Inertial forward-backward methods with subgradient-based corrections [*]

Milan Barik[†] Suvendu Ranjan Pattanaik[‡]

Abstract

Shi et al. [34] propose acceleration methods to solve smooth convex optimization problems. In our work, we focus on the general unconstrained composite non-smooth convex optimization problem. We provide an inertial forward-backward algorithm with subgradient correction, derived through time discretization of the ODE, as studied by Shi et al. We achieve the rate of convergence of the objective gap as $\mathcal{O}\left(\frac{1}{t^2}\right)$ and the $o\left(\frac{1}{t^3}\right)$ rate of convergence of the squared subdifferential norm for $\alpha \geq 3$. When $\alpha > 3$, the rate of objective gap is improved to $o\left(\frac{1}{t^2}\right)$, and also the iterative sequence generated by the algorithm converges to a minimal point. Furthermore, we analyze the inexact version of the proposed algorithm. The effectiveness of the proposed method has been demonstrated through various numerical studies.

Keywords. *Composite Convex optimization problem, Inertial forward-backward algorithm, Inexactness, Subgradient correction*

MSC codes. *37N40, 49M37, 65K05, 90C25*

1 Introduction

In this work, we focus on the unconstrained optimization problem

$$\min_{u \in \mathbb{R}^d} f(u), \tag{1}$$

where f is a real valued function on $\mathbb{R}^d$ with $f = g + h$ such that $g : \mathbb{R}^d \to \mathbb{R}$ is continuously differentiable and convex, and $h : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is a lower semicontinuous (l.s.c.) proper convex function. For application of Problem (1), one may see [17, 18, 31, 36].

---

[*] Submitted to the editors DATE.

[†] Department of Mathematics, NIT Rourkela, India (milanbarik7008@gmail.com).

[‡] Department of Mathematics, NIT Rourkela, India (suvendu.pattanaik@gmail.com).

A numerical technique for determining the optimal approximate solution to Problem (1) is the proximal gradient method (PGM). The general update rule of PGM is given by

$$u_{t+1} = \mathrm{Prox}_{\lambda h}(u_t - \lambda \nabla g(u_t)), \tag{2}$$

where λ is a positive real number and Prox denotes the proximal operator [9, 13] on $\mathbb{R}^d$ defined in Sec. (2). Under the assumption that g is convex with ∇g being L_g Lipschitz continuous, and h is an l.s.c. proper convex function, it is observed that the iterates generated by (2) satisfy [10]

$$f(u_t) - \min f = \mathcal{O}\Big(\frac{1}{t}\Big). \tag{3}$$

Beck and Teboulle [10] applied Nesterov's acceleration [28] on the iterative shrinkage-thresholding algorithm (ISTA) [14, 19] to propose an accelerated first-order method called the fast iterative shrinkage-thresholding algorithm (C-FISTA). The iterative sequence provided by the C-FISTA algorithm is

$$\begin{aligned} \omega_t &= u_t + \Big(\frac{t-3}{t}\Big)(u_t - u_{t-1}) \\ u_{t+1} &= \mathrm{Prox}_{\lambda h}(u_t - \lambda \nabla g(u_t)), \end{aligned} \tag{4}$$

and whenever $\lambda \in (0, \frac{1}{L_g})$ the rate of objective value gap is $\mathcal{O}\Big(\frac{1}{t^2}\Big)$ with the same presumptions as the PGM approach.

Su et al. [37] proposed an ODE for the discreatization

$$(\mathrm{AVD})_\alpha : \qquad \ddot{X} + \frac{\alpha}{t}\dot{X} + \nabla f(X(t)) = 0, \tag{5}$$

where $\alpha > 0$. The ODE (5) state a relation between first-order accelerated algorithm and second-order differential equation, where $\frac{\alpha}{t}$ is represented as the viscous damping coefficient. Attouch and Peypouquet [6] established an accelerated forward-backward (AFB) algorithm by discretizing the ODE $(\mathrm{AVD})_\alpha$. In [6], for $\alpha > 3$, they obtained $o\Big(\frac{1}{t^2}\Big)$ rate of decrement of objective residual and the iterates converge to some minimum point. Dynamical-system techniques are widely recognized as useful methods for addressing optimization problems; see, for instance, [2, 12, 23, 27, 34]. Specifically, accelerated forward-backward (AFB) algorithms are produced by suitable temporal discretizations of inertial dynamic systems with diminishing damping of the form $\frac{\alpha}{t}$: [1, 5, 8, 15, 22, 25, 41, 42]. Attouch et al. [7] studied the ODE which is the perturbed version of (5) that is:

$$\ddot{X} + \frac{\alpha}{t}\dot{X} + \nabla f(X(t)) = f_0(t), \tag{6}$$

where $f_0 : [t_0, +\infty] \to \mathbb{R}$ is an integrable source term. They showed the same rate of convergence as [6] for both smooth and non-smooth convex optimization problems under some condition on f_0.

There was a hitherto unknown result: in Nesterov's acceleration, the squared gradient norm has an inverse cubic rate in smooth convex optimization function. First time, Shi et al. [34, 35] answered the question. They derived a high-resolution ODE through dimensional analysis of the Nesterov gradient algorithm for smooth convex functions. The dynamical system is given by

$$\ddot{X} + \frac{3}{t}\dot{X} + \sqrt{s}\nabla^2 f(X(t))\dot{X} + \Big(1 + \frac{3\sqrt{s}}{2t}\Big)\nabla f(X(t)) = 0. \tag{7}$$

They extended the dynamical system (7) to a general form of high-resolution ODE, which is given by

$$\text{HR-ODE :} \qquad \ddot{X} + \frac{\alpha}{t}\dot{X} + \beta\sqrt{s}\nabla^2 f(X(t))\dot{X} + \Big(1 + \frac{\alpha\sqrt{s}}{2t}\Big)\nabla f(X(t)) = 0. \tag{8}$$

By employing symplectic Euler discretization on (HR-ODE), they established that for smooth convex function f and $\alpha \geq 3$, the convergence rates meet [34, 35]

$$f(u_t) - \min f = \mathcal{O}\left(\frac{1}{t^2}\right), \qquad \min_{1\leq j\leq t} \|\nabla f(u_j)\|^2 = \mathcal{O}\left(\frac{1}{t^3}\right). \tag{9}$$

Additionally, they obtained the better rate $o\left(\frac{1}{t^2}\right)$ for the objective residual, and the iterates converge to a minimal point when $\alpha > 3$ and $\beta > \frac{1}{2}$ for smooth convex class of objective functions.

Nevertheless, oscillations will occur in the iterates produced by the dynamical system $(\text{AVD})_\alpha$ for the Rosenbrock function, as stated in [3]. Attouch et al. [7, 3] studied the class of inertial dynamic system

$$(\text{DIN-AVD})_{(\alpha,\beta,b)} : \qquad \ddot{X} + \frac{\alpha}{t}\dot{X} + \beta(t)\nabla^2 g(X(t))\dot{X} + b(t)\nabla g(X(t)) = 0. \tag{10}$$

They assume b as a time scale parameter and β as a damping parameter; both depend on t. Attouch et al. [3] compared the $(\text{DIN-AVD})_{(\alpha,\beta,1)}$ with $(\text{AVD})_\alpha$, and see that Hessian-driven damping (HDD) term $\nabla^2 f(X(t))\dot{X}(t)$ neutralized the wild oscillations of $(\text{AVD})_\alpha$. They investigated the system $(\text{DIN-AVD})_{(\alpha,\beta,1+\frac{\beta}{t})}$ and established the rate $\mathcal{O}\big(\frac{1}{t^2}\big)$ for objective residual and $\mathcal{O}\big(\frac{1}{t^3}\big)$ for squared gradient norm when $\alpha \geq 3$. Furthermore, Attouch et al. [4] used system $(\text{DIN-AVD})_{(\alpha,\beta,1+\frac{\beta}{t})}$ and showed that the iterates generated by the proposed accelerated methods converge to some minimal point and proved that the improved rate of objective gap is $o\big(\frac{1}{t^2}\big)$ in the discrete case and also the continuous dynamics when $\alpha > 3$ for smooth convex function .

Wang et al. [40] discretized the $(\text{DIN-AVD})_{(\alpha,\beta(t),1)}$ system and induced the parameter θ to handle the ratio of implicit-explicit temporal discretization of velocity term $\dot{X}(t)$, and they obtained an $o\left(\frac{1}{t^2}\right)$ rate of convergence of $f(u_t) - \min f$, and iterates converge to a minimal point for non-smooth composite convex functions. He and Fang [21] introduced the accelerated forward-backward algorithm with subgradient correction (AFBA) for non-smooth convex functions based on the temporal discretization of (DIN-AVD)$_{(\alpha,\beta,1+\frac{\beta}{t})}$. They obtained $f(u_t) - \min f = \mathcal{O}\big(\frac{1}{t^2}\big)$ and $\min_{1\leq i\leq t} \operatorname{dist}^2(0, \partial f(u_i)) = \mathcal{O}\big(\frac{1}{t^3}\big)$ for $\alpha \geq 3$. The convergence rate of the function values improves to $o\left(\frac{1}{t^2}\right)$ when $\alpha > 3$.

We see the (HR-ODE) (contain $\mathcal{O}(\sqrt{s})$ term) is obtained by neglecting the $\mathcal{O}((\sqrt{s})^2)$ term [34], whereas (DIN-AVD)$_{(\alpha,\beta,1+\frac{\beta}{t})}$ (contain $\mathcal{O}(1)$ term) is obtained by neglecting the $\mathcal{O}(\sqrt{s})$ term [3]. It implies that (HR-ODE) is more stable than (DIN-AVD)$_{(\alpha,\beta,1+\frac{\beta}{t})}$ [34, 39]. We discretize the (HR-ODE) using an implicit-explicit approach, incorporating subgradient-based corrections. Additionally, we analyze the convergence of the proposed inertial first-order algorithms, considering both exact and inexact versions.

Main Contribution

(i) Shi et al. [34] discretize the ODE (16) for to solve smooth convex optimization problem. We develop an inertial forward-backward algorithm with subgradient correction (IFBASC) designed to address the non-smooth convex composite optimization Problem (1). This is achieved by discretizing the ODE (16) using an implicit-explicit approach.

(ii) In their discretization [34, 35] of ODE (16) and FISTA [7], the required step size is $\lambda \leq \frac{1}{L_{\nabla g}}$; we improve the step size to $\lambda \leq \frac{2\beta+1}{(\beta+1)L_{\nabla g}}$ for $\beta > 0$.

(iii) We demonstrate that for $\alpha \geq 3$, the convergence rates obey $f(u_t) - \min_u f(u) = \mathcal{O}\left(\frac{1}{t^2}\right)$. This improves to $o\left(\frac{1}{t^2}\right)$ when $\alpha > 3$ and $\beta > 0$, where our parameter β is more flexible, taking values in $(0, +\infty)$, compared to the results in [34] where $\beta > \frac{1}{2}$. Additionally, we have established that $\min_{1\leq i\leq t} \operatorname{dist}^2(0, \partial f(u_i)) = o\left(\frac{1}{t^3}\right)$ for $\alpha \geq 3$ and $\beta > 0$.

(iv) We also propose an inexact version of Algorithm (1) that incorporates error sequences $\{\vartheta_t, \varepsilon_t\} \in \mathbb{R}^d \times \mathbb{R}_+$. We achieve an $o\left(\frac{1}{t^2}\right)$ rate of convergence for the objective gap and an $o\left(\frac{1}{t^3}\right)$ convergence rate for $\min_{1\leq i\leq t} \operatorname{dist}^2(0, \partial_{\varepsilon_{t+1}} f(u_i))$, applicable for $\alpha \geq 3$ and $\beta > 0$.

(v) To demonstrate the effectiveness and practical usefulness of the proposed strategy, we investigate three composite convex optimisation problems: the lasso problem, image deblurring via wavelet transform and least squares with regularisation. It is shown that the proposed methods, Algorithms (1) and (2), yield better results than existing algorithms. Furthermore, we obtain better effectiveness of Algorithm (1) as compared with AFBA [21].

2 Preliminaries

We present certain terminology and key concepts to make the analysis that follows easier. For a comprehensive on convex optimization, we suggest the reader to [9, 32]. The standard Euclidean norm and inner product on $\mathbb{R}^d$ are denoted by $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$, respectively, while $\|\cdot\|_1$ represents the ℓ_1-norm. Let M be a nonempty closed subset of $\mathbb{R}^d$. We specify $\operatorname{dist}^2(0, M) := \min_{u\in M}\|u\|^2$. For a mapping $h : \mathbb{R}^d \to \mathbb{R}\cup\{+\infty\}$, its domain is defined as $\operatorname{dom}(h) := \{u \in \mathbb{R}^d \mid h(u) < +\infty\}$. The notation $[x]_+ := \max\{x, 0\}$. The function h is said to be proper if $\operatorname{dom}(h) \neq \varnothing$. Moreover, h is l.s.c. at u if $h(u) \leq \liminf_{y\to u} h(y)$. The proximal operator associated with a convex function h is defined, for $\lambda > 0$, as

$$\operatorname{Prox}_{\lambda h}(w) := \arg\min_{u\in\mathbb{R}^d}\left\{h(u) + \frac{1}{2\lambda}\|w-u\|^2\right\}, \quad \forall\, w \in \mathbb{R}^d, \tag{11}$$

see [9].

Lemma 1 ((Theorem 2.1.5, [30]), (Lemma 3, [24])). *Let $g : \mathbb{R}^d \to \mathbb{R}$ be a convex function whose gradient ∇g is Lipschitz continuous with Lipschitz constant $L_{\nabla g}$, then we have*

(i). $\langle\nabla g(a), b-c\rangle \geq g(b) - g(c) - \frac{L_{\nabla g}}{2}\|b-a\|^2$.

