$$

and

$$-\delta s\sqrt{s}\beta_k\langle \nabla\phi_k(x_{k+1}), \nabla f(x_{k+1})\rangle = \frac{-\delta s\sqrt{s}\beta_k}{2}\left(\|\nabla\phi_k(x_{k+1})\|^2 + \|\nabla f(x_{k+1})\|^2 - \frac{c^2}{\beta_k^2}\|x_{k+1}\|^2\right),$$

we obtain

$$\begin{aligned}
&\langle \dot{v}_k, v_{k+1}\rangle \\
&\leq \frac{1}{2}\left(\tau-\alpha\sqrt{s}+\frac{\tau}{a}\right)\|\dot{x}_k\|^2+\frac{\tau}{2}(\tau^2-sc)\left\|x_{k+1}-x_{\beta_k}\right\|^2+\frac{\tau}{2}(2a+1)\frac{\dot{\beta}_{k-1}^2}{\beta_k^2}\|x^*\|^2 \\
&\quad -s\beta_k\left(\phi_k(x_{k+1})-\phi_k(x_k)\right)+\frac{s\delta}{2}\left(\frac{\tau\delta}{a}-\delta(\tau-\alpha\sqrt{s})-\sqrt{s}\beta_k\right)\|\nabla f(x_{k+1})\|^2 \\
&\quad +\frac{1}{2}(\tau-\alpha\sqrt{s})\left\|\dot{x}_k+\delta\sqrt{s}\nabla f(x_{k+1})\right\|^2-\frac{\delta s\sqrt{s}\beta_k}{2}\|\nabla\phi_k(x_{k+1})\|^2+\frac{\delta c^2 s\sqrt{s}}{2\beta_k}\|x_{k+1}\|^2 \\
&\quad -s\tau\beta_k\left(\phi_k(x_{k+1})-\phi_k(x_{\beta_k})\right)+\tau(\tau-\alpha\sqrt{s})\left\langle\dot{x}_k,x_{k+1}-x_{\beta_k}\right\rangle. \qquad (2.16)
\end{aligned}$$

Combining (2.16) with (2.8), and using that $\left\|x_{\beta_{k+1}}\right\|\leq\|x^*\|$, we obtain

$$\begin{aligned}
E_{k+1}-E_k &\leq \frac{1}{2}\left(\tau-\alpha\sqrt{s}+\frac{\tau}{a}\right)\|\dot{x}_k\|^2+\frac{\tau}{2}(\tau^2-sc)\left\|x_{k+1}-x_{\beta_k}\right\|^2 \\
&\quad +\frac{1}{2}\left(\tau(2a+1)\frac{\dot{\beta}_{k-1}^2}{\beta_k^2}+sc\left(\frac{\beta_{k+1}}{\beta_k}-1\right)\right)\|x^*\|^2 \\
&\quad +\frac{s\delta}{2}\left(\frac{\tau\delta}{a}-\delta(\tau-\alpha\sqrt{s})-\sqrt{s}\beta_k\right)\|\nabla f(x_{k+1})\|^2 \qquad (2.17) \\
&\quad +\frac{1}{2}(\tau-\alpha\sqrt{s})\left\|\dot{x}_k+\delta\sqrt{s}\nabla f(x_{k+1})\right\|^2+\frac{sc}{2\beta_k}(\delta c\sqrt{s}-\beta_{k+1}+\beta_k)\|x_{k+1}\|^2 \\
&\quad +s\left(\beta_{k+1}-(1+\tau)\beta_k\right)\left(\phi_k(x_{k+1})-\phi_k(x_{\beta_k})\right)-\frac{\delta s\sqrt{s}\beta_k}{2}\|\nabla\phi_k(x_{k+1})\|^2 \\
&\quad +\tau(\tau-\alpha\sqrt{s})\left\langle\dot{x}_k,x_{k+1}-x_{\beta_k}\right\rangle.
\end{aligned}$$

Let $\mu>0$. From the definition of E_k, we conclude that

$$\begin{aligned}
\mu E_{k+1} &= \mu s\beta_{k+1}\left(\phi_{k+1}(x_{k+1})-\phi_{k+1}(x_{\beta_{k+1}})\right)+\frac{\mu}{2}\left\|\tau(x_{k+1}-x_{\beta_k})+\dot{x}_k+\delta\sqrt{s}\nabla f(x_{k+1})\right\|^2 \\
&= \mu s\beta_{k+1}\left(\phi_{k+1}(x_{k+1})-\phi_{k+1}(x_{\beta_{k+1}})\right)+\frac{\mu\tau^2}{2}\left\|x_{k+1}-x_{\beta_k}\right\|^2 \\
&\quad +\frac{\mu}{2}\left\|\dot{x}_k+\delta\sqrt{s}\left(\nabla f(x_{k+1})+\frac{pc}{\beta_k}x_{k+1}\right)\right\|^2+\mu\tau\left\langle x_{k+1}-x_{\beta_k},\dot{x}_k+\delta\sqrt{s}\nabla f(x_{k+1})\right\rangle.
\end{aligned}$$

Using the fact that $\nabla f(x_{k+1})=\nabla\phi_k(x_{k+1})-\dfrac{c}{\beta_k}x_{k+1}$, we have

$$\begin{aligned}
\mu E_{k+1} &= \mu s\beta_{k+1}\left(\phi_k(x_{k+1})-\phi_k(x_{\beta_k})\right) \\
&\quad +\mu s\beta_{k+1}\left(\phi_{k+1}(x_{k+1})-\phi_k(x_{k+1})+\phi_k(x_{\beta_k})-\phi_{k+1}(x_{\beta_{k+1}})\right) \\
&\quad +\frac{\mu\tau^2}{2}\left\|x_{k+1}-x_{\beta_k}\right\|^2+\frac{\mu}{2}\left\|\dot{x}_k+\delta\sqrt{s}\nabla f(x_{k+1})\right\|^2+\mu\tau\left\langle x_{k+1}-x_{\beta_k},\dot{x}_k\right\rangle \\
&\quad +\mu\tau\delta\sqrt{s}\left\langle x_{k+1}-x_{\beta_k},\nabla\phi_k(x_{k+1})\right\rangle-\frac{\mu\tau\delta c\sqrt{s}}{\beta_k}\left\langle x_{k+1}-x_{\beta_k},x_{k+1}\right\rangle.
\end{aligned}$$