(ii). $\langle\nabla g(b) - \nabla g(c), b-c\rangle \geq \frac{L_{\nabla g}}{2}\|\nabla g(b) - \nabla g(c)\|^2$, for all a, b and c belongs to $\mathbb{R}^d$.

Lemma 2 (Theorem 2.1.5, [29]). *Let $g : \mathbb{R}^d \to \mathbb{R}$ be a convex function whose gradient ∇g is Lipschitz continuous with Lipschitz constant $L_{\nabla g}$, then we have*

$$\frac{1}{L_{\nabla g}}\|\nabla g(a) - \nabla g(b)\|^2 \leq \langle\nabla g(a) - \nabla g(b), a-b\rangle, \text{ for all } a, b \in \mathbb{R}^d. \tag{12}$$

Lemma 3. *[21] For any $a, b \in \mathbb{R}^d$ and $s > 0$, the following holds*

$$\langle u, a-b\rangle - \frac{1}{2}\|a-b\|^2 = \frac{1}{2}\|a\|^2 - \frac{1}{2}\|v\|^2, \tag{13}$$

$$-\frac{s}{2}\|a\|^2 - \frac{1}{2s}\|b\|^2 \leq \langle a, b\rangle \leq \frac{s}{2}\|a\|^2 + \frac{1}{2s}\|b\|^2. \tag{14}$$

Lemma 4 (Lemma 5.31, [9]). *Let $\{a_t\}_{t\geq1} \subset \mathbb{R}$ be a sequence that is bounded below, and let $\{b_t\}_{t\geq1}$ and $\{c_t\}_{t\geq1}$ be sequences of nonnegative real numbers satisfying $\sum_{t=1}^{\infty} c_t < \infty$. Suppose that, for every $t \geq 1$, $a_{t+1} \leq a_t - b_t + c_t$. Then $\{a_t\}_{t\geq1}$ is convergent, and in addition, $\sum_{t=1}^{\infty} b_t < \infty$.*

Lemma 5 (Lemma 2.47, [9]). *Let $\{a_t\}_{t\geq1} \subset \mathbb{R}^d$ be a sequence, and let $S \subset \mathbb{R}^d$ be a nonempty set. Suppose that:*

(i) For every $a^* \in S$, the limit $\lim_{t\to\infty}\|a_t - a^*\|$ exists.

(ii) All cluster points of the sequence $\{a_t\}$ lie in S.

Then $\{a_t\}_{t\geq 1}$ converges to a point in S.

Lemma 6 (Lemma 5.14, [5])**.** *Let $\{(a_t, b_t)\}_{t\geq 1}$ be non-negative sequence such that $\sum_{t=1}^{+\infty} b_t < +\infty$ and*

$$a_t^2 \leq c^2 + \sum_{j=1}^{t} b_j a_j,$$

for all $t \geq 1$, where $c \geq 0$. Then $a_t \leq c + \sum_{j=1}^{+\infty} b_j, \qquad \forall t \geq 1.$

Assumption 1. *We assume the following hypothesis hold:*

- *(a) $g : \mathbb{R}^d \to \mathbb{R}$ is a $\mathcal{C}^1$ function and gradient ∇g is Lipschitz continuous with constant $L_{\nabla g}$.*
- *(b) $h : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is a proper l.s.c. convex function.*
- *(c) $\arg\min_{u\in\mathbb{R}^d} f(u) = S \neq \varnothing$.*

3 Discretization of ODE

In [34], they proposed a high-resolution ordinary differential equation (HR-ODE)

$$\ddot{X} + \frac{\alpha}{t}\dot{X} + \beta\sqrt{s}\nabla^2 f(X(t))\dot{X} + \Big(1 + \frac{\alpha\sqrt{s}}{2t}\Big)\nabla f(X(t)) = 0. \tag{15}$$

To analyze the discrete version of (15), they slightly modified in the ODE (15), see [34, 35], they replace the third term $\Big(1 + \frac{\alpha\sqrt{s}}{2t}\Big)$ with $\Big(1 + \frac{\alpha\sqrt{s}}{t}\Big)$. Then, Equation (15) becomes

$$\ddot{X} + \frac{\alpha}{t}\dot{X} + \beta\sqrt{s}\nabla^2 f(X(t))\dot{X} + \Big(1 + \frac{\alpha\sqrt{s}}{t}\Big)\nabla f(X(t)) = 0. \tag{16}$$

Now, we discretize the ODE (16) in a implicit-explicit way by taking the step size $\sqrt{s}$ and $t = (t-1)\sqrt{s}$,

$$\begin{aligned}
&\frac{u_{t+1} - 2u_t + u_{t-1}}{s} + \frac{\alpha}{(t-1)s}(u_t - u_{t-1}) \\
&+ \frac{\beta\sqrt{s}}{\sqrt{s}}(\nabla h(u_{t+1}) + \nabla g(u_{t+1}) - (\nabla h(u_t) + \nabla g(u_t))) \\
&+ (\nabla h(u_{t+1}) + \nabla g(u_{t+1})) + \frac{\alpha\sqrt{s}}{(t-1)\sqrt{s}}(\nabla h(u_t) + \nabla g(u_t)) = 0.
\end{aligned} \tag{17}$$

Multiplying the Equation (17) with s, we obtain

$$\begin{aligned}&u_{t+1} - 2u_t + u_{t-1} + \frac{\alpha}{(t-1)}(u_t - u_{t-1})\\&+ \beta s(\nabla h(u_{t+1}) + \nabla g(u_{t+1}) - (\nabla h(u_t) + \nabla g(u_t)))\\&+ s(\nabla h(u_{t+1}) + \nabla g(u_{t+1})) + \frac{\alpha s}{t-1}(\nabla h(u_t) + \nabla g(u_t)) = 0.\end{aligned} \tag{18}$$

Then,

$$\begin{aligned}u_{t+1} = u_t &+ \Big(1 - \frac{\alpha}{t-1}\Big)(u_t - u_{t-1}) - (s + \beta s)(\nabla h(u_{t+1}) + \nabla g(u_{t+1}))\\&+ s\Big(\beta - \frac{\alpha}{t-1}\Big)(\nabla h(u_t) + \nabla g(u_t)).\end{aligned} \tag{19}$$

Since to evaluate u_{t+1}, we need to require u_{t+1}, that is not possible. We replace $\nabla g(u_{t+1})$ with $\nabla g(\omega_t)$, where

$$\omega_t = u_t + \Big(1 - \frac{\alpha}{t-1}\Big)(u_t - u_{t-1}) + s\Big(\beta - \frac{\alpha}{t-1}\Big)(\nabla h(u_t) + \nabla g(u_t)). \tag{20}$$

Equation (19) becomes

$$u_{t+1} = \omega_t - (s + \beta s)(\nabla h(u_{t+1}) + \nabla g(\omega_t)).$$

Taking $\lambda = s + \beta s$ and solving the above equation, we obtain

$$u_{t+1} + \lambda \nabla h(u_{t+1}) = \omega_t - \lambda \nabla g(\omega_t). \tag{21}$$

Since g is only convex function, we get

$$u_{t+1} = \mathrm{Prox}_{\lambda h}(\omega_t - \lambda \nabla g(\omega_t)). \tag{22}$$

Using Euqation (22), we obtain

$$-\nabla g(\omega_t) - \frac{1}{\lambda}(u_{t+1} - \omega_t) \in \partial h(u_{t+1}).$$

In order to manage the non-smooth term $\nabla h(u_t)$ in (20), we substitute $\nabla h(u_t)$ by σ_t and $\sigma_t \in \partial h(u_t)$. We can decide

$$\sigma_{t+1} = -\nabla g(\omega_t) - \frac{1}{\lambda}(u_{t+1} - \omega_t).$$

Algorithm 1 Inertial forward-backward algorithm with subgradient correction (IFBASC)

```
Initialize: Let $u_1 = u_2 = \omega_1 \in \mathbb{R}^d$, $\sigma_2 \in \partial h(u_1)$, $\alpha \geq 3$, $s > 0$, $\beta \geq 0$, $\lambda = s + \beta s$.
for $t = 2, \dots$ do
    $\omega_t = u_t + \frac{t-1-\alpha}{t-1}(u_t - u_{t-1}) + s\Big(\beta - \frac{\alpha}{t-1}\Big)\big(\sigma_t + \nabla g(\omega_{t-1})\big)$,
    $u_{t+1} = \mathrm{Prox}_{\lambda h}\big(\omega_t - \lambda \nabla g(\omega_t)\big)$,
    $\sigma_{t+1} = -\nabla g(\omega_t) - \frac{1}{\lambda}(u_{t+1} - \omega_t)$.
    until a termination criterion is satisfied
end for
return $(u_t, \omega_t, \sigma_t)$
```

4 Lyapunov function

Let us define the energy function

$$\mathcal{E}_t(\gamma) := s\big[(t+\alpha-\beta-1)(t-\gamma-1)+\eta(t\beta-\alpha)\big](f(u_t)-f(u^*))+\frac{1}{2}\|\pi_t(\gamma)\|^2+\frac{\gamma\eta}{2}\|u_t-u^*\|^2, \tag{23}$$

where

$$\pi_t(\gamma) := \gamma(u_t - u^*) + (t-1-\alpha)(u_t - u_{t-1}) + s(\beta(t-1)-\alpha)(\sigma_t + \nabla g(\omega_{t-1})). \tag{24}$$

Lemma 7. *Let Assumption 1 hold, and the sequences* $\{(u_t, \omega_t, \sigma_t)\}_{t\geq 1}$ *be generated by IFBASC, and let* $u^* \in S$. *The corresponding Lyapunov function* $\mathcal{E}_t(\gamma)$ *is given in* (23). *Then, we have*

$$\begin{aligned}
&\mathcal{E}_{t+1}(\gamma) - \mathcal{E}_t(\gamma)\\
&\leq s\big[\eta\beta + t(-\gamma+2) + (\alpha-\beta)(-\gamma+1) - 1\big](f(u_{t+1}) - f(u^*))\\
&- \Big[\eta(t-\alpha) + \frac{\gamma\eta}{2} + \frac{\eta^2}{2}\Big]\|u_{t+1}-u_t\|^2\\
&- \frac{s^2(t+\alpha-\beta-1)}{2}\Big[t(2\beta+1) - \alpha - \beta - 1 - L_{\nabla g}\lambda(\beta+1)\Big(t - 1 + \frac{\eta(t\beta-\alpha)}{t+\alpha-\beta-1}\Big)\Big]\\
&\times \|\sigma_{t+1} + \nabla g(\omega_t)\|^2.
\end{aligned}$$

Proof. Using Equation (13),

$$\begin{aligned}
&\frac{1}{2}\|\pi_{t+1}(\gamma)\|^2 - \frac{1}{2}\|\pi_t(\gamma)\|^2 + \frac{\gamma\eta}{2}\|u_{t+1}-u^*\|^2 - \frac{\gamma\eta}{2}\|u_t - u^*\|^2\\
&= \langle \pi_{t+1}(\gamma), \pi_{t+1}(\gamma) - \pi_t(\gamma)\rangle - \frac{1}{2}\|\pi_{t+1}(\gamma) - \pi_t(\gamma)\|^2\\
&+ \gamma\eta\langle u_{t+1} - u^*, u_{t+1} - u_t\rangle - \frac{\gamma\eta}{2}\|u_{t+1}-u^*\|^2.
\end{aligned} \tag{25}$$

From Equation (24), we have

$$\begin{aligned}
&\pi_{t+1}(\gamma) - \pi_t(\gamma)\\
&= \gamma(u_{t+1} - u^*) + (t-\alpha)(u_{t+1} - u_t) + s(\beta t - \alpha)(\sigma_{t+1} + \nabla g(\omega_t))\\
&- \gamma(u_t - u^*) - (t-1-\alpha)(u_t - u_{t-1}) - s(\beta(t-1) - \alpha)(\sigma_t + \nabla g(\sigma_{t-1}))\\
&= \gamma(u_{t+1} - u_t) + (t-\alpha)(u_{t+1} - u_t) + s(\beta t - \alpha)(\sigma_{t+1} + \nabla g(\omega_t))\\
&- (t-1-\alpha)(u_t - u_{t-1}) - s(\beta(t-1) - \alpha)(\sigma_t + \nabla g(\sigma_{t-1})).
\end{aligned} \tag{26}$$

Since $\eta = \alpha - 1 - \gamma$, Equation (26) becomes

$$\begin{aligned}
\pi_{t+1}(\gamma) - \pi_t(\gamma) &= -\eta(u_{t+1} - u_t) + (t-1)(u_{t+1} - u_t) + s(\beta t - \alpha)(\sigma_{t+1} + \nabla g(\omega_t))\\
&\quad - (t-1-\alpha)(u_t - u_{t-1}) - s(\beta(t-1) - \alpha)(\sigma_t + \nabla g(\omega_t)).
\end{aligned}$$

Using step 3 of Algorithm (1),

$$\begin{aligned}
&\pi_{t+1}(\gamma) - \pi_t(\gamma)\\
&= -\eta(u_{t+1} - u_t) + (t-1)(u_{t+1} - u_t) + s(\beta t - \alpha)(\sigma_{t+1} + \nabla g(\omega_t)) - (t-1)(\omega_t - u_t)\\
&= -\eta(u_{t+1} - u_t) + (t-1)(u_{t+1} - \omega_t) + s(\beta t - \alpha)(\sigma_{t+1} + \nabla g(\omega_t)).
\end{aligned} \tag{27}$$

Utilizing $\lambda = s(\beta + 1)$ and step 5 of Algorithm (1), Equation (27) becomes

$$\begin{aligned}
&\pi_{t+1}(\gamma) - \pi_t(\gamma)\\
&= -\eta(u_{t+1} - u_t) - s(\beta+1)(t-1)(\sigma_{t+1} + \nabla g(\omega_t)) + s(\beta t - \alpha)(\sigma_{t+1} + \nabla g(\omega_t))\\
&= -\eta(u_{t+1} - u_t) - s(t + \alpha - \beta - 1)(\sigma_{t+1} + \nabla g(\omega_t)).
\end{aligned} \tag{28}$$

Now,

$$\begin{aligned}
\|\pi_{t+1}(\gamma) - \pi_t(\gamma)\|^2 &= \|\eta(u_{t+1} - u_t) + s(t + \alpha - \beta - 1)(\sigma_{t+1} + \nabla g(\omega_t))\|^2\\
&= \eta^2\|u_{t+1} - u_t\|^2 + s^2(t + \alpha - \beta - 1)^2\|\sigma_{t+1} + \nabla g(\omega_t)\|^2\\
&+ 2\eta s(t + \alpha - \beta - 1)\langle u_{t+1} - u_t, \sigma_{t+1} + \nabla g(\omega_t)\rangle.
\end{aligned} \tag{29}$$

From Equation (24) and (28), we evaluate

$$\begin{aligned}
&\langle \pi_{t+1}(\gamma), \pi_{t+1}(\gamma) - \pi_t(\gamma)\rangle\\
&= \langle \gamma(u_{t+1} - u^*) + (t-\alpha)(u_{t+1} - u_t) + s(\beta t - \alpha)(\sigma_{t+1} + \nabla g(\omega_t)),\\
&- \eta(u_{t+1} - u_t) - s(t + \alpha - \beta - 1)(\sigma_{t+1} + \nabla g(\omega_t))\\
&= -\gamma\eta\langle u_{t+1} - u^*, u_{t+1} - u_t\rangle - \gamma s(t + \alpha - \beta - 1)\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle\\
&- \eta(t-\alpha)\|u_{t+1} - u_t\|^2 - s(t-\alpha)(t + \alpha - \beta - 1)\langle u_{t+1} - u_t, \sigma_{t+1} + \nabla g(\omega_t)\rangle\\
&- \eta s(t\beta - \alpha)\langle u_{t+1} - u_t, \sigma_{t+1} + \nabla g(\omega_t)\rangle\\
&- s^2(t\beta - \alpha)(t + \alpha - \beta - 1)\|\sigma_{t+1} + \nabla g(\omega_t)\|^2\\
&= -\gamma\eta\langle u_{t+1} - u^*, u_{t+1} - u_t\rangle - \gamma s(t + \alpha - \beta - 1)\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle\\
&- s\big[(t-\alpha)(t + \alpha - \beta - 1) + \eta(t\beta - \alpha)\big]\langle u_{t+1} - u_t, \sigma_{t+1} + \nabla g(\omega_t)\rangle\\
&- \eta(t-\alpha)\|u_{t+1} - u_t\|^2 - s^2(t\beta - \alpha)(t + \alpha - \beta - 1)\|\sigma_{t+1} + \nabla g(\omega_t)\|^2.
\end{aligned} \tag{30}$$

From Equation (29) and (30), Equation (25) becomes

$$
\begin{aligned}
&\frac{1}{2}\|\pi_{t+1}(\gamma)\|^2 - \frac{1}{2}\|\pi_t(\gamma)\|^2 + \frac{\gamma\eta}{2}\|u_{t+1} - u^*\|^2 - \frac{\gamma\eta}{2}\|u_t - u^*\|^2 \\
&= \langle \pi_{t+1}(\gamma), \pi_{t+1}(\gamma) - \pi_t(\gamma)\rangle - \frac{1}{2}\|\pi_{t+1}(\gamma) - \pi_t(\gamma)\|^2 \\
&+ \gamma\eta\langle u_{t+1} - u^*, u_{t+1} - u_t\rangle - \frac{\gamma\eta}{2}\|u_{t+1} - u^*\|^2 \\
&= -\gamma\eta\langle u_{t+1} - u^*, u_{t+1} - u_t\rangle - \gamma s(t+\alpha-\beta-1)\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle \\
&- s\big[(t-\alpha)(t+\alpha-\beta-1) + \eta(t\beta-\alpha)\big]\langle u_{t+1} - u_t, \sigma_{t+1} + \nabla g(\omega_t)\rangle \\
&- \eta(t-\alpha)\|u_{t+1} - u_t\|^2 - s^2(t\beta-\alpha)(t+\alpha-\beta-1)\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 \\
&- \frac{\eta^2}{2}\|u_{t+1} - u_t\|^2 - \frac{s^2(t+\alpha-\beta-1)^2}{2}\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 \\
&- \eta s(t+\alpha-\beta-1)\langle u_{t+1} - u_t, \sigma_{t+1} + \nabla g(\omega_t)\rangle \\
&+ \gamma\eta\langle u_{t+1} - u^*, u_{t+1} - u_t\rangle - \frac{\gamma\eta}{2}\|u_{t+1} - u^*\|^2 \\
&= -\gamma s(t+\alpha-\beta-1)\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle \\
&- s\big[(t+\alpha-\beta-1)(t-\alpha+\eta) + \eta(t\beta-\alpha)\big]\langle u_{t+1} - u_t, \sigma_{t+1} + \nabla g(\omega_t)\rangle \\
&- \big[\eta(t-\alpha) + \frac{\gamma\eta}{2} + \frac{\eta^2}{2}\big]\|u_{t+1} - u_t\|^2 \\
&- \frac{s^2(t+\alpha-\beta-1)}{2}(t(2\beta+1) - \alpha - \beta - 1)\|\sigma_{t+1} + \nabla g(\omega_t)\|^2. \qquad (31)
\end{aligned}
$$

Using step 4 and step 5 of Algorithm (1), we obtain

$$\sigma_{t+1} \in \partial h(u_{t+1}).$$

Consequently, considering the convexity of h for any $u \in \mathbb{R}^d$, we obtain

$$\langle \sigma_{t+1}, u_{t+1} - u\rangle \geq h(u_{t+1}) - h(u). \qquad (32)$$

Lemma (1) (i), step 5 of Algorithm (1) implies that

$$
\begin{aligned}
\langle \nabla g(\omega_t), u_{t+1} - u\rangle &\geq g(u_{t+1}) - g(u) - \frac{L_{\nabla g}}{2}\|u_{t+1} - \omega_t\|^2 \\
&= g(u_{t+1}) - g(u) - \frac{L_{\nabla g}\lambda^2}{2}\|\sigma_{t+1} + \nabla g(\omega_t)\|^2. \qquad (33)
\end{aligned}
$$

Adding Equation (32) and (33), we obtain

$$
\begin{aligned}
\langle u_{t+1} - u, \sigma_{t+1} + \nabla g(\omega_t)\rangle &\geq g(u_{t+1}) + h(u_{t+1}) - g(u) - h(u) - \frac{L_{\nabla g}\lambda^2}{2}\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 \\
&= f(u_{t+1}) - f(u) - \frac{L_{\nabla g}\lambda^2}{2}\|\sigma_{t+1} + \nabla g(\omega_t)\|^2. \qquad (34)
\end{aligned}
$$

Replacing u in Equation (34) with u^* and u_t, respectively, we get

$$\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle \geq f(u_{t+1}) - f(u^*) - \frac{L_{\nabla g}\lambda^2}{2}\|\sigma_{t+1} + \nabla g(\omega_t)\|^2, \tag{35}$$

and

$$\langle u_{t+1} - u_t, \sigma_{t+1} + \nabla g(\omega_t)\rangle \geq f(u_{t+1}) - f(u_t) - \frac{L_{\nabla g}\lambda^2}{2}\|\sigma_{t+1} + \nabla g(\omega_t)\|^2. \tag{36}$$

Now, we obtain

$$\begin{aligned}
&\gamma s(t+\alpha-\beta-1)\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle \\
&+ s\big[(t+\alpha-\beta-1)(t-\alpha+\eta) + \eta(t\beta-\alpha)\big]\langle u_{t+1} - u_t, \sigma_{t+1} + \nabla g(\omega_t)\rangle \\
&\overset{(35),(36)}{\geq} s(t+\alpha-\beta-1)(f(u_{t+1}) - f(u^*)) \\
&+ s\big[(t+\alpha-\beta-1)(t-\alpha+\eta) + \eta(t\beta-\alpha)\big](f(u_{t+1}) - f(u_t)) \\
&- \frac{L_{\nabla g}\lambda^2}{2}s\big[\gamma(t+\alpha-\beta-1) + (t+\alpha-\beta-1)(t-\alpha+\eta) + \eta(t\beta-\alpha)\big]\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 \\
&= s\big[(t+\alpha-\beta-1)(t-\alpha+\eta+\gamma) + \eta(t\beta-\alpha)\big](f(u_{t+1}) - f(u^*)) \\
&- s\big[(t+\alpha-\beta-1)(t-\alpha+\eta) + \eta(t\beta-\alpha)\big](f(u_t) - f(u^*)) \\
&- \frac{L_{\nabla g}\lambda^2}{2}s\big[(t+\alpha-\beta-1)(t-\alpha+\eta+\gamma) + \eta(t\beta-\alpha)\big]\|\sigma_{t+1} + \nabla g(\omega_t)\|^2.
\end{aligned} \tag{37}$$

Since $\eta = \alpha - 1 - \gamma$, Equation (37) becomes

$$\begin{aligned}
&\gamma s(t+\alpha-\beta-1)\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle \\
&+ s\big[(t+\alpha-\beta-1)(t-\alpha+\eta) + \eta(t\beta-\alpha)\big]\langle u_{t+1} - u_t, \sigma_{t+1} + \nabla g(\omega_t)\rangle \\
&\geq \underbrace{s\big[(t+\alpha-\beta-1)(t-1) + \eta(t\beta-\alpha)\big](f(u_{t+1}) - f(u^*))}_{J_1} \\
&- \underbrace{s\big[(t+\alpha-\beta-1)(t-\gamma-1) + \eta(t\beta-\alpha)\big](f(u_t) - f(u^*))}_{J_2} \\
&\underbrace{- \frac{L_{\nabla g}\lambda^2}{2}s\big[(t+\alpha-\beta-1)(t-1) + \eta(t\beta-\alpha)\big]\|\sigma_{t+1} + \nabla g(\omega_t)\|^2}_{J_3}.
\end{aligned} \tag{38}$$

From Equation (23), we get

$$\begin{aligned}
&\mathcal{E}_{t+1}(\gamma) - \mathcal{E}_t(\gamma) \\
&= s\big[(t+\alpha-\beta)(t-\gamma) + \eta((t+1)\beta-\alpha)\big](f(u_{t+1}) - f(u^*)) \\
&- s\big[(t+\alpha-\beta-1)(t-\gamma-1) + \eta(t\beta-\alpha)\big](f(u_t) - f(u^*)) \\
&+ \frac{1}{2}\|\pi_{t+1}(\gamma)\|^2 + \frac{\gamma\eta}{2}\|u_{t+1} - u^*\|^2 - \frac{1}{2}\|\pi_t(\gamma)\|^2 - \frac{\gamma\eta}{2}\|u_t - u^*\|^2.
\end{aligned} \tag{39}$$

From Equation (39), we obtain

$$\begin{aligned}
&\mathcal{E}_{t+1}(\gamma) - \mathcal{E}_t(\gamma) \\
&\overset{(31),(38)}{\leq} \; s\big[(t+\alpha-\beta)(t-\gamma) + \eta((t+1)\beta-\alpha)\big](f(u_{t+1}) - f(u^*)) \\
&- s\big[(t+\alpha-\beta-1)(t-\gamma-1) + \eta(t\beta-\alpha)\big](f(u_t) - f(u^*)) \\
&- s\big[(t+\alpha-\beta-1)(t-1) + \eta(t\beta-\alpha)\big](f(u_{t+1}) - f(u^*)) \\
&+ s\big[(t+\alpha-\beta-1)(t-\gamma-1) + \eta(t\beta-\alpha)\big](f(u_t) - f(u^*)) \\
&+ \frac{L_{\nabla g}\lambda^2}{2} s\big[(t+\alpha-\beta-1)(t-1) + \eta(t\beta-\alpha)\big]\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 \\
&- \big[\eta(t-\alpha) + \frac{\gamma\eta}{2} + \frac{\eta^2}{2}\big]\|u_{t+1} - u_t\|^2 \\
&- \frac{s^2(t+\alpha-\beta-1)}{2}(t(2\beta+1)-\alpha-\beta-1)\|\sigma_{t+1} + \nabla g(\omega_t)\|^2. \qquad (40)
\end{aligned}$$

Now,

$$\begin{aligned}
&s\big[(t+\alpha-\beta)(t-\gamma) + \eta((t+1)\beta-\alpha)\big] - s\big[(t+\alpha-\beta-1)(t-1) + \eta(t\beta-\alpha)\big] \\
&= s\big[\eta\beta + (t+\alpha-\beta)(t-\gamma-t+1) + (t-1)\big] \\
&= s\big[\eta\beta + (t+\alpha-\beta)(-\gamma+1)\big] \\
&= s\big[\eta\beta + t(-\gamma+2) + (\alpha-\beta)(-\gamma+1) - 1\big], \qquad (41)
\end{aligned}$$

and as $\lambda = s(\beta+1)$,

$$\begin{aligned}
&-\frac{s^2(t+\alpha-\beta-1)}{2}(t(2\beta+1)-\alpha-\beta-1) + \frac{L_{\nabla g}\lambda^2}{2} s\big[(t+\alpha-\beta-1)(t-1) + \eta(t\beta-\alpha)\big] \\
&= -\frac{s^2(t+\alpha-\beta-1)}{2}\Big[t(2\beta+1)-\alpha-\beta-1 - L_{\nabla g}\lambda(\beta+1)\Big(t-1+\frac{\eta(t\beta-\alpha)}{t+\alpha-\beta-1}\Big)\Big]. \qquad (42)
\end{aligned}$$

From Equation (41) and (42), Equation (40) becomes

$$\begin{aligned}
&\mathcal{E}_{t+1}(\gamma)-\mathcal{E}_t(\gamma)\\
&\le \underbrace{s\big[\eta\beta+t(-\gamma+2)+(\alpha-\beta)(-\gamma+1)-1\big](f(u_{t+1})-f(u^*))}_{F_1}\\
&\underbrace{-\big[\eta(t-\alpha)+\frac{\gamma\eta}{2}+\frac{\eta^2}{2}\big]\|u_{t+1}-u_t\|^2}_{X_1}\\
&-\underbrace{\frac{s^2(t+\alpha-\beta-1)}{2}\Big[t(2\beta+1)-\alpha-\beta-1-L_{\nabla g}\lambda(\beta+1)\Big(t-1+\frac{\eta(t\beta-\alpha)}{t+\alpha-\beta-1}\Big)\Big]}_{G_1}\\
&\underbrace{\times\|\sigma_{t+1}+\nabla g(\omega_t)\|^2}_{}. \qquad (43)
\end{aligned}$$

□

Theorem 8. *Suppose Assumption 1 holds. Let* $\{(u_t,\omega_t,\sigma_t)\}_{t\ge1}$ *be the sequences generated by Algorithm 1 and* $u^*\in S$. *Assuming* $2\beta+1-L_{\nabla g}\lambda(\beta+1)>0$, *we have*

(i). $f(u_t)-\min f=\mathcal{O}\Big(\frac{1}{t^2}\Big)$.