Since $\phi_{k+1}(x_{k+1}) - \phi_k(x_{k+1}) = \frac{c}{2}\left(\frac{1}{\beta_{k+1}} - \frac{1}{\beta_k}\right)\|x_{k+1}\|^2$, according to Lemma 2.2 *i*), we obtain

$$\begin{aligned}\mu E_{k+1} \le\ & \mu s\beta_{k+1}\left(\phi_k(x_{k+1}) - \phi_k(x_{\beta_k})\right) + \frac{\mu sc}{2}\left(\frac{\beta_{k+1}}{\beta_k} - 1\right)\left(\left\|x_{\beta_{k+1}}\right\|^2 - \|x_{k+1}\|^2\right) \\ & + \frac{\mu\tau^2}{2}\left\|x_{k+1} - x_{\beta_k}\right\|^2 + \frac{\mu}{2}\left\|\dot{x}_k + \delta\sqrt{s}\nabla f(x_{k+1})\right\|^2 + \mu\tau\left\langle x_{k+1} - x_{\beta_k}, \dot{x}_k\right\rangle \\ & + \mu\tau\delta\sqrt{s}\left\langle x_{k+1} - x_{\beta_k}, \nabla\phi_k(x_{k+1})\right\rangle - \frac{\mu\tau\delta c\sqrt{s}}{\beta_k}\left\langle x_{k+1} - x_{\beta_k}, x_{k+1}\right\rangle.\end{aligned}$$

Note that

$$\begin{aligned}& \frac{\mu\tau^2}{2}\left\|x_{k+1} - x_{\beta_k}\right\|^2 + \mu\tau\delta\sqrt{s}\left\langle x_{k+1} - x_{\beta_k}, \nabla\phi_k(x_{k+1})\right\rangle \\ & \le \mu\tau^2\left\|x_{k+1} - x_{\beta_k}\right\|^2 + \frac{\mu s\delta^2}{2}\|\nabla\phi_k(x_{k+1})\|^2.\end{aligned}$$

It follows that

$$\begin{aligned}\mu E_{k+1} \le\ & \mu s\beta_{k+1}\left(\phi_k(x_{k+1}) - \phi_k(x_{\beta_k})\right) + \frac{\mu sc}{2}\left(\frac{\beta_{k+1}}{\beta_k} - 1\right)\left(\|x^*\|^2 - \|x_{k+1}\|^2\right) \\ & + \mu\tau^2\left\|x_{k+1} - x_{\beta_k}\right\|^2 + \frac{\mu}{2}\left\|\dot{x}_k + \delta\sqrt{s}\nabla f(x_{k+1})\right\|^2 + \mu\tau\left\langle x_{k+1} - x_{\beta_k}, \dot{x}_k\right\rangle \\ & + \frac{\mu s\delta^2}{2}\|\nabla\phi_k(x_{k+1})\|^2 - \frac{\mu\tau\delta c\sqrt{s}}{\beta_k}\left\langle x_{k+1} - x_{\beta_k}, x_{k+1}\right\rangle. \qquad (2.18)\end{aligned}$$

By adding (2.17) with (2.18), we see that

$$\begin{aligned}& E_{k+1} - E_k + \mu E_{k+1} \\ & \le \frac{1}{2}\left(\tau - \alpha\sqrt{s} + \frac{\tau}{a}\right)\|\dot{x}_k\|^2 + \frac{\tau}{2}(2\mu\tau + \tau^2 - sc)\left\|x_{k+1} - x_{\beta_k}\right\|^2 \\ & + \frac{1}{2}\left(\tau(2a+1)\frac{\dot{\beta}_{k-1}^2}{\beta_k^2} + sc(\mu+1)\left(\frac{\beta_{k+1}}{\beta_k} - 1\right)\right)\|x^*\|^2 \\ & + \frac{s\delta}{2}\left(\mu\delta - \sqrt{s}\beta_k\right)\|\nabla\phi_k(x_{k+1})\|^2 \\ & + \frac{s\delta}{2}\left(\frac{\tau\delta}{a} - \delta(\tau - \alpha\sqrt{s}) - \sqrt{s}\beta_k\right)\|\nabla f(x_{k+1})\|^2 \\ & + \frac{1}{2}(\mu + \tau - \alpha\sqrt{s})\left\|\dot{x}_k + \delta\sqrt{s}\nabla f(x_{k+1})\right\|^2 \\ & + \tau(\mu + \tau - \alpha\sqrt{s})\left\langle \dot{x}_k, x_{k+1} - x_{\beta_k}\right\rangle \\ & + \frac{sc}{2\beta_k}\left(\delta c\sqrt{s} - (\mu+1)(\beta_{k+1} - \beta_k)\right)\|x_{k+1}\|^2 - \frac{\mu\tau\delta c\sqrt{s}}{\beta_k}\left\langle x_{k+1} - x_{\beta_k}, x_{k+1}\right\rangle \\ & + s\left((\mu+1)\beta_{k+1} - (1+\tau)\beta_k\right)\left(\phi_k(x_{k+1}) - \phi_k(x_{\beta_k})\right).\end{aligned}$$

In addition, we have

$$-\frac{\mu\tau\delta c\sqrt{s}}{\beta_k}\left\langle x_{k+1} - x_{\beta_k}, x_{k+1}\right\rangle = \frac{-\mu\tau\delta c\sqrt{s}}{2\beta_k}\left(\|x_{k+1}\|^2 + \left\|x_{k+1} - x_{\beta_k}\right\|^2 - \left\|x_{\beta_k}\right\|^2\right).$$

Hence

$$
\begin{aligned}
&(1+\mu)E_{k+1}-E_k \leq \\
&\frac{1}{2}\Bigg(\underbrace{\tau-\alpha\sqrt{s}+\frac{\tau}{a}}_{\mathbf{A}}\Bigg)\|\dot{x}_k\|^2+\frac{\tau}{2}\Bigg(\underbrace{2\mu\tau+\tau^2-sc}_{\mathbf{F}}\underbrace{-\frac{\mu\delta c\sqrt{s}}{\beta_k}}_{\leq 0}\Bigg)\left\|x_{k+1}-x_{\beta_k}\right\|^2 \\
&+\frac{sc}{2\beta_k}\Bigg(\underbrace{\delta c\sqrt{s}-(\mu+1)(\beta_{k+1}-\beta_k)-\frac{\mu\tau\delta}{\sqrt{s}}}_{\mathbf{B}}\Bigg)\|x_{k+1}\|^2 \\
&+\frac{1}{2}\left(\tau(2a+1)\frac{\dot{\beta}_{k-1}^2}{\beta_k^2}+sc(\mu+1)\left(\frac{\beta_{k+1}}{\beta_k}-1\right)+\frac{\mu\tau\delta c\sqrt{s}}{\beta_k}\right)\|x^*\|^2 \\
&+\frac{s\delta}{2}\Bigg(\underbrace{\frac{\tau\delta}{a}-\delta(\tau-\alpha\sqrt{s})-\sqrt{s}\beta_k}_{\mathbf{C}}\Bigg)\|\nabla f(x_{k+1})\|^2+\frac{1}{2}(\underbrace{\mu+\tau-\alpha\sqrt{s}}_{\mathbf{E}})\left\|\dot{x}_k+\delta\sqrt{s}\nabla f(x_{k+1})\right\|^2 \\
&+\frac{s\delta}{2}\Bigg(\underbrace{\mu\delta-\sqrt{s}\beta_k}_{\mathbf{D}}\Bigg)\|\nabla\phi_k(x_{k+1})\|^2+\tau(\underbrace{\mu+\tau-\alpha\sqrt{s}}_{\mathbf{E}})\left\langle\dot{x}_k,x_{k+1}-x_{\beta_k}\right\rangle \\
&+s\Bigg(\underbrace{(\mu+1)\beta_{k+1}-(1+\tau)\beta_k}_{\mathbf{G}}\Bigg)\Bigg(\underbrace{\phi_k(x_{k+1})-\phi_k(x_{\beta_k})}_{\geq 0}\Bigg).
\end{aligned}
$$

Taking $\mu=\dfrac{\alpha\sqrt{s}}{a+1}$ and $\tau=a\mu=\dfrac{a}{a+1}\alpha\sqrt{s}$, we obtain

- $\mathbf{A}=\mathbf{E}=0$;
- From condition $(\mathbf{H}_0)$, $\lim_{k\to+\infty}\beta_k=+\infty$ and $\lim_{k\to+\infty}\beta_{k+1}-\beta_k=+\infty$, then $\mathbf{C}\leq 0$, $\mathbf{B}\leq 0$ and $\mathbf{D}\leq 0$, for k large enough ;
- $\mathbf{G}=(\mu+1)\beta_k\left(\dfrac{\beta_{k+1}}{\beta_k}-\dfrac{a\mu+1}{\mu+1}\right)=(\mu+1)\beta_k\left(\dfrac{\beta_{k+1}}{\beta_k}-\dfrac{a(\alpha\sqrt{s}+1)+1}{a+\alpha\sqrt{s}+1}\right)\leq 0$, for k large enough (here we use (2.10));
- $\mathbf{F}=2\mu\tau+\tau^2-sc=s\left(\dfrac{a^2+2a}{(a+1)^2}\alpha^2-c\right)\leq 0$, because $c\geq\alpha^2\geq\dfrac{a^2+2a}{(a+1)^2}\alpha^2$.

Consequently, there exists $k_1\geq k_0$ such that, for all $k\geq k_1$,

$$(1+\mu)E_{k+1}-E_k\leq\lambda_k\|x^*\|^2, \tag{2.19}$$

where $\lambda_k=\dfrac{1}{2}\left(a\mu(2a+1)\dfrac{\dot{\beta}_{k-1}^2}{\beta_k^2}+(\mu+1)sc\left(\dfrac{\beta_{k+1}}{\beta_k}-1\right)+\dfrac{a\mu^2\delta c\sqrt{s}}{\beta_k}\right)$.

Inequality (2.19) is equivalent to $E_{k+1}+\left(-1+\dfrac{\mu}{1+\mu}\right)E_k\leq\dfrac{\lambda_k}{1+\mu}\|x^*\|^2$. Let us set $\rho:=$

$\frac{\mu}{1+\mu}$. By multiplying the last inequality by $e^{\rho(k+1)}$, we obtain that

$$e^{\rho(k+1)}E_{k+1}+(-1+\rho)e^{\rho(k+1)}E_k \leq \frac{e^{\rho(k+1)}\lambda_k}{1+\mu}\|x^*\|^2.$$

Therefore, for all $k \geq k_1$,

$$e^{\rho(k+1)}E_{k+1}-e^{\rho k}E_k+e^{\rho(k+1)}(e^{-\rho}-1+\rho)E_k \leq \frac{e^{\rho(k+1)}\lambda_k}{1+\mu}\|x^*\|^2.$$

Remarking that, for all $y \in \mathbb{R} : e^{-y}-1+y \geq 0$, we have $e^{\rho k}(e^{-\rho}-1+\rho)E_k \geq 0$, and then, for all $k \geq k_1$,

$$e^{\rho(k+1)}E_{k+1}-e^{\rho k}E_k \leq e^{\rho(k+1)}\theta_k\|x^*\|^2,$$

where

$$\theta_k = \frac{\lambda_k}{1+\mu} = \frac{1}{2}\left(a\rho(2a+1)\frac{\dot{\beta}_{k-1}^2}{\beta_k^2}+sc\left(\frac{\beta_{k+1}}{\beta_k}-1\right)+\frac{a\rho^2\delta c\sqrt{s}}{(1-\rho)\beta_k}\right).$$

Summing the above inequalities between k_1 and $k > k_1$, we obtain

$$e^{\rho(k+1)}E_{k+1}-e^{\rho k_1}E_{k_1} \leq \|x^*\|^2\left(\sum_{j=k_1}^{k} e^{\rho(j+1)}\theta_j\right).$$

Finally, by dividing by $e^{\rho(k+1)}$, we obtain (2.7). □

3. Strong Convergence of (IPATTH)

Throughout the rest of this paper, we note as in the previous proof $\dot{\beta}_k := \beta_{k+1}-\beta_k$.

Under suitable assumptions on (β_k), in this section, we can ensure: the convergence of values to $\min f$, the strong convergence of iterates towards x^*, and the convergence of the gradients and velocities to zero.

Theorem 3.1. *Let $f : \mathscr{H} \to \mathbb{R}$ be a convex function satisfying* ($\mathbf{H}_0'$)*, and (x_k) be a sequence generated by the algorithm* (**IPATTH**) *with $c \geq \alpha^2$. Suppose* ($\mathbf{H}_0$) *and*

$$\lim_{k\longrightarrow+\infty}\frac{\dot{\beta}_{k+1}}{\dot{\beta}_k} = \lim_{k\longrightarrow+\infty}\frac{\beta_{k+1}}{\beta_k} := \ell \in [1, 1+\alpha\sqrt{s}[. \qquad (\mathbf{H}_2)$$

Then, for k large enough,

i) $f(x_k)-\min_{\mathscr{H}} f = \mathscr{O}\left(\frac{1}{\beta_k}\right)$.