(ii). $(\alpha-3)\sum_{t=1}^{+\infty}t(f(u_{t+t_{\alpha,\beta}+1})-f(u^*))<+\infty$..

(iii). $\sum_{t=t_{\alpha,\beta}+1}^{+\infty}(t-t_{\alpha,\beta})^2\|\sigma_{t+1}+\nabla g(\omega_t)\|^2<+\infty,\qquad \sum_{t=t_{\alpha,\beta}+1}^{+\infty}(t-t_{\alpha,\beta})^2\|\sigma_{t+1}+\nabla g(u_{t+1})\|^2<+\infty$.

(iv). $\sum_{t=1}^{+\infty}t^2\mathrm{dist}^2(0,\partial f(u_{t+t_{\alpha,\beta}+1}))<+\infty,\qquad \min_{1\le i\le t}\mathrm{dist}^2(0,\partial f(u_{i+t_{\alpha,\beta}+1}))=o\Big(\frac{1}{t^3}\Big)$.

(v). If $\alpha>3$, $\sum_{t=1}^{+\infty}t\|u_{t+T_{\alpha,\beta}+1}-u_{t+T_{\alpha,\beta}}\|^2<+\infty$.

Proof. (i). Choose $\gamma=\alpha-1$ then $\eta=0$. From Equation (43), for all $t\ge1$ we obtain

$$\begin{aligned}
&\mathcal{E}_{t+1}(\alpha-1)-\mathcal{E}_t(\alpha-1)\\
&\le s\big[t(3-\alpha)+(\alpha-\beta)(2-\alpha)-1\big](f(u_{t+1})-f(u^*))\\
&-\frac{s^2(t+\alpha-\beta-1)}{2}\Big[t(2\beta+1)-\alpha-\beta-1-L_{\nabla g}\lambda(\beta+1)(t-1)\Big]\|\sigma_{t+1}+\nabla g(\omega_t)\|^2. \qquad (44)
\end{aligned}$$

To show

$$\mathcal{E}_{t+1}(\alpha-1)-\mathcal{E}_t(\alpha-1)\le 0,$$

we require $t(3-\alpha)+(\alpha-\beta)(2-\alpha)-1\le 0$, $t+\alpha-\beta-1\ge 0$ and $t(2\beta+1)-\alpha-\beta-1-L_{\nabla g}\lambda(\beta+1)(t-1)\ge 0$. Simplifying the said requirements, we obtain $t\ge \frac{(2-\alpha)(\alpha-\beta)-1}{\alpha-3}$, $t\ge \beta-\alpha+1$ and $t\ge \frac{(\alpha+\beta+1)-L_{\nabla g}\lambda(\beta+1)}{2\beta+1-L_{\nabla g}\lambda(\beta+1)}$. Therefore, using $\lambda=s(\beta+1)$ we choose

$$t_{\alpha,\beta}=\max\Big\{0,\Big\lfloor\frac{(2-\alpha)(\alpha-\beta)-1}{\alpha-3}\Big\rfloor+1,\lfloor\beta-\alpha+1\rfloor+1,\Big\lfloor\frac{(\alpha+\beta+1)-L_{\nabla g}s(\beta+1)^2}{2\beta+1-L_{\nabla g}\lambda(\beta+1)}\Big\rfloor+1\Big\}. \tag{45}$$

Then, $t_{\alpha,\beta}(\alpha-3)\ge(2-\alpha)(\alpha-\beta)-1$, $t_{\alpha,\beta}\ge\beta-\alpha+1$ and $t_{\alpha,\beta}(2\beta+1-L_{\nabla g}\lambda(\beta+1))\ge(\alpha+\beta+1)-L_{\nabla g}s(\beta+1)^2$. Now,

$$\begin{aligned}t(3-\alpha)+(\alpha-\beta)(2-\alpha)-1&=-t(\alpha-3)+((\alpha-\beta)(2-\alpha)-1)\\&\le -t(\alpha-3)+t_{\alpha,\beta}(\alpha-3)\\&=-(\alpha-3)(t-t_{\alpha,\beta}).\end{aligned}\tag{46}$$

Similarly, $t+\alpha-\beta-1=t-(\beta-\alpha+1)\ge t-t_{\alpha,\beta}$ and

$$\begin{aligned}&\frac{s^2(t+\alpha-\beta-1)}{2}\Big[t(2\beta+1)-\alpha-\beta-1-L_{\nabla g}\lambda(\beta+1)(t-1)\Big]\\&\ge\frac{s^2(t-t_{\alpha,\beta})}{2}\Big[t((2\beta+1)-L_{\nabla g}\lambda(\beta+1))-\big(\alpha+\beta+1-L_{\nabla g}s(\beta+1)^2\big)\Big]\\&\ge\frac{s^2(t-t_{\alpha,\beta})}{2}\Big[t((2\beta+1)-L_{\nabla g}\lambda(\beta+1))-t_{\alpha,\beta}(2\beta+1-L_{\nabla g}\lambda(\beta+1))\Big]\\&=\frac{s^2(t-t_{\alpha,\beta})^2}{2}[(2\beta+1)-L_{\nabla g}\lambda(\beta+1)].\end{aligned}\tag{47}$$

Using Equation (46) and (47), Equation (44) becomes

$$\begin{aligned}\mathcal{E}_{t+1}(\alpha-1)-\mathcal{E}_t(\alpha-1)&\le -s(\alpha-3)(t-t_{\alpha,\beta})(f(u_{t+1})-f(u^*))\\&\quad-\frac{s^2(t-t_{\alpha,\beta})^2}{2}[(2\beta+1)-L_{\nabla g}\lambda(\beta+1)]\|\sigma_{t+1}+\nabla g(\omega_t)\|^2.\end{aligned}\tag{48}$$

Since $\alpha\ge 3$ and $(2\beta+1)-L_{\nabla g}\lambda(\beta+1)>0$, then

$$\begin{aligned}\mathcal{E}_{t+1}(\alpha-1)-\mathcal{E}_t(\alpha-1)&\le -s(\alpha-3)(t-t_{\alpha,\beta})(f(u_{t+1})-f(u^*))\\&\quad-\frac{s^2(t-t_{\alpha,\beta})^2}{2}[(2\beta+1)-L_{\nabla g}\lambda(\beta+1)]\|\sigma_{t+1}+\nabla g(\omega_t)\|^2\\&\le 0,\end{aligned}\tag{49}$$

for all $t\ge t_{\alpha,\beta}+1$. Thus, the sequence $\mathcal{E}_t(\alpha-1)$ is positive and non-increasing, and this implies it is bounded. From Equation (23), we obtain

$$s\big(t+\alpha-\beta-1)(t-\alpha)\big](f(u_t)-f(u^*))+\frac{1}{2}\|\pi_t(\alpha-1)\|^2\le\mathcal{E}_{t_{\alpha,\beta}}(\alpha-1),$$

also we have

$$s(t+\alpha-\beta-1)(t-\alpha)(f(u_t)-f(u^*)) \leq \mathcal{E}_{t_{\alpha,\beta}}(\alpha-1).$$

Then,

$$f(u_t)-f(u^*) \leq \frac{\mathcal{E}_{t_{\alpha,\beta}}(\alpha-1)}{s(t+\alpha-\beta-1)(t-\alpha)}. \tag{50}$$

(ii). From Equation (49), we get

$$\mathcal{E}_{t+1}(\alpha-1)-\mathcal{E}_t(\alpha-1)+s(\alpha-3)(t-t_{\alpha,\beta})(f(u_{t+1})-f(u^*)) \leq 0. \tag{51}$$

By summing Equation (51) for $t = t_{\alpha,\beta}+1, t_{\alpha,\beta}+2, \cdots, n$, we obtain

$$\mathcal{E}_{n+1}(\alpha-1)+s(\alpha-3)\sum_{t=t_{\alpha,\beta}+1}^{n}(t-t_{\alpha,\beta})(f(u_{t+1})-f(u^*)) \leq \mathcal{E}_{t_{\alpha,\beta}+1}(\alpha-1).$$

Also,

$$(\alpha-3)\sum_{t=t_{\alpha,\beta}+1}^{n}(t-t_{\alpha,\beta})(f(u_{t+1})-f(u^*)) \leq \frac{\mathcal{E}_{t_{\alpha,\beta}+1}(\alpha-1)}{s}.$$

As $\{\mathcal{E}_t(\alpha-1)\}_{t\geq 1}$ is bounded, for $n \to +\infty$,

$$(\alpha-3)\sum_{t=t_{\alpha,\beta}+1}^{+\infty}(t-t_{\alpha,\beta})(f(u_{t+1})-f(u^*)) < +\infty. \tag{52}$$

Replace $t-t_{\alpha,\beta}$ with t, then

$$(\alpha-3)\sum_{t=1}^{+\infty}t(f(u_{t+t_{\alpha,\beta}+1})-f(u^*)) < +\infty. \tag{53}$$

(iii). From Equation (49), as $2\beta+1-L_{\nabla g}\lambda(\beta+1) > 0$ we get

$$\mathcal{E}_{t+1}(\alpha-1)-\mathcal{E}_t(\alpha-1)+\frac{s^2(t-t_{\alpha,\beta})^2}{2}\|\sigma_{t+1}+\nabla g(\omega_t)\|^2 < 0. \tag{54}$$

By summing Equation (54) for $t = t_{\alpha,\beta}+1, t_{\alpha,\beta}+2, \cdots, N$, we obtain

$$\mathcal{E}_N(\alpha-1)+\frac{s^2}{2}\sum_{t=t_{\alpha,\beta}+1}^{N}(t-t_{\alpha,\beta})^2\|\sigma_{t+1}+\nabla g(\omega_t)\|^2 \leq \mathcal{E}_{t_{\alpha,\beta}+1}(\alpha-1).$$

Also,

$$\sum_{t=t_{\alpha,\beta}+1}^{N}(t-t_{\alpha,\beta})^2\|\sigma_{t+1}+\nabla g(\omega_t)\|^2 \leq \frac{2\mathcal{E}_{t_{\alpha,\beta}+1}(\alpha-1)}{s^2}.$$

As $\{\mathcal{E}_t(\alpha-1)\}_{t\geq 1}$ is bounded, for $N \to +\infty$,

$$\sum_{t=t_{\alpha,\beta}+1}^{+\infty} (t-t_{\alpha,\beta})^2 \|\sigma_{t+1} + \nabla g(\omega_t)\|^2 < +\infty. \tag{55}$$

Replace $t - t_{\alpha,\beta}$ with t, then

$$\sum_{t=1}^{+\infty} t^2 \|\sigma_{t+t_{\alpha,\beta}+1} + \nabla g(\omega_{t+t_{\alpha,\beta}})\|^2 < +\infty. \tag{56}$$

Using the Lipschitz continuity of ∇g, we have

$$\begin{aligned}\|\sigma_{t+1} + \nabla g(u_{t+1})\|^2 &= \|\sigma_{t+1} + \nabla g(\omega_t) + \nabla g(u_{t+1}) - \nabla g(\omega_t)\|^2 \\ &\leq 2(\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 + \|\nabla g(u_{t+1}) - \nabla g(\omega_t)\|^2) \\ &\leq 2(\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 + L_{\nabla g}^2 \|u_{t+1} - \omega_t\|^2). \end{aligned} \tag{57}$$

From step 5 of Algorithm (1), the above equation becomes

$$\|\sigma_{t+1} + \nabla g(u_{t+1})\|^2 \leq 2(1 + L_{\nabla g}^2 \lambda^2) \|\sigma_{t+1} + \nabla g(\omega_t)\|^2. \tag{58}$$