If moreover $\ell = 1$, then we have simultaneously, the convergence of values to $\min_{\mathscr{H}} f$, the strong convergence of iterates to x^*, and the strong convergence of gradient and velocities to 0*, with the following rates:*

$$ii)\quad f(x_k)-\min_{\mathscr{H}} f = o\left(\frac{1}{\beta_k}\right);\ \|x_k-x_{\beta_k}\|^2 = \mathscr{O}\left(\frac{\dot{\beta}_k}{\beta_k}+e^{-\rho k}\right);$$

$$\|\nabla f(x_k)\|^2 = \mathscr{O}\left(\frac{\dot{\beta}_k}{\beta_k}+e^{-\rho k}+\frac{1}{\beta_k^2}\right);\ \textit{and}\ \|\dot{x}_{k-1}\|^2 = \mathscr{O}\left(\frac{\dot{\beta}_k}{\beta_k}+e^{-\rho k}+\frac{1}{\beta_k^2}\right).$$

Proof. From (2.7), we have

$$E_{k+1} \leq e^{(k_1-k-1)\rho} E_{k_1} + N_k \|x^*\|^2 \tag{3.1}$$

with

$$N_k = \frac{sc}{2e^{\rho(k+1)}} \sum_{j=k_1+1}^{k} e^{\rho(j+1)} \left(r\frac{\dot{\beta}_{j-1}^2}{\beta_j^2} + \frac{\dot{\beta}_j}{\beta_j} + \frac{b}{\beta_j} \right), \tag{3.2}$$

where $r := \dfrac{a\rho(2a+1)}{sc}$ and $b := \dfrac{a\rho^2\delta}{(1-\rho)\sqrt{s}}$. Set, for each $j > k_1$,

$$\Gamma_j = \frac{1}{r\dfrac{\dot{\beta}_{j-1}^2 \beta_{j+1}}{\beta_j^2 \dot{\beta}_{j+1}} + \dfrac{\dot{\beta}_j \beta_{j+1}}{\dot{\beta}_{j+1}\beta_j} + \dfrac{b\beta_{j+1}}{\dot{\beta}_{j+1}\beta_j}} \times \left(e^\rho - \frac{\dot{\beta}_j \beta_{j+1}}{\dot{\beta}_{j+1}\beta_j} \right).$$

Let us show that, for j large enough, $\Gamma_j > 0$. Using conditions $(\mathbf{H}_0)$ and $(\mathbf{H}_2)$, we obtain

$$\begin{aligned}
&\lim_{j\to+\infty} r\frac{\dot{\beta}_{j-1}^2 \beta_{j+1}}{\beta_j^2 \dot{\beta}_{j+1}} + \frac{\dot{\beta}_j \beta_{j+1}}{\dot{\beta}_{j+1}\beta_j} + \frac{b\beta_{j+1}}{\dot{\beta}_{j+1}\beta_j} \\
&= \lim_{j\to+\infty} \frac{\beta_{j+1}}{\beta_j} \left(r\left(\frac{\dot{\beta}_{j-1}}{\dot{\beta}_j}\right)^2 \times \frac{\dot{\beta}_j}{\dot{\beta}_{j+1}} \times \frac{\dot{\beta}_j}{\beta_j} + \frac{\dot{\beta}_j}{\dot{\beta}_{j+1}} + \frac{b}{\dot{\beta}_{j+1}} \right) \\
&= \ell \left(\frac{r}{\ell^2} \times \frac{1}{\ell} \times (\ell - 1) + \frac{1}{\ell} \right) \\
&= \frac{r(\ell-1)}{\ell^2} + 1 \geq 1,
\end{aligned}$$

and we also have $\displaystyle\lim_{j\to+\infty} \left(e^\rho - \frac{\dot{\beta}_j \beta_{j+1}}{\dot{\beta}_{j+1}\beta_j} \right) = e^\rho - 1 > 0$. Therefore,

$$\lim_{j\to+\infty} \Gamma_j = \frac{1}{1 + \frac{r(\ell-1)}{\ell^2}} (e^\rho - 1) > 0.$$

Hence, for j large enough $\Gamma_j > 0$. Consequently, There exists $m > 0$ such that, for j large enough, $0 < m \leq \Gamma_j$, which implies that, for j large enough,

$$m \left(r\frac{\dot{\beta}_{j-1}^2 \beta_{j+1}}{\beta_j^2 \dot{\beta}_{j+1}} + \frac{\dot{\beta}_j \beta_{j+1}}{\dot{\beta}_{j+1}\beta_j} + \frac{b\beta_{j+1}}{\dot{\beta}_{j+1}\beta_j} \right) \leq e^\rho - \frac{\dot{\beta}_j \beta_{j+1}}{\dot{\beta}_{j+1}\beta_j}.$$

Multiplying by $\dfrac{\dot{\beta}_{j+1}}{m\beta_{j+1}} e^{\rho(j+1)}$, we obtain

$$e^{\rho(j+1)} \left(r\frac{\dot{\beta}_{j-1}^2}{\beta_j^2} + \frac{\dot{\beta}_j}{\beta_j} + \frac{b}{\beta_j} \right) \leq \frac{\dot{\beta}_{j+1}}{m\beta_{j+1}} e^{\rho(j+2)} - \frac{\dot{\beta}_j}{m\beta_j} e^{\rho(j+1)}.$$

Using (3.2), we deduce that there exists $k_2 > k_1 + 1$ such that, for all $k \geq k_2$,

$$\begin{aligned} N_k &\leq \frac{sc}{2e^{\rho(k+1)}}\left[C_1 + \sum_{j=k_2}^{k}\left(\frac{\dot\beta_{j+1}}{m\beta_{j+1}}e^{\rho(j+2)} - \frac{\dot\beta_j}{m\beta_j}e^{\rho(j+1)}\right)\right] \\ &= \frac{sc}{2e^{\rho(k+1)}}\left(C_1 + \frac{\dot\beta_{k+1}}{m\beta_{k+1}}e^{\rho(k+2)} - \frac{\dot\beta_{k_2}}{m\beta_{k_2}}e^{\rho(k_2+1)}\right) \\ &\leq \frac{C_1 sc}{2}e^{-\rho(k+1)} + \frac{sce^{\rho}}{2m}\frac{\dot\beta_{k+1}}{\beta_{k+1}}. \end{aligned}$$

where

$$C_1 = \sum_{j=k_1+1}^{k_2-1} e^{\rho(j+1)}\left(r\frac{\dot\beta_{j-1}^2}{\beta_j^2} + \frac{\dot\beta_j}{\beta_j} + \frac{b}{\beta_j}\right).$$

Returning to (3.1), we deduce that for k large enough

$$E_{k+1} = \mathcal{O}\left(\frac{\dot\beta_{k+1}}{\beta_{k+1}} + e^{-\rho(k+1)}\right). \tag{3.3}$$

i) From (2.4) and (3.3), there exist a positive constant C_2 such that for k large enough

$$f(x_k) - \min_{\mathcal{H}} f \leq \frac{1}{\beta_k}\left(E_k + \frac{c}{2}\|x^*\|^2\right) \leq \frac{1}{\beta_k}\left[C_2\left(\frac{\dot\beta_k}{\beta_k} + e^{-\rho k}\right) + \frac{c}{2}\|x^*\|^2\right].$$