Using the result (58), we obtain

$$\sum_{t=t_{\alpha,\beta}+1}^{+\infty} (t-t_{\alpha,\beta})^2 \|\sigma_{t+1} + \nabla g(u_{t+1})\|^2 \leq \sum_{t=t_{\alpha,\beta}+1}^{+\infty} (t-t_{\alpha,\beta})^2 \|\sigma_{t+1} + \nabla g(\omega_t)\|^2 < +\infty. \tag{59}$$

Replace $t - t_{\alpha,\beta}$ with t, we have

$$\sum_{t=1}^{+\infty} t^2 \|\sigma_{t+t_{\alpha,\beta}+1} + \nabla g(u_{t+t_{\alpha,\beta}+1})\|^2 < \infty. \tag{60}$$

(iv). Since h is convex and $\sigma_{t+1} \in \partial h(u_{t+1})$, we have

$$\sum_{t=1}^{+\infty} t^2 \mathrm{dist}^2(0, \partial f(u_{t+t_{\alpha,\beta}+1})) \leq \sum_{t=1}^{+\infty} t^2 \|\sigma_{t+t_{\alpha,\beta}+1} + \nabla g(u_{t+t_{\alpha,\beta}+1})\|^2 < +\infty. \tag{61}$$

Therefore, there exists a positive constant C such that [16, 26]

$$0 \leq \lim_{t\to+\infty} \left(t^3 \min_{1\leq i\leq t} \mathrm{dist}^2(0, \partial f(u_{i+t_{\alpha,\beta}+1})) \right) \leq C \lim_{t\to+\infty} \sum_{i=\lfloor \frac{t}{2} \rfloor}^{t} t^2 \mathrm{dist}^2(0, \partial f(u_{i+t_{\alpha,\beta}+1})) = 0. \tag{62}$$

This implies

$$\min_{1\le i\le t} \operatorname{dist}^2(0, \partial f(u_{i+t_{\alpha,\beta}+1})) = o\Big(\frac{1}{t^3}\Big). \tag{63}$$

(v). Since $\alpha > 3$, we have $\gamma \in (2, \alpha-1)$, and then $\eta = \alpha - 1 - \gamma > 0$. To show $\mathcal{E}_{t+1}(\gamma) - \mathcal{E}_t(\gamma) \le 0$ for $\gamma \in (2, \alpha-1)$, we require $\eta\beta + t(-\gamma+2) + (\alpha-\beta)(-\gamma+1) - 1 \le 0$, $t - \alpha \ge 0$, $t + \alpha - \beta - 1 \ge 0$ and $t(2\beta+1) - \alpha - \beta - 1 - L_{\nabla g}\lambda(\beta+1)\Big(t - 1 + \frac{\eta(t\beta-\alpha)}{t+\alpha-\beta-1}\Big) \ge 0$. Now, we define a sequence $p(t) := \frac{\eta(t\beta-\alpha)}{t+\alpha-\beta-1}$, then we observe that the sequence $\{p(t)\}_{t\ge1}$ is a convergent sequence, thus $\{p(t)\}_{t\ge1}$ is bounded. Let $|p(t)| \le C$ for all $t \ge 1$, where $C \in \mathbb{R}$. Then, using $p(t) = \frac{\eta(t\beta-\alpha)}{t+\alpha-\beta-1}$, we have

$$\begin{aligned}
&t(2\beta+1) - \alpha - \beta - 1 - L_{\nabla g}\lambda(\beta+1)\Big(t - 1 + \frac{\eta(t\beta-\alpha)}{t+\alpha-\beta-1}\Big)\\
&= t(((2\beta+1)) - L_{\nabla g}\lambda(\beta+1)) - \alpha - \beta - 1 + L_{\nabla g}\lambda(\beta+1) - L_{\nabla g}\lambda(\beta+1)p(k)\\
&\ge t(((2\beta+1)) - L_{\nabla g}\lambda(\beta+1)) - \alpha - \beta - 1 + L_{\nabla g}\lambda(\beta+1) - L_{\nabla g}\lambda(\beta+1)C\\
&= t(((2\beta+1)) - L_{\nabla g}\lambda(\beta+1)) - \alpha - \beta - 1 - L_{\nabla g}\lambda(\beta+1)(C-1).
\end{aligned} \tag{64}$$

Then we obtain $t \ge \frac{(\alpha-\beta)(-\gamma+1)-1+\eta\beta}{\gamma-2}$, $t \ge \alpha$, $t \ge \beta - \alpha + 1$ and $t \ge \frac{\alpha+\beta+1+L_{\nabla g}\lambda(\beta+1)(C-1)}{(2\beta+1))-L_{\nabla g}\lambda(\beta+1)}$. We choose

$$\begin{aligned}
T_{\alpha,\beta} = &\max\Big\{0, \Big\lfloor \frac{(\alpha-\beta)(-\gamma+1) - 1 + \eta\beta}{\gamma-2} \Big\rfloor + 1, \lfloor\alpha\rfloor + 1, \lfloor\beta - \alpha + 1\rfloor + 1,\\
&\Big\lfloor \frac{\alpha+\beta+1+L_{\nabla g}\lambda(\beta+1)(C-1)}{(2\beta+1)) - L_{\nabla g}\lambda(\beta+1)} \Big\rfloor + 1\Big\}.
\end{aligned} \tag{65}$$

Thus, $T_{\alpha,\beta}(\gamma-2) \ge (\alpha-\beta)(-\gamma+1) - 1 + \eta\beta$, $T_{\alpha,\beta} \ge \alpha$, $T_{\alpha,\beta} \ge \beta - \alpha + 1$ and $T_{\alpha,\beta}((2\beta+1)) - L_{\nabla g}\lambda(\beta+1)) \ge \alpha + \beta + 1 + L_{\nabla g}\lambda(\beta+1)(C-1)$. Now,

$$\begin{aligned}
\eta\beta + t(-\gamma+2) + (\alpha-\beta)(-\gamma+1) - 1 &= -t(\gamma-2) + (\alpha-\beta)(-\gamma+1) - 1 + \eta\beta\\
&\le -t(\gamma-2) + T_{\alpha,\beta}(\gamma-2)\\
&= -(\gamma-2)(t - T_{\alpha,\beta}).
\end{aligned} \tag{66}$$

Using Equation (64),

$$\begin{aligned}
&\frac{s^2(t+\alpha-\beta-1)}{2}\Big[t(2\beta+1)-\alpha-\beta-1-L_{\nabla g}\lambda(\beta+1)\Big(t-1+\frac{\eta(t\beta-\alpha)}{t+\alpha-\beta-1}\Big)\Big]\\
&\geq \frac{s^2(t-T_{\alpha,\beta})}{2}\big[t((2\beta+1)-L_{\nabla g}\lambda(\beta+1))-\alpha-\beta-1-L_{\nabla g}\lambda(\beta+1)(C-1)\big]\\
&= \frac{s^2(t-T_{\alpha,\beta})}{2}[t((2\beta+1)-L_{\nabla g}\lambda(\beta+1))-(\alpha+\beta+1+L_{\nabla g}\lambda(\beta+1)(C-1))]\\
&\geq \frac{s^2(t-T_{\alpha,\beta})}{2}\big[t((2\beta+1)-L_{\nabla g}\lambda(\beta+1))-T_{\alpha,\beta}(((2\beta+1))-L_{\nabla g}\lambda(\beta+1))\big]\\
&= \frac{s^2(t-T_{\alpha,\beta})^2}{2}((2\beta+1)-L_{\nabla g}\lambda(\beta+1)). \qquad (67)
\end{aligned}$$

From Equation (66), (67) and (43), we obtain

$$\begin{aligned}
\mathcal{E}_{t+1}(\gamma)-\mathcal{E}_t(\gamma) \leq & -s(\gamma-2)(t-T_{\alpha,\beta})(f(u_{t+1})-f(u^*))\\
& -\big[\eta(t-T_{\alpha,\beta})+\frac{\gamma\eta}{2}+\frac{\eta^2}{2}\big]\|u_{t+1}-u_t\|^2\\
& -\frac{s^2(t-T_{\alpha,\beta})^2}{2}((2\beta+1)-L_{\nabla g}\lambda(\beta+1))\|\sigma_{t+1}+\nabla g(\omega_t)\|^2. \qquad (68)
\end{aligned}$$

Since $\gamma-2>0$ and $(2\beta+1))-L_{\nabla g}\lambda(\beta+1)>0$, we get

$$\mathcal{E}_{t+1}(\gamma)-\mathcal{E}_t(\gamma) \leq -\big[\eta(t-T_{\alpha,\beta})+\frac{\gamma\eta}{2}+\frac{\eta^2}{2}\big]\|u_{t+1}-u_t\|^2, \qquad (69)$$

for all $t \geq T_{\alpha,\beta}+1$. Then $\{\mathcal{E}_t(\gamma)\}$ is positive non-increasing sequence, hence, it is bounded for all $\gamma \in (2, \alpha-1)$. Summing Equation (69) for $t = T_{\alpha,\beta}+1, T_{\alpha,\beta}+2, \cdots, N$, we get

$$\mathcal{E}_{N+1}(\gamma)+\eta\sum_{t=T_{\alpha,\beta}+1}^{N}(t-T_{\alpha,\beta})\|u_{t+1}-u_t\|^2 \leq \mathcal{E}_{T_{\alpha,\beta}+1}(\gamma). \qquad (70)$$

Also,

$$\sum_{t=T_{\alpha,\beta}+1}^{N}(t-T_{\alpha,\beta})\|u_{t+1}-u_t\|^2 \leq \frac{\mathcal{E}_{T_{\alpha,\beta}+1}(\gamma)}{\eta}. \qquad (71)$$

Since $\{\mathcal{E}_t(\gamma)\}$ is bounded, as $N \to +\infty$, we obtain

$$\sum_{t=T_{\alpha,\beta}+1}^{+\infty}(t-T_{\alpha,\beta})\|u_{t+1}-u_t\|^2 < +\infty. \qquad (72)$$

Replace $t - T_{\alpha,\beta}$ with t, we get

$$\sum_{t=1}^{+\infty} t\|u_{t+T_{\alpha,\beta}+1} - u_{t+T_{\alpha,\beta}}\|^2 < +\infty. \tag{73}$$

□

Remark 1. *The ODE in* (16) *can be simplified by setting* $s = 1$ *and letting* $\alpha = \beta$ *only in the coefficient of* $\nabla f(X(t))$*. This leads to a reduction of the ODE* (16) *to the form (DIN-AVD)*$_{(\alpha,\beta,1+\frac{\beta}{t})}$*. By discretizing the ODE in a similar manner, we can derive an algorithm that resembles that of [21], with modifications only in the coefficients of* $(u_t - u_{t-1})$ *and* $(\sigma_t + \nabla g(\omega_{t-1}))$*.*

5 Convergence of the iterates

Lemma 9. *Let* $\alpha \geq 3$*, and let* $\{a_t\}_{t\geq 1}$ *and* $\{b_t\}_{t\geq 1}$ *be two sequences in* $[0, +\infty)$ *such that*

$$a_{t+1} \leq \frac{t-1-\alpha}{t-1} a_t + b_t,$$

for all $t \geq 1$*. If* $\sum_{t=1}^{+\infty} t b_t < +\infty$*, then* $\sum_{t=1}^{+\infty} a_t < +\infty$*.*

Proof. Using $\alpha \geq 3$, we get $t - 1 - \alpha \leq t - 3$. Therefore, we write

$$a_{t+1} \leq \frac{t-3}{t-1} a_t + b_t. \tag{74}$$

Multiplying $(t+1)^2$ in Equation (74), we get

$$\begin{aligned}(t+1)^2 a_{t+1} &\leq \frac{(t+1)^2(t-3)}{t-1} a_t + (t+1)^2 b_t \\ &\leq t^2 a_t + (t+1)^2 b_t. \end{aligned} \tag{75}$$

Equation (75) summation for $i = 1, 2 \ldots, t-1$, we obtain

$$t^2 a_t \leq a_1 + \sum_{i=1}^{t-1} (i+1)^2 b_i.$$

Dividing by t^2, and summing for $t = 2, \ldots, N$, we get

$$\sum_{t=2}^{N} a_t \leq a_1 \sum_{t=2}^{N} \frac{1}{t^2} + \sum_{t=2}^{N} \frac{1}{t^2} \sum_{i=1}^{t-1} (i+1)^2 b_i. \tag{76}$$

Observe that

$$\sum_{t=i+1}^{+\infty} \frac{1}{t^2} \leq \int_{t=i+1}^{+\infty} \frac{1}{i^2} di = \frac{1}{i+1}. \tag{77}$$

Applying Fubini's Theorem to Equation (76)'s last sum, we obtain

$$\sum_{t=2}^{N} a_d \leq a_1 \sum_{t=2}^{N} \frac{1}{t^2} + \sum_{i=1}^{N-1} \Big(\sum_{t=i+1}^{+\infty} \frac{1}{t^2} \Big) (i+1)^2 b_i.$$

Since $\sum_{t=1}^{+\infty} t b_t < +\infty$, and from Equation (77), we get

$$\begin{aligned} \sum_{t=2}^{N} a_t &\leq a_1 \sum_{t=2}^{N} \frac{1}{t^2} + \sum_{i=1}^{N-1} \frac{(i+1)^2}{i+1} b_i \\ &\leq a_1 \sum_{t=2}^{N} \frac{1}{t^2} + \sum_{i=1}^{N-1} (i+1) b_i < +\infty. \end{aligned} \tag{78}$$

Hence, proved. □

Theorem 10. *Let Assumption* (1) *hold,* $\alpha > 3$ *and* $2\beta + 1 - L_{\nabla g}\lambda(\beta+1) > 0$*. Let the sequences* $\{u_t, \omega_t, \sigma_t\}_{t \geq 1}$ *be produced by IFBASC. Then*

$$\|u_{t+1} - u_t\| = \mathcal{O}\Big(\frac{1}{t}\Big), \tag{79}$$

and $\{u_t\}_{t\geq 1}$ *converges, with* $\lim_{t\to+\infty} u_t = u^* \in S$.

Proof. From Equation (69), the Lyapunov function $\{\mathcal{E}_t(\gamma)\}$ is bounded for any $\gamma \in (2, \alpha - 1)$. By connecting this with $\eta = \alpha - 1 - \gamma > 0$ and (23), we deduce that $\{u_t\}_{t\geq 1}$ is bounded. Utilizing Equation (55), we obtain

$$\lim_{t\to+\infty} (t - t_{\alpha,\beta}) \|\sigma_{t+1} + \nabla g(\omega_t)\| = 0. \tag{80}$$

This implies $\{s(t\beta - \alpha)\}(\sigma_{t+1} + \nabla g(\omega_t))$ is also bounded. From Equation (49), using (23) and (24), we get that the bouundedness of $\{\mathcal{E}_t(\alpha - 1)\}_{t\geq 1}$ implies that $\{(\alpha-1)(u_{t+1} - u^*) + (t-\alpha)(u_{t+1} - u_t) + s(\beta t - \alpha)(\sigma_{t+1} + \nabla g(\omega_t))\}_{t\geq 1}$ is bounded. Since $\{u_t\}_{t\geq 1}$ and $\{s(t\beta - \alpha)\}(\sigma_{t+1} + \nabla g(\omega_t))$ are bounded, this implies that $\{(t-\alpha)(u_{t+1} - u_t)\}_{t\geq 1}$ is bounded. Hence,

$$\|u_{t+1} - u_t\| = \mathcal{O}\Big(\frac{1}{t}\Big). \tag{81}$$

Since h is lower-semicontinuous, the function f is also lower-semicontinuous, and for each sequential cluster point u^∞ of $\{u_t\}_{t\geq 1}$ such that $\lim_{t_j\to+\infty} u_{t_j} = u^\infty$, we have

$$f(u^\infty) \leq \lim_{t_j\to+\infty} \inf f(u_{t_j}) \overset{\text{Theorem(8)}(i)}{=} \min f.$$

Therefore, each sequential cluster point of $\{u_t\}_{t\geq 1}$ belongs to S. Now, fix $u^* \in S$, and we will show that $\{\|u_t - u^*\|\}_{t\geq 1}$ converges. Let $h_t = \frac{1}{2}\|u_t - u^*\|^2$, using (13), we get

$$h_{t+1} - h_t = \langle u_{t+1} - u^*, u_{t+1} - u_t\rangle - \frac{1}{2}\|u_{t+1} - u_t\|^2. \tag{82}$$

Using step 3 of Algorithm (1), the above equation becomes

$$\begin{aligned} h_{t+1} - h_t = {}& \langle u_{t+1} - u^*, u_{t+1} - \omega_t\rangle + \frac{t-1-\alpha}{t-1}\langle u_{t+1} - u^*, u_t - u_{t-1}\rangle \\ &+ \frac{s(\beta(t-1)-\alpha)}{t-1}\langle u_{t+1} - u^*, \sigma_t + \nabla g(\sigma_{t-1})\rangle - \frac{1}{2}\|u_{t+1} - u_t\|^2. \end{aligned} \tag{83}$$

As like Equation (82), we have

$$h_t - h_{t-1} = \langle u_t - u^*, u_t - u_{t-1}\rangle - \frac{1}{2}\|u_t - u_{t-1}\|^2, \tag{84}$$

and using (83) and (84), we get

$$\begin{aligned} & h_{t+1} - h_t - \frac{t-1-\alpha}{t-1}(h_t - h_{t-1}) \\ &= \langle u_{t+1} - u^*, u_{t+1} - \omega_t\rangle + \frac{t-1-\alpha}{t-1}\langle u_{t+1} - u^*, u_t - u_{t-1}\rangle \\ &+ \frac{s(\beta(t-1)-\alpha)}{t-1}\langle u_{t+1} - u^*, \sigma_t + \nabla g(\sigma_{t-1})\rangle - \frac{1}{2}\|u_{t+1} - u_t\|^2 \\ &- \frac{t-1-\alpha}{t-1}\langle u_t - u^*, u_t - u_{t-1}\rangle + \frac{t-1-\alpha}{2(t-1)}\|u_t - u_{t-1}\|^2. \end{aligned} \tag{85}$$

Utilizing step 5 of Algorithm (1) in the above equation, we obtain

$$\begin{aligned} & h_{t+1} - h_t - \frac{t-1-\alpha}{t-1}(h_t - h_{t-1}) \\ &\leq -\lambda\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle + \frac{t-1-\alpha}{t-1}\langle u_{t+1} - u_t, u_t - u_{t-1}\rangle \\ &+ \frac{s(\beta(t-1)-\alpha)}{t-1}\langle u_{t+1} - u^*, \sigma_t + \nabla g(\sigma_{t-1})\rangle + \frac{t-1-\alpha}{2(t-1)}\|u_t - u_{t-1}\|^2. \end{aligned} \tag{86}$$

Let

$$\Delta_{t+1} := h_{t+1} - h_t + s\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle, \tag{87}$$

we have

$$\begin{aligned}
&\Delta_{t+1} - \frac{t-1-\alpha}{t-1}\Delta_t \\
&\overset{(87)}{=} h_{t+1} - h_t - \frac{t-1-\alpha}{t-1}(h_t - h_{t-1}) \\
&+ s\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle - \frac{s(t-1-\alpha)}{t-1}\langle u_t - u^*, \sigma_t + \nabla g(\sigma_{t-1})\rangle \\
&\overset{(86)}{\leq} -\lambda\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle + \frac{t-1-\alpha}{t-1}\langle u_{t+1} - u_t, u_t - u_{t-1}\rangle \\
&+ \frac{s(\beta(t-1)-\alpha)}{t-1}\langle u_{t+1} - u_t + u_t - u^*, \sigma_t + \nabla g(\sigma_{t-1})\rangle + \frac{t-1-\alpha}{2(t-1)}\|u_t - u_{t-1}\|^2 \\
&+ s\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle - \frac{s(t-1-\alpha)}{t-1}\langle u_t - u^*, \sigma_t + \nabla g(\sigma_{t-1})\rangle. \qquad (88)
\end{aligned}$$

Since $\lambda = s + \beta s$, the above equation becomes

$$\begin{aligned}
&\Delta_{t+1} - \frac{t-1-\alpha}{t-1}\Delta_t \\
&\leq -\beta s\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle + \frac{t-1-\alpha}{t-1}\langle u_{t+1} - u_t, u_t - u_{t-1}\rangle \\
&+ \frac{s(\beta(t-1)-\alpha)}{t-1}\langle u_{t+1} - u_t, \sigma_t + \nabla g(\sigma_{t-1})\rangle + \frac{t-1-\alpha}{2(t-1)}\|u_t - u_{t-1}\|^2 \\
&+ \frac{s}{t-1}(\beta t - \beta - \alpha - t + 1 + \alpha)\langle u_t - u^*, \sigma_t + \nabla g(\sigma_{t-1})\rangle \\
&= -\beta s\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle + \frac{t-1-\alpha}{t-1}\langle u_{t+1} - u_t, u_t - u_{t-1}\rangle \\
&+ \frac{s(\beta(t-1)-\alpha)}{t-1}\langle u_{t+1} - u_t, \sigma_t + \nabla g(\sigma_{t-1})\rangle + \frac{t-1-\alpha}{2(t-1)}\|u_t - u_{t-1}\|^2 \\
&+ s(\beta-1)\langle u_t - u^*, \sigma_t + \nabla g(\sigma_{t-1})\rangle. \qquad (89)
\end{aligned}$$

Using the Cauchy-Schwarz inequality in the above equation and as $\frac{t-1-\alpha}{t-1} \leq 1$, we obtain

$$\begin{aligned}
&\Delta_{t+1} - \frac{t-1-\alpha}{t-1}\Delta_t \\
&\leq -\beta s\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle + \|u_{t+1} - u_t\|\|u_t - u_{t-1}\| \\
&+ s\beta\|u_{t+1} - u_t\|\|\sigma_t + \nabla g(\omega_{t-1})\| + \frac{1}{2}\|u_t - u_{t-1}\|^2 + s(\beta-1)\|u_t - u^*\|\|\sigma_t + \nabla g(\omega_{t-1})\|. \\
&\qquad (90)
\end{aligned}$$

Since $0 \in \partial f(u^*) = \nabla g(u^*) + \partial h(u^*)$, we have $-\nabla g(u^*) \in \partial h(u^*)$. Using the fact $\sigma_{t+1} \in \partial h(u_{t+1})$ and monotonicity of ∂h, we obtain

$$\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(u^*)\rangle \geq 0.$$

Then, we establish

$$\begin{aligned}\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle &= \langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(u^*) + \nabla g(\omega_t) - \nabla g(u^*)\rangle \\ &\geq \langle u_{t+1} - u^*, \nabla g(\omega_t) - \nabla g(u^*)\rangle \\ &= \langle u_{t+1} - \omega_t, \nabla g(\omega_t) - \nabla g(u^*)\rangle + \langle \omega_t - u^*, \nabla g(\omega_t) - \nabla g(u^*)\rangle.\end{aligned} \tag{91}$$

Utilizing Cauchy-Schwarz inequality, Equation (91) becomes

$$\begin{aligned}\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle &\geq \langle u_{t+1} - u^*, \nabla g(\omega_t) - \nabla g(u^*) \\ &\overset{(12)}{\geq} -\|u_{t+1} - \omega_t\|\|\nabla g(\omega_t) - \nabla g(u^*)\| + \frac{1}{L_{\nabla g}}\|\nabla g(\omega_t) - \nabla g(u^*)\|^2.\end{aligned} \tag{92}$$

We have

$$\|x\|\|y\| \leq \frac{1}{2}(\|x\|^2 + \|y\|^2). \tag{93}$$

Put $x = \frac{\|\nabla g(\omega_t) - \nabla g(u^*)\|}{\sqrt{\frac{L_{\nabla g}}{2}}}$ and $y = \sqrt{\frac{L_{\nabla g}}{2}}\|u_{t+1} - \omega_t\|$ in Equation (93), we obtain

$$-\|u_{t+1} - \omega_t\|\|\nabla g(\omega_t) - \nabla g(u^*)\| + \frac{1}{L_{\nabla g}}\|\nabla g(\omega_t) - \nabla g(u^*)\|^2 \geq \frac{-L_{\nabla g}}{4}\|u_{t+1} - \omega_t\|^2. \tag{94}$$

Equation (92) becomes

$$\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle \geq \frac{-L_{\nabla g}}{4}\|u_{t+1} - \omega_t\|^2. \tag{95}$$

Putting Equation (95) in (90), and from stage 5 of Algorithm 1, we obtain

$$\Delta_{t+1} - \frac{t - 1 - \alpha}{t - 1}\Delta_t \leq e_t, \tag{96}$$

where

$$\begin{aligned}e_t :\overset{(93)}{=} &\frac{L_{\nabla g}\lambda^2\beta s}{4}\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 + \frac{1 + \beta s}{2}\|u_{t+1} - u_t\|^2 \\ &+ \|u_t - u_{t-1}\|^2 + \frac{\beta s}{2}\|\sigma_t + \nabla g(\sigma_{t-1})\|^2 + s(\beta - 1)\|u_t - u^*\|\|\sigma_t + \nabla g(\sigma_{t-1})\|.\end{aligned} \tag{97}$$

From Equation (56) and (73), we obtain

$$\sum_{t=1}^{+\infty} t^2\|\sigma_{t+t_{\alpha,\beta}+1} + \nabla g(\omega_{t+t_{\alpha,\beta}})\|^2 < +\infty, \quad \sum_{t=1}^{+\infty} t\|u_{t+T_{\alpha,\beta}+1} - u_{t+T_{\alpha,\beta}}\|^2 < +\infty. \tag{98}$$

Since $\{u_t\}_{t\geq 1}$ is bounded, this implies $\{\|u_t-u^*\|\}_{t\geq 1}$ is bounded. Let $\|u_t-u^*\|\leq c$ for all $t\geq 1$. Then

$$\|u_t-u^*\|\|\sigma_t+\nabla g(\sigma_{t-1})\|\leq c\|\sigma_t+\nabla g(\sigma_{t-1})\|.$$

Furthermore,

$$\begin{aligned}\sum_{t=1}^{+\infty} t\|u_t-u^*\|\|\sigma_t+\nabla g(\sigma_{t-1})\| &\leq c\sum_{t=1}^{+\infty} t\|\sigma_t+\nabla g(\sigma_{t-1})\|\\ &\leq c\sum_{t=1}^{+\infty} t^2\|\sigma_t+\nabla g(\sigma_{t-1})\|^2<+\infty. \end{aligned} \tag{99}$$

From Theorem 8 (v), Equation (98) and (99), we obtain

$$\sum_{t=1}^{+\infty} te_t<+\infty. \tag{100}$$

We can further derive

$$[\Delta_{t+1}]_+\leq \frac{t-1-\alpha}{t-1}[\Delta_t]_++e_t, \tag{101}$$

utilizing Equation (100) and applying Lemma (9), we obtain

$$\sum_{t=1}^{+\infty}[\Delta_t]_+<+\infty. \tag{102}$$

From Equation (87), we get

$$\sum_{t=1}^{+\infty}[h_{t+1}-h_t]_+\leq \sum_{t=1}^{+\infty}[\Delta_{t+1}]_++s\sum_{t=1}^{+\infty}\|u_t-u^*\|\|\sigma_t+\nabla g(\sigma_{t-1})\|. \tag{103}$$

Using the boundedness of $\{u_t\}_{t\geq 1}$, we have

$$\sum_{t=1}^{+\infty}\|u_t-u^*\|\|\sigma_t+\nabla g(\sigma_{t-1})\| \overset{(93)}{\leq} \sum_{t=1}^{+\infty}\Big(\frac{\|u_t-u^*\|^2}{2t^2}+\frac{t^2}{2}\|\sigma_t+\nabla g(\sigma_{t-1})\|^2\Big) \overset{(56)}{<} +\infty. \tag{104}$$

Now, Equation (103) becomes

$$\sum_{t=1}^{+\infty}[h_{t+1}-h_t]_+ \overset{(102),(104)}{<} +\infty. \tag{105}$$

We have

$$h_{t+1}\leq h_t+[h_{t+1}-h_t]_+, \tag{106}$$

applying Lemma (4), we conclude $\lim_{t\to+\infty} h_t = \frac{1}{2}\lim_{t\to+\infty}\|u_t - u^*\|^2$ exists. As all sequential cluster point of $\{u_t\}_{t\geq 1}$ lie in S, we conclude that $\lim_{t\to+\infty} u_t = u^* \in S$ using Lemma (5). □

Theorem 11. *Let Assumption* (1) *hold,* $\alpha > 3$, $2\beta+1-L_{\nabla g}\lambda(\beta+1) > 0$. *Let the sequences* $\{(u_t, \omega_t, \sigma_t)\}_{t\geq 1}$ *be produced by IFBASC. Then, as* $t \to +\infty$, *we have*

(i). $f(u_t) - \min f = o\Big(\frac{1}{t^2}\Big)$.