Condition **(H_2)** means that $\left(\frac{\dot\beta_k}{\beta_k}\right)$ is bounded, and we have $\lim_{k\to+\infty} e^{-\rho k} = 0$. Thus, for k large enough, $f(x_k) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{\beta_k}\right)$.

ii) Now, assume that $\lim_{k\to+\infty}\frac{\beta_{k+1}}{\beta_k} = 1$ $\left(\text{i.e. } \lim_{k\to+\infty}\frac{\dot\beta_k}{\beta_k} = 0\right)$, which implies that $\lim_{k\to+\infty} E_k = 0$. From Lemma 2.1, we have $\left\|x_k - x_{\beta_k}\right\|^2 \leq \frac{2}{sc}E_k$. Thus it follows from (3.3) that, for k large enough,

$$\left\|x_k - x_{\beta_k}\right\|^2 = \mathcal{O}\left(\frac{\dot\beta_k}{\beta_k} + e^{-\rho k}\right), \tag{3.4}$$

which ensures the strong convergence of (x_k) to x^* and consequently the convergence to zero of $\left(\|x^*\|^2 - \|x_k\|^2\right)$. Using again Lemma 2.1, we have

$$f(x_k) - \min_{\mathcal{H}} f = o\left(\frac{1}{\beta_k}\right), \text{ as } k \to +\infty.$$

Applying the Lipschitz continuity of ∇f, (2.1) and (2.2), we have

$$\begin{aligned} \|\nabla f(x_k)\| &\leq \|\nabla f(x_k) - \nabla f(x_{\beta_k})\| + \|\nabla f(x_{\beta_k})\| \\ &\leq L\|x_k - x_{\beta_k}\| + \frac{c}{\beta_k}\|x_{\beta_k}\| \\ &\leq L\|x_k - x_{\beta_k}\| + \frac{c}{\beta_k}\|x^*\|. \end{aligned}$$

Using (3.4), we conclude

$$\|\nabla f(x_k)\|^2 = \mathcal{O}\left(\frac{\dot{\beta}_k}{\beta_k} + e^{-\rho k} + \frac{1}{\beta_k^2}\right), \; as\; k \to +\infty.$$

It follows that

$$\begin{aligned}\left\|\dot{x}_{k-1} + \delta\sqrt{s}\nabla f(x_k)\right\|^2 &\le 2\left\|\tau\left(x_k - x_{\beta_{k-1}}\right) + \dot{x}_{k-1} + \delta\sqrt{s}\nabla f(x_k)\right\|^2 + 2\tau^2\left\|x_k - x_{\beta_{k-1}}\right\|^2 \\ &\le 2\left\|\tau\left(x_k - x_{\beta_{k-1}}\right) + \dot{x}_{k-1} + \delta\sqrt{s} f(x_k)\right\|^2 \\ &\quad + 4\tau^2\left(\left\|x_k - x_{\beta_k}\right\|^2 + \left\|x_{\beta_k} - x_{\beta_{k-1}}\right\|^2\right).\end{aligned}$$

By definition of E_k and (3.3), we obtain as $k \to +\infty$

$$2\left\|\tau\left(x_k - x_{\beta_{k-1}}\right) + \dot{x}_{k-1} + \delta\sqrt{s}\nabla f(x_k)\right\|^2 \le 4E_k = \mathcal{O}\left(\frac{\dot{\beta}_k}{\beta_k} + e^{-\rho k}\right). \tag{3.5}$$

From Lemma 2.2, we have $\left\|x_{\beta_k} - x_{\beta_{k-1}}\right\| \le \frac{\dot{\beta}_{k-1}}{\beta_k}\|x^*\|$. Therefore, by using (3.4), we obtain

$$\left\|\dot{x}_{k-1} + \delta\sqrt{s}\nabla f(x_k)\right\|^2 \le C_3\left(\frac{\dot{\beta}_k}{\beta_k} + e^{-\rho k}\right) + 4\tau^2\left(\frac{\dot{\beta}_{k-1}}{\beta_k}\right)^2\|x^*\|^2,$$

where C_3 is a positive constant. For k large enough, $\left(\frac{\dot{\beta}_{k-1}}{\beta_k}\right)^2 = \mathcal{O}\left(\frac{\dot{\beta}_k}{\beta_k}\right)$. Indeed, using **(H$_2$)**, we obtain

$$\begin{aligned}\lim_{k\to+\infty}\left(\frac{\dot{\beta}_{k-1}}{\beta_k}\right)^2 \times \frac{\beta_k}{\dot{\beta}_k} = \lim_{k\to+\infty}\frac{\dot{\beta}_{k-1}}{\dot{\beta}_k} \times \frac{\dot{\beta}_{k-1}}{\beta_k} &= \lim_{k\to+\infty}\frac{\dot{\beta}_{k-1}}{\dot{\beta}_k} \times \left(1 - \frac{\beta_{k-1}}{\beta_k}\right) \\ &= \frac{1}{\ell}\left(1 - \frac{1}{\ell}\right) \le 1.\end{aligned}$$

Therefore, we conclude that, for k large enough,

$$\left\|\dot{x}_{k-1} + \delta\sqrt{s}\nabla f(x_k)\right\|^2 = \mathcal{O}\left(\frac{\dot{\beta}_k}{\beta_k} + e^{-\rho k}\right). \tag{3.6}$$

Returning to

$$\|\dot{x}_{k-1}\|^2 \le 2\left\|\dot{x}_{k-1} + \delta\sqrt{s}\nabla f(x_k)\right\|^2 + 2s\delta^2\|\nabla f(x_k)\|^2,$$

and using (3) and (3.6), we conclude that $\|\dot{x}_{k-1}\|^2 = \mathcal{O}\left(\frac{\dot{\beta}_k}{\beta_k} + e^{-\rho k} + \frac{1}{\beta_k^2}\right)$, *as* $k \to +\infty$. □

4. Particular Rases

In this section, we investigate two particular examples of β_k to illustrate the results of the previous sections.

4.1. **Case $\beta_k = k^r \ln^q(k)$.** Let us investigate the case $\beta_k = k^r \ln^q(k)$ for $r \geq 1$ and $q \geq 0$.

Proposition 4.1. *Let (x_k) be a sequence generated via the algorithm (IPATTH). If either $\beta_k = k^r \ln^q(k)$, where $r > 1$ or $r = 1$ and $q \neq 0$, $\alpha > 0$, and $c \geq \alpha^2$. Then, as $k \to +\infty$,*

$$i)\ f(x_k) - \min_{\mathscr{H}} f = o\left(\frac{1}{k^r \ln^q(k)}\right) \qquad ii)\ \left\|x_k - x_{\beta_k}\right\|^2 = \mathscr{O}\left(\frac{1}{k}\right)$$

$$iii)\ \|\nabla f(x_k)\|^2 = \mathscr{O}\left(\frac{1}{k}\right) \qquad iv)\ \|\dot{x}_{k-1}\|^2 = \mathscr{O}\left(\frac{1}{k}\right).$$

Proof. To fulfil the conditions of Theorem 3.1, we first note that when $k \to +\infty$, and for all $a > 0$, we have

$$\left(1+\frac{a}{k}\right)^r = 1 + \frac{ar}{k} + o\left(\frac{1}{k}\right) \quad \text{and} \quad \ln\left(1+\frac{a}{k}\right) = \frac{a}{k} + o\left(\frac{1}{k}\right).$$

Then

$$\begin{aligned}\left(1+\tfrac{a}{k}\right)^r \left(1+\tfrac{\ln(1+\frac{a}{k})}{\ln(k)}\right)^q &= \left(1+\tfrac{ar}{k}+o\left(\tfrac{1}{k}\right)\right)\left(1+\tfrac{a}{k\ln(k)}+o\left(\tfrac{1}{k\ln(k)}\right)\right)^q \\ &= \left(1+\tfrac{ar}{k}+o\left(\tfrac{1}{k}\right)\right)\left(1+\tfrac{aq}{k\ln(k)}+o\left(\tfrac{1}{k\ln(k)}\right)\right);\end{aligned}$$

which gives

$$\left(1+\frac{a}{k}\right)^r \left(1+\frac{\ln(1+\frac{a}{k})}{\ln(k)}\right)^q = 1 + \frac{ar}{k} + \frac{aq}{k\ln(k)} + o\left(\frac{1}{k\ln(k)}\right). \tag{4.1}$$

Now let us check that (β_k) satisfies all conditions $(\mathbf{H}_0)$ and $(\mathbf{H}_2)$. We have

- $\lim_{k\to+\infty} \beta_k = \lim_{k\to+\infty} k^r \ln^q(k) = +\infty$;
- Using (4.1), when $k \to +\infty$, we obtain

$$\begin{aligned}\beta_{k+1} - \beta_k &= (k+1)^r \ln^q(k+1) - k^r \ln^q(k) \\ &= k^r \ln^q(k)\left(\left(1+\frac{1}{k}\right)^r \left(1+\frac{\ln\left(1+\frac{1}{k}\right)}{\ln(k)}\right)^q - 1\right) \\ &= k^r \ln^q(k)\left(\frac{r}{k} + \frac{q}{k\ln(k)} + o\left(\frac{1}{k\ln(k)}\right)\right) \\ &= rk^{r-1}\ln^q(k) + qk^{r-1}\ln^{q-1}(k) + o\left(k^{r-1}\ln^{q-1}(k)\right).\end{aligned}$$

  Hence, if $r > 1$ or $(r = 1 \text{ and } q > 0)$, then $\lim_{k\to+\infty} \beta_{k+1} - \beta_k = +\infty$.

Therefore, $(\mathbf{H}_0)$ is verified.

- $\dfrac{\beta_{k+1}}{\beta_k} = \dfrac{(k+1)^r \ln^q(k+1)}{k^r \ln^q(k)} = \left(1+\dfrac{1}{k}\right)^r \left(1+\dfrac{\ln(1+\frac{1}{k})}{\ln(k)}\right)^q$;

  using (4.1), we get, when $k \to +\infty$, $\dfrac{\beta_{k+1}}{\beta_k} = 1 + \frac{r}{k} + \frac{q}{k\ln(k)} + o\left(\frac{1}{k\ln(k)}\right) \underset{+\infty}{\sim} 1.$

- On the other hand,

$$\frac{\dot{\beta}_{k+1}}{\dot{\beta}_k} = \frac{(k+2)^r \ln^q(k+2) - (k+1)^r \ln^q(k+1)}{(k+1)^r \ln^q(k+1) - k^r \ln^q(k)}$$
$$= \frac{\left(1+\frac{2}{k}\right)^r \left(1+\frac{\ln(1+\frac{2}{k})}{\ln(k)}\right)^q - \left(1+\frac{1}{k}\right)^r \left(1+\frac{\ln(1+\frac{1}{k})}{\ln(k)}\right)^q}{\left(1+\frac{1}{k}\right)^r \left(1+\frac{\ln(1+\frac{1}{k})}{\ln(k)}\right)^q - 1}.$$

Therefore, by using again (4.1), we see that when $k \to +\infty$

$$\frac{\dot{\beta}_{k+1}}{\dot{\beta}_k} = \frac{\left(1+\frac{2r}{k}+\frac{2q}{k\ln(k)}+o\left(\frac{1}{k\ln(k)}\right)\right) - \left(1+\frac{r}{k}+\frac{q}{k\ln(k)}+o\left(\frac{1}{k\ln(k)}\right)\right)}{\left(1+\frac{r}{k}+\frac{q}{k\ln(k)}+o\left(\frac{1}{k\ln(k)}\right)\right) - 1}$$
$$= \frac{r\ln(k)+q+o(1)}{r\ln(k)+q+o(1)}.$$

Thus $\lim\limits_{k\to+\infty} \dfrac{\dot{\beta}_{k+1}}{\dot{\beta}_k} = 1.$

Consequently condition $(\mathbf{H}_2)$ is satisfied. Then by applying Theorem 3.1, we obtain the assertions i), ii), iii), and iv). □

4.2. **Case $\beta_k = e^{\gamma k^r}$.** Let us now treat the case $\beta_k = e^{\gamma k^r}$ with $r \in]0,1]$ and $\gamma > 0$. Remark that, when $k \to +\infty$,

$$(k+1)^r - k^r = k^r \left[\left(1+\frac{1}{k}\right)^r - 1\right] = rk^{r-1} + o(k^{r-1}). \tag{4.2}$$

Hence, for all $a > 0$,

$$e^{\gamma((k+a)^r - k^r)} = e^{a\gamma r k^{r-1} + o(k^{r-1})} = 1 + a\gamma r k^{r-1} + o(k^{r-1}). \tag{4.3}$$

We have

- $\dot{\beta}_k \neq 0$ *for all* $k > 0$;
- $(\beta_k)_{k\geq 0}$ *is a nondecreasing sequence, with* $\lim\limits_{k\to+\infty} \beta_k = +\infty$;
- It follows fro (4.2) that

$$\frac{\beta_{k+1}}{\beta_k} = e^{\gamma((k+1)^r - k^r)} = e^{\gamma r k^{r-1} + o(k^{r-1})} \xrightarrow[k\to+\infty]{} \begin{cases} e^{\gamma} & \text{if } r = 1 \\ 1 & \text{if } 0 < r < 1. \end{cases}$$