(ii). $\|u_{t+1} - u_t\| = o\Big(\frac{1}{t}\Big)$.

Proof. From Theorem (10), we have the sequence $\{u_t\}_{t\geq 1}$ converges in S. Consequently, set $u^* = \lim_{t\to+\infty} u_t \in S$. We have

$$\mathcal{E}_t(\alpha-1) \overset{(23)}{\geq} 0, \text{ and } \mathcal{E}_{t+1}(\alpha-1) \overset{(48)}{\leq} \mathcal{E}_t(\alpha-1), \tag{107}$$

for all $t \geq t_{\alpha,\beta}+1$. This yields $\lim_{t\to+\infty}\mathcal{E}_t(\alpha-1)$ exists. From Equation (55) and Theorem (10), we obtain

$$\lim_{t\to+\infty}(t - t_{\alpha,\beta})(\sigma_{t+1} + \nabla g(\omega_t)) = 0, \text{ and } \sup_{t\geq 1} t\|u_{t+1} - u_t\| < +\infty. \tag{108}$$

Combining Equation (108) with $u_t \to u^*$, Equation (23) indicates that

$$\begin{aligned}\lim_{t\to+\infty}\mathcal{E}_t(\alpha-1) = \lim_{t\to+\infty} & s(t+\alpha-\beta-1)(t-\alpha)(f(u_t) - f(u^*)) \\ & + \frac{(t-1-\alpha)^2}{2}\|u_t - u_{t-1}\|^2 \text{ exists.}\end{aligned} \tag{109}$$

We set

$$\lim_{t\to+\infty} s(t+\alpha-\beta-1)(t-\alpha)(f(u_t) - f(u^*)) + \frac{(t-1-\alpha)^2}{2}\|u_t - u_{t-1}\|^2 = l \geq 0. \tag{110}$$

Now, we have to show that $l = 0$. Assume that $l > 0$, then there exists $t_0 \geq 1$ such that

$$s(t+\alpha-\beta-1)(t-\alpha)(f(u_t) - f(u^*)) + \frac{(t-1-\alpha)^2}{2}\|u_t - u_{t-1}\|^2 \geq \frac{l}{2} > 0,$$

for all $t \geq t_0$. This condition yields

$$\begin{aligned}&\sum_{t=t_0}^{+\infty}\Big[s(t+\alpha-\beta-1)(f(u_t) - f(u^*)) + \frac{(t-1-\alpha)}{2}\|u_t - u_{t-1}\|^2\Big] \\ &\geq \sum_{t=t_0}^{+\infty}\frac{1}{t}\Big[s(t+\alpha-\beta-1)(t-\alpha)(f(u_t) - f(u^*)) + \frac{(t-1-\alpha)^2}{2}\|u_t - u_{t-1}\|^2\Big] \\ &\geq \frac{l}{2}\sum_{t=t_0}^{+\infty}\frac{1}{t} = +\infty.\end{aligned} \tag{111}$$

Nevertheless, when we combine Theorem (8) (ii) with (v), we get

$$\sum_{t=1}^{+\infty}\Big[s(t+\alpha-\beta-1)(f(u_t)-f(u^*))+\frac{(t-1-\alpha)}{2}\|u_t-u_{t-1}\|^2\Big]<+\infty, \tag{112}$$

which is a contradiction to Equation (111). As a result, $l=0$, and our result follows. $\square$

6 Inexact accelerated forward-backward algorithm

An alternative way to model inexactness in the proximal step (Step 4 of IFBASC) is to compute u_{t+1} such that

$$0\in\partial_{\varepsilon_{t+1}}h(u_{t+1})+\nabla g(\omega_t)+\frac{1}{\lambda}(u_{t+1}-\omega_t),$$

which is equivalent to requiring

$$\Psi_t(u_{t+1})\leq\inf_u\Psi_t(u)+\varepsilon_{t+1},$$

where

$$\Psi_t(u):=h(u)+\frac{1}{2\lambda}\left\|u-\big(\omega_t-\lambda\nabla g(\omega_t)\big)\right\|^2.$$

Here, $\varepsilon_{t+1}\geq 0$ denotes the admissible error, and $\partial_\varepsilon h(u)$ with $\varepsilon>0$ is the ε-subdifferential of h at $u\in\operatorname{dom}(h)$, defined by

$$\partial_\varepsilon h(u)=\left\{w\in\mathbb{R}^d\;\middle|\;h(z)\geq h(u)+\langle w,z-u\rangle-\varepsilon,\forall z\in\mathbb{R}^d\right\}. \tag{113}$$

This notion of inexactness was introduced in [38], where it is shown that inexact AFB schemes retain the convergence rate $\mathcal{O}(1/t^2)$ provided that $\varepsilon_t=\mathcal{O}(1/t^q)$ with $q>3$.

A further type of inexactness in the proximal step (Step 4 of IFBASC) consists in computing u_{t+1} such that

$$\vartheta_{t+1}\in\partial h(u_{t+1})+\nabla g(\omega_t)+\frac{1}{\lambda}(u_{t+1}-\omega_t),$$

which can be equivalently expressed as

$$\operatorname{dist}\big(0,\partial\Psi_t(u_{t+1})\big)\leq\|\vartheta_{t+1}\|.$$

Here, $\vartheta_{t+1}\in\mathbb{R}^d$ denotes an error vector. This notion of approximate proximal computation was originally introduced in [33] and further studied in [5, 20]. In particular, Attouch et al. [5] showed that the AFB method achieves a convergence rate $\mathcal{O}(1/t^2)$ of objective gap provided that $\sum_{t=1}^{+\infty}t\|\vartheta_{t+1}\|<+\infty$.

Combining the above ideas, one can characterize an estimated solution to Stage 4 of IFBASC using a pair of residuals $(\vartheta_{t+1}, \varepsilon_{t+1}) \in \mathbb{R}^d \times \mathbb{R}_+$ satisfying

$$\vartheta_{t+1} \in \partial_{\varepsilon_{t+1}} h(u_{t+1}) + \nabla g(\omega_t) + \frac{1}{\lambda}(u_{t+1} - \omega_t).$$

Motivated by this, we introduce the following inexact scheme. A related approach was considered in [11], where the explicit estimation of the error sequence ϑ_{t+1} is required for the correction step. In contrast, this approach only requires prescribing a tolerance $\|\vartheta_{t+1}\|$ such that

$$\operatorname{dist}\Big(0, \partial_{\varepsilon_{t+1}} h(u_{t+1}) + \nabla g(\omega_t) + \frac{1}{\lambda}(u_{t+1} - \omega_t)\Big) \leq \|\vartheta_{t+1}\|.$$

In this setting, there exists σ_{t+1} such that

$$\sigma_{t+1} \in \partial_{\varepsilon_{t+1}} h(u_{t+1}) - \vartheta_{t+1},$$

which shows that the classical subgradient correction is replaced by an ε_{t+1}-subgradient correction involving an additional error vector.

Now, we propose the inexact version of Algorithm (1).

Algorithm 2 Inexact inertial forward-backward algorithm with subgradient correction (I-IFBASC)

```
Initialize: Let $u_1 = u_2 = \omega_1 \in \mathbb{R}^d$, $\sigma_2 \in \partial h(u_1)$, $\alpha \geq 3$, $s > 0$, $\beta \geq 0$, $\lambda = s + \beta s$.
for $t = 2, \ldots$ do
    $\omega_t = u_t + \frac{t-1-\alpha}{t-1}(u_t - u_{t-1}) + s\Big(\beta - \frac{\alpha}{t-1}\Big)\big(\sigma_t + \nabla g(\omega_{t-1})\big)$,
    Set error $(\vartheta_{t+1}, \varepsilon_{t+1}) \in \mathbb{R}^d \times \mathbb{R}_+$. Find $u_{t+1}$ satisfy
        $\vartheta_{t+1} \in \partial_{\varepsilon_{t+1}} h(u_{t+1}) + \nabla g(\omega_t) + \frac{1}{\lambda}(u_{t+1} - \omega_t)$.
    $\sigma_{t+1} = -\nabla g(\omega_t) - \frac{1}{\lambda}(u_{t+1} - \omega_t)$.
    until a termination criterion is satisfied
end for
return $(u_t, \omega_t, \sigma_t)$
```

6.1 Analyzing the convergence of inexact version

Consider the sequence $\{(u_t, \omega_t, \sigma_t)\}_{t\geq 1}$ generated by Algorithm (2) with the error sequence $\{(\vartheta_t, \varepsilon_t)\}_{t\geq 1}$ and $u^* \in S$. For $\gamma \in [2, \alpha - 1]$ and $\eta = \alpha - \gamma - 1$, we derive the new energy function $\mathcal{E}_t^{\vartheta}(\gamma)$ which is given by;

$$\mathcal{E}_t^{\vartheta}(\gamma) := \mathcal{E}_t(\gamma) - \sum_{i=1}^{t} s(i + \alpha - \beta - 2)\langle \vartheta_i, \pi_i(\gamma)\rangle, \tag{114}$$

where the sequences $\{\mathcal{E}_t(\gamma)\}_{t\geq 1}$ and $\{\pi_t(\gamma)\}_{t\geq 1}$ is defined on Equation (23) and (24), respectively.

Lemma 12. *Let Assumption 1 hold, and* $\{(u_t, \omega_t, \sigma_t)\}_{t\geq 1}$ *be the sequences produced by Algorithm* (2) *with* $\{(\vartheta_t, \varepsilon_t)\}_{t\geq 1}$ *as the error sequence, and let* $u^* \in S$*. Consider the energy function* $\mathcal{E}_t^{\vartheta}(\gamma)$ *in* (114)*. Then, we have*

$$\begin{aligned}
&\mathcal{E}_{t+1}^{\vartheta}(\gamma) - \mathcal{E}_t^{\vartheta}(\gamma)\\
&\leq s\big[\eta\beta + t(-\gamma+2) + (\alpha-\beta)(-\gamma+1) - 1\big](f(u_{t+1}) - f(u^*))\\
&- \big[\eta(t-\alpha) + \frac{\gamma\eta}{2} + \frac{\eta^2}{2}\big]\|u_{t+1} - u_t\|^2\\
&- \frac{s^2(t+\alpha-\beta-1)}{2}\Big[t(2\beta+1) - \alpha - \beta - 1 - L_{\nabla g}\lambda(\beta+1)\Big(t - 1 + \frac{\eta(t\beta-\alpha)}{t+\alpha-\beta-1}\Big)\Big]\\
&\times \|\sigma_{t+1} + \nabla g(\omega_t)\|^2\\
&+ s[(t+\alpha-\beta-1)(t-1) + \eta(t\beta-\alpha)]\varepsilon_{t+1}\\
&+ \eta s(\beta+1)(t-1)\langle \vartheta_{t+1}, u_{t+1} - u_t\rangle - s^2(t+\alpha-\beta-1)(t\beta-\alpha)\langle \vartheta_{t+1}, \sigma_{t+1} + \nabla g(\omega_t)\rangle.
\end{aligned}$$

Proof. From stages 4 and 5 of Algorithm 2, we derive

$$\sigma_{t+1} + \vartheta_{t+1} \in \partial h_{\varepsilon_{t+1}}(u_{t+1}). \tag{115}$$

From Equation (113), for any $u \in \mathbb{R}^d$, we have

$$\begin{aligned}
\langle \sigma_{t+1}, u_{t+1} - u\rangle &= \langle \sigma_{t+1} + \vartheta_{t+1}, u_{t+1} - u\rangle - \langle \vartheta_{t+1}, u_{t+1} - u\rangle\\
&\geq h(u_{t+1}) - h(u) - \langle \vartheta_{t+1}, u_{t+1} - u\rangle - \varepsilon_{t+1}. \qquad (116)
\end{aligned}$$