- If $0 < r < 1$, by using (4.3), we see then, when $k \to +\infty$,

$$\frac{\dot{\beta}_{k+1}}{\dot{\beta}_k} = \frac{e^{\gamma(k+2)^r} - e^{\gamma(k+1)^r}}{e^{\gamma(k+1)^r} - e^{\gamma k^r}} = \frac{e^{\gamma((k+2)^r - k^r)} - e^{\gamma((k+1)^r - k^r)}}{e^{\gamma((k+1)^r - k^r)} - 1}$$
$$= \frac{\gamma r k^{r-1} + o(k^{r-1})}{\gamma r k^{r-1} + o(k^{r-1})}.$$

Therefore, $\lim\limits_{k\to+\infty} \dfrac{\dot{\beta}_{k+1}}{\dot{\beta}_k} = 1$. In the case that $r = 1$,

$$\lim_{k\to+\infty} \frac{\dot{\beta}_{k+1}}{\dot{\beta}_k} = \lim_{k\to+\infty} \frac{e^{\gamma(k+1)}}{e^{\gamma k}} = e^{\gamma}.$$

- $\beta_{k+1}-\beta_k = e^{\gamma k^r}\left(e^{\gamma((k+1)^r-k^r)}-1\right) = e^{\gamma k^r}\left(\gamma r k^{r-1}+o(k^{r-1})\right)$.
  Hence $\lim_{k\to+\infty}\beta_{k+1}-\beta_k = +\infty$.

Consequently, (β_k) fulfils assumption $(\mathbf{H}_0)$ and condition $(\mathbf{H}_2)$ is verified if and only if $0 < r < 1$ or $(r=1, \text{ and } \gamma < \ln(1+\alpha\sqrt{s}))$. Then by applying Theorem 3.1, we have the following proposition.

Proposition 4.2. *Let f satisfy $(\mathbf{H}_0')$, and let (x_k) be a sequence generated by the algorithm* (**IPATTH**)*, where $\beta_k = e^{\gamma k^r}$, $r \in]0,1]$, $\gamma > 0$, $\alpha > 0$ and $c \geq \alpha^2$. Then:*

- *If $r = 1$, $\gamma < \ln(1+\alpha\sqrt{s})$, then $f(x_k) - \min_{\mathscr{H}} f = \mathscr{O}\left(e^{-\gamma k}\right)$ as $k \to +\infty$.*
- *If $0 < r < 1$, then, as $k \to +\infty$,*

$$f(x_k) - \min_{\mathscr{H}} f = o\left(e^{-\gamma k^r}\right); \ \left\|x_k - x_{\beta_k}\right\|^2 = \mathscr{O}\left(\frac{1}{k^{1-r}}\right);$$

$$\|\nabla f(x_k)\|^2 = \mathscr{O}\left(\frac{1}{k^{1-r}}\right); \ \|\dot{x}_{k-1}\|^2 = \mathscr{O}\left(\frac{1}{k^{1-r}}\right).$$

5. Numerical Example

Consider the convex differentiable function $f : \mathbb{R}_+^* \times \mathbb{R}_+^* \to \mathbb{R}$ defined by

$$f(x,y) = 2x^2 + x - \frac{3}{2}\ln(x) + \frac{1}{2}y^2 - 3\ln(y).$$

We have $\nabla f(x,y) = \begin{pmatrix} 4x+1-\frac{3}{2x} \\ y - \frac{3}{y} \end{pmatrix}$ and $x^* = \left(\frac{1}{2}, \sqrt{3}\right)$ is the unique global minimizer of f, moreover $\min_{(\mathbb{R}_+^*)^2} f = \frac{5}{2} + \frac{3}{2}\ln\left(\frac{2}{3}\right)$. Further, for all $\lambda > 0$,

$$\operatorname{prox}_{\lambda f}(x,y) = \left(\frac{x-\lambda+\sqrt{(x-\lambda)^2+6\lambda(1+4\lambda)}}{2(1+4\lambda)}, \frac{y+\sqrt{y^2+12\lambda(1+\lambda)}}{2(1+\lambda)}\right).$$

In Figure 1, we compare the convergence rate associated with our algorithm (**IPATTH**), under different choices of β_k, to that studied by Attouch, Balhag, Chbani and Riahi in [33]. The latter represents an implicit discretization of system (1.2) and is defined as follows

$$\begin{cases} \text{Step } k: \text{ set } d_k = \dfrac{1}{1+\alpha\sqrt{s\varepsilon_k}+s\varepsilon_k} \ \text{ and } \ \lambda_k = \dfrac{s+\delta\sqrt{s}}{1+\alpha\sqrt{s\varepsilon_k}+s\varepsilon_k} \\ y_k = x_k + d_k(x_k - x_{k-1} + \delta\sqrt{s}\nabla f(x_k)) \\ x_{k+1} = \operatorname{prox}_{\lambda_k f}(y_k - s\varepsilon_k d_k x_k), \end{cases} \tag{5.1}$$

where $\varepsilon_k = t^{-r}$ with $1 < r < 2$.

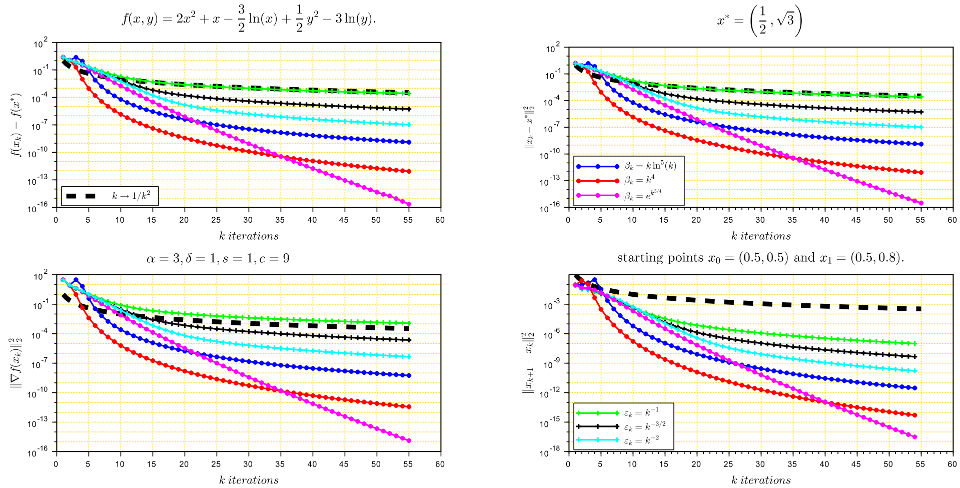

FIGURE 1. Comparison of convergence rates for solutions associated with the algorithms (**IPATTH**) and (5.1). We also add the path trajectory x_k for different sequence β_k.

REFERENCES

[1] B. Martinet, Régularisation d'inéquations variationnelles par approximations successives, Rev. Franaise Informat. Recherche Opérationnelle 4 (1970), 154-158 (in French).

[2] RT. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Control Optim. 14 (1976), 877–898.

[3] O. Güler, On the convergence of the proximal point algorithm for convex minimization, SIAM J. Control Optim. 29 (1991), 403-419.

[4] S. Kamimura, W. Takahashi, Strong convergence of a proximal-type algorithm in a Banach space, SIAM J. Optim. 13 (2003), 938-945.

[5] N. Lehdili, A. Moudafi, Combining the proximal algorithm and Tikhonov regularization, Optimization 37 (1996), 239-252.

[6] S. Reich, Strong convergence theorems for resolvents of accretive operators in Banach spaces, J. Math. Anal. Appl. 183 (1994), 118-120.

[7] MV. Solodov, BF. Svaiter, Forcing strong convergence of proximal point iterations in a Hilbert space, Math. Program., Ser. A 87 (2000), 189-202.

[8] X. Qin, S.M. Kang, Y.J. Cho, Approximating zeros of monotone operators by proximal point algorithms, J. Glob. Optim. 46 (2010), 75-87.

[9] Y. Wang, H.K. Xu, Strong convergence for the proximal-gradient method, J Nonlinear Convex Anal. 15 (2014) 581–593.

[10] H. Attouch, A. Damlamian, Strong solutions for parabolic variational inequalities, Nonlinear Anal. 2 (1978), 329–353.

[11] V. Barbu, Nonlinear Differential Equations of Monotone Types in Banach Spaces, Springer, 2010.

[12] A. Haraux, Nonlinear evolution equations-Global behavior of solutions, Lecture note in Mathematics, 841, Springer-Verlag, 1982.

[13] R.E. Showalter, Monotone operators in Banach space and nonlinear partial differential equation, Math. Surveys Monogr. 49, Amer. Math. Soc., 1997.

[14] B. Polyak, Some methods of speeding up the convergence of iteration methods, USSR Comput. Math. Math. Phys. 4 (1964), 1-17.

[15] B. Polyak, Introduction to Optimization, Optimization Software-Inc, New York, NY, 1987.

[16] A. Alvarez, On the minimizing property of a second order dissipative system in Hilbert space, SIAM J. Control Optim. 38 (2000), 1102–1119.

[17] H. Attouch, M.-O. Czarnecki, Asymptotic control and stabilization of nonlinear oscillators with non-isolated equilibria, J. Differential Equations 179 (2002), 278–310.
[18] H. Attouch, M.-O. Czarnecki, Asymptotic behavior of gradient-like dynamical systems involving inertia and multiscale aspects, J. Differential Equations 262 (2017), 2745-2770.
[19] H. Attouch, Z. Chbani, H. Riahi, Combining fast inertial dynamics for convex optimization with Tikhonov regularization, J. Math. Anal. Appl, 457 (2018), 1065–1094.
[20] W. Su, S. Boyd, EJ. Candès, A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights, J. Mach. Learn. Res. 17 (2016), 5312-5354.
[21] Y. Nesterov, A method of solving a convex programming problem with convergence rate $O(1/k^2)$, Soviet Math. Dokl. 27 (1983), 372-376.
[22] Y. Nesterov, Introductory lectures on convex optimization: A basic course, volume 87 of Applied Optimization, Kluwer Academic Publishers, Boston, MA, 2004.
[23] Y. Nesterov, How to Make the Gradients Small, Optima 88, Mathematical Optimization Society Newsletter, May 2012.
[24] Y. Nesterov, Gradient methods for minimizing composite functions, Math. Program. 140 (2013) 125-161.
[25] V. Apidopoulos, Ch. Aujol, JF. Dossal, Convergence rate of inertial Forward-Backward algorithm beyond Nesterov's rule, Math. Program. Ser. B 180 (2020), 137–156 .
[26] H. Attouch, Z. Chbani, J. Peypouquet, P. Redont, Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity, Math. Program. Ser. B 168 (2018), 123-175.
[27] H. Attouch, Z. Chbani, H. Riahi, Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$, ESAIM Control Optim. Calc. Var. 25 (2019), 2.
[28] H. Attouch, J. Peypouquet, P. Redont, The rate of convergence of Nesterov's accelerated forward–backward method is actually faster than $1/k^2$, SIAM J. Optim. 26 (2016), 1824–1834.
[29] A. Chambolle, C. Dossal, On the convergence of the iterates of the Fast Iterative Shrinkage Thresholding Algorithm, J. Optim. Theory Appl. 166 (2015), 968–982.
[30] R. May, Asymptotic for a second order evolution equation with convex potential and vanishing damping term, Turk. J. Math. 41 (2017), 681-685.
[31] H. Attouch H, S. László, Convex optimization via inertial algorithms with vanishing Tikhonov regularization: fast convergence to the minimum norm solution, arXiv:2104.11987v1, 2021.
[32] H. Attouch, A. Balhag, Z. Chbani, H. Riahi, Damped inertial dynamics with vanishing Tikhonov regularization: Strong asymptotic convergence towards the minimum norm solution, J. Differential Equations 311 (2022), 29–58.
[33] H. Attouch, A. Balhag, Z. Chbani, H. Riahi, Accelerated gradient methods combining Tikhonov regularization with geometric damping driven by the Hessian, Appl. Math. Optim. 88 (2023), 29.
[34] A.C.Bagy, Z. Chbani, H. Riahi, Strong convergence of trajectories via inertial dynamics combining Hessian-driven damping and Tikhonov regularization for general convex minimizations, Numer. Funct. Anal. Optim. 44 (2023), 1481-1509.
[35] AC. Bagy, Z. Chbani, H. Riahi, The Heavy ball method regularized by Tikhonov term. Simultaneous convergence of values and trajectories, Evol. Equ. Control Theory 12 (2023), 687–702.
[36] H. Attouch, A. Balhag, Z. Chbani, H. Riahi, Fast convex optimization via inertial dynamics combining viscous and Hessian-driven damping with time rescaling, Evol. Equ. Control Theory 11 (2022), 487-514.
[37] H. Attouch, Z. Chbani, H. Riahi, Fast convex optimization via time scaling of damped inertial gradient dynamics, SIAM J.Optim. 29 (3) (2019), 2227-2256.