Adding Equation (33) and (116), we yields

$$\begin{aligned}
&\langle u_{t+1} - u, \sigma_{t+1} + \nabla g(\omega_t)\rangle\\
&\geq g(u_{t+1}) + h(u_{t+1}) - g(u) - h(u) - \frac{L_{\nabla g}\lambda^2}{2}\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 - \langle \vartheta_{t+1}, u_{t+1} - u\rangle - \varepsilon_{t+1}\\
&= f(u_{t+1}) - f(u) - \frac{L_{\nabla g}\lambda^2}{2}\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 - \langle \vartheta_{t+1}, u_{t+1} - u\rangle - \varepsilon_{t+1}. \qquad (117)
\end{aligned}$$

Implementing u^* and u_t in place of u in Equation (117), we obtain

$$\begin{aligned}
&\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle\\
&\geq f(u_{t+1}) - f(u^*) - \frac{L_{\nabla g}\lambda^2}{2}\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 - \langle \vartheta_{t+1}, u_{t+1} - u^*\rangle - \varepsilon_{t+1}, \qquad (118)
\end{aligned}$$

and

$$\begin{aligned}
&\langle u_{t+1} - u_t, \sigma_{t+1} + \nabla g(\omega_t)\rangle \\
&\geq f(u_{t+1}) - f(u_t) - \frac{L_{\nabla g}\lambda^2}{2}\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 - \langle \vartheta_{t+1}, u_{t+1} - u_t\rangle - \varepsilon_{t+1}. \qquad (119)
\end{aligned}$$

Since $\eta = \alpha - 1 - \gamma$, we have

$$\begin{aligned}
&- \gamma s(t+\alpha-\beta-1)\langle \vartheta_{t+1}, u_{t+1} - u^*\rangle - \gamma s(t+\alpha-\beta-1)\varepsilon_{t+1} \\
&- s[(t+\alpha-\beta-1)(t-\alpha+\eta) + \eta(t\beta-\alpha)]\langle \vartheta_{t+1}, u_{t+1} - u_t\rangle \\
&- s[(t+\alpha-\beta-1)(t-\alpha+\eta) + \eta(t\beta-\alpha)]\varepsilon_{t+1} \\
&= -s[(t+\alpha-\beta-1)(t-1) + \eta(t\beta-\alpha)]\varepsilon_{t+1} - \gamma s(t+\alpha-\beta-1)\langle \vartheta_{t+1}, u_{t+1} - u^*\rangle \\
&- s[(t+\alpha-\beta-1)(t-\alpha+\eta) + \eta(t\beta-\alpha)]\langle \vartheta_{t+1}, u_{t+1} - u_t\rangle. \qquad (120)
\end{aligned}$$

Now, we determine

$$\begin{aligned}
&\gamma s(t+\alpha-\beta-1)\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle \\
&+ s\big[(t+\alpha-\beta-1)(t-\alpha+\eta) + \eta(t\beta-\alpha)\big]\langle u_{t+1} - u_t, \sigma_{t+1} + \nabla g(\omega_t)\rangle \\
&\overset{(118),(119)}{\geq} s(t+\alpha-\beta-1)(f(u_{t+1}) - f(u^*)) \\
&+ s\big[(t+\alpha-\beta-1)(t-\alpha+\eta) + \eta(t\beta-\alpha)\big](f(u_{t+1}) - f(u_t)) \\
&- \frac{L_{\nabla g}\lambda^2}{2}s\big[\gamma(t+\alpha-\beta-1) + (t+\alpha-\beta-1)(t-\alpha+\eta) + \eta(t\beta-\alpha)\big]\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 \\
&- \gamma s(t+\alpha-\beta-1)\langle \vartheta_{t+1}, u_{t+1} - u^*\rangle - \gamma s(t+\alpha-\beta-1)\varepsilon_{t+1} \\
&- s[(t+\alpha-\beta-1)(t-\alpha+\eta) + \eta(t\beta-\alpha)]\langle \vartheta_{t+1}, u_{t+1} - u_t\rangle \\
&- s[(t+\alpha-\beta-1)(t-\alpha+\eta) + \eta(t\beta-\alpha)]\varepsilon_{t+1} \\
&\overset{(38),(120)}{\geq} J_1 - J_2 - J_3 \\
&- s[(t+\alpha-\beta-1)(t-1) + \eta(t\beta-\alpha)]\varepsilon_{t+1} - \gamma s(t+\alpha-\beta-1)\langle \vartheta_{t+1}, u_{t+1} - u^*\rangle \\
&- s[(t+\alpha-\beta-1)(t-\alpha+\eta) + \eta(t\beta-\alpha)]\langle \vartheta_{t+1}, u_{t+1} - u_t\rangle. \qquad (121)
\end{aligned}$$

Putting Equation (121) in (31), we obtain

$$\begin{aligned}
\mathcal{E}_{t+1}(\gamma) - \mathcal{E}_t(\gamma) &\overset{(43)}{\leq} F_1 + X_1 + G_1 + \gamma s(t+\alpha-\beta-1)\langle \vartheta_{t+1}, u_{t+1} - u^*\rangle \\
&+ s[(t+\alpha-\beta-1)(t-1) + \eta(t\beta-\alpha)]\varepsilon_{t+1} \\
&+ s[(t+\alpha-\beta-1)(t-\alpha+\eta) + \eta(t\beta-\alpha)]\langle \vartheta_{t+1}, u_{t+1} - u_t\rangle. \qquad (122)
\end{aligned}$$

Applying (114), we have

$$\mathcal{E}^{\vartheta}_{t+1}(\gamma) - \mathcal{E}^{\vartheta}_t(\gamma) = \mathcal{E}_{t+1}(\gamma) - \mathcal{E}_t(\gamma) - s(t+\alpha-\beta-1)\langle \vartheta_{t+1}, \pi_{t+1}(\gamma)\rangle. \qquad (123)$$

Equation 123 becomes

$$\begin{aligned}
&\mathcal{E}^{\vartheta}_{t+1}(\gamma) - \mathcal{E}^{\vartheta}_{t}(\gamma) \\
&\overset{(24),(122)}{\leq} F_1 + X_1 + G_1 + \gamma s(t+\alpha-\beta-1)\langle \vartheta_{t+1}, u_{t+1} - u^* \rangle \\
&+ s[(t+\alpha-\beta-1)(t-1) + \eta(t\beta-\alpha)]\varepsilon_{t+1} \\
&+ s[(t+\alpha-\beta-1)(t-\alpha+\eta) + \eta(t\beta-\alpha)]\langle \vartheta_{t+1}, u_{t+1} - u_t \rangle \\
&- \gamma s(t+\alpha-\beta-1)\langle \vartheta_{t+1}, u_{t+1} - u^* \rangle - s(t-\alpha)(t+\alpha-\beta-1)\langle \vartheta_{t+1}, u_{t+1} - u_t \rangle \\
&- s^2(t+\alpha-\beta-1)(t\beta-\alpha)\langle \vartheta_{t+1}, \sigma_{t+1} + \nabla g(\omega_t) \rangle \\
&\overset{(43)}{\leq} s\big[\eta\beta + t(-\gamma+2) + (\alpha-\beta)(-\gamma+1) - 1\big](f(u_{t+1}) - f(u^*)) \\
&- \big[\eta(t-\alpha) + \frac{\gamma\eta}{2} + \frac{\eta^2}{2}\big]\|u_{t+1} - u_t\|^2 \\
&- \frac{s^2(t+\alpha-\beta-1)}{2}\Big[t(2\beta+1) - \alpha - \beta - 1 - L_{\nabla g}\lambda(\beta+1)\Big(t - 1 + \frac{\eta(t\beta-\alpha)}{t+\alpha-\beta-1}\Big)\Big] \\
&\times \|\sigma_{t+1} + \nabla g(\omega_t)\|^2 \\
&+ s[(t+\alpha-\beta-1)(t-1) + \eta(t\beta-\alpha)]\varepsilon_{t+1} + \eta s(\beta+1)(t-1)\langle \vartheta_{t+1}, u_{t+1} - u_t \rangle \\
&- s^2(t+\alpha-\beta-1)(t\beta-\alpha)\langle \vartheta_{t+1}, \sigma_{t+1} + \nabla g(\omega_t) \rangle. \qquad (124)
\end{aligned}$$

□

Theorem 13. *Let Assumption* (1) *hold,* $\alpha \geq 3$ *and* $2\beta + 1 - L_{\nabla g}\lambda(\beta+1) > 0$. *Let the sequences* $\{(u_t, \omega_t, \sigma_t)\}_{t\geq 1}$ *be induced by I-IFBASC. Suppose the error sequence* $\{(\vartheta_t, \varepsilon_t)\}_{t\geq 1}$ *in I-IFBASC satisfies*

$$\|\vartheta_t\| = \mathcal{O}\Big(\frac{1}{t^q}\Big), \qquad \varepsilon_t = \mathcal{O}\Big(\frac{1}{t^{q+1}}\Big),$$

with $q > 2$ *as* $t \to +\infty$. *Then*

(i). $f(u_t) - \min f = \mathcal{O}\Big(\frac{1}{t^2}\Big)$, *as* $t \to +\infty$.

(ii). $\sum_{t=1}^{+\infty}(t - t'_{\alpha,\beta})^2 \mathrm{dist}^2(0, \partial_{\varepsilon_{t+1}} f(u_{t+1})) < +\infty$, $\min_{1\leq i\leq t} \mathrm{dist}^2(0, \partial_{\varepsilon_{t+1}} f(u_{t+1})) = o\Big(\frac{1}{t^3}\Big)$.

Furthermore, when $\alpha > 3$

(iii). $\sum_{t=T'_{\alpha,\beta}+1}^{+\infty}(t - T'_{\alpha,\beta})(f(u_{t+1}) - f(u^*)) < +\infty$, $\sum_{t=2T'_{\alpha,\beta}+1}^{+\infty}(t - 2T'_{\alpha,\beta} + 1)\|u_{t+1} - u_t\|^2 < +\infty$.

(iv). $\lim_{t\to+\infty} u_t = u^* \in S$.

(v). $f(u_t) - \min f = o\Big(\frac{1}{t^2}\Big)$, $\|u_{t+1} - u_t\| = o\Big(\frac{1}{t}\Big)$.

Proof. (i). Set $\gamma = \alpha - 1$, then $\eta = \alpha - \gamma - 1 = 0$. From Lemma (12), we get

$$\begin{aligned}
&\mathcal{E}^{\vartheta}_{t+1}(\alpha-1) - \mathcal{E}^{\vartheta}_{t}(\alpha-1)\\
&\le s\big[t(3-\alpha) + (\alpha-\beta)(2-\alpha) - 1\big](f(u_{t+1}) - f(u^*))\\
&- \frac{s^2(t+\alpha-\beta-1)}{2}[t(2\beta+1) - \alpha - \beta - 1 - L_{\nabla g}\lambda(\beta+1)] \times \|\sigma_{t+1} + \nabla g(\omega_t)\|^2\\
&+ s[(t+\alpha-\beta-1)(t-1)]\varepsilon_{t+1} - s^2(t+\alpha-\beta-1)(t\beta-\alpha)\langle \vartheta_{t+1}, \sigma_{t+1} + \nabla g(\omega_t)\rangle. \quad (125)
\end{aligned}$$

Let $t'_{\alpha,\beta} := \max\{t_{\alpha,\beta}, \lfloor \frac{\alpha}{\beta} \rfloor + 1\}$ then $t\beta - \alpha \ge \beta(t - t'_{\alpha,\beta})$. Using Equation (46) and (47) in Equation (125), we get

$$\begin{aligned}
\mathcal{E}^{\vartheta}_{t+1}(\alpha-1) - \mathcal{E}^{\vartheta}_{t}(\alpha-1) &\le -s(\alpha-3)(t-t'_{\alpha,\beta})(f(u_{t+1}) - f(u^*))\\
&- \frac{s^2(t-t'_{\alpha,\beta})^2}{2}[(2\beta+1) - L_{\nabla g}\lambda(\beta+1)]\|\sigma_{t+1} + \nabla g(\omega_t)\|^2\\
&+ s[(t+\alpha-\beta-1)(t-1)]\varepsilon_{t+1} - s^2\beta(t-t'_{\alpha,\beta})^2\langle \vartheta_{t+1}, \sigma_{t+1} + \nabla g(\omega_t)\rangle.
\end{aligned} \tag{126}$$

Putting $s = \frac{2\beta}{\delta}$, $u = \vartheta_{t+1}$ and $v = \sigma_{t+1} + \nabla g(\omega_t)$ in Equation (14), we obtain

$$\langle \vartheta_{t+1}, \sigma_{t+1} + \nabla g(\omega_t)\rangle \ge -\frac{\beta}{\delta}\|\vartheta_{t+1}\|^2 - \frac{\delta}{4\beta}\|\sigma_{t+1} + \nabla g(\omega_t)\|^2, \tag{127}$$

where $\delta = (2\beta+1) - L_{\nabla g}\lambda(\beta+1) > 0$. Applying the result (127) in (126), we get

$$\begin{aligned}
\mathcal{E}^{\vartheta}_{t+1}(\alpha-1) - \mathcal{E}^{\vartheta}_{t}(\alpha-1) &\le -s(\alpha-3)(t-t'_{\alpha,\beta})(f(u_{t+1}) - f(u^*))\\
&- \frac{s^2(t-t'_{\alpha,\beta})^2}{2}[(2\beta+1) - L_{\nabla g}\lambda(\beta+1)]\|\sigma_{t+1} + \nabla g(\omega_t)\|^2\\
&+ s[(t+\alpha-\beta-1)(t-1)]\varepsilon_{t+1} + \frac{s^2\beta^2(t-t'_{\alpha,\beta})^2}{\delta}\|\vartheta_{t+1}\|^2\\
&+ \frac{s^2(t-t'_{\alpha,\beta})^2}{4}[(2\beta+1) - L_{\nabla g}\lambda(\beta+1)]\|\sigma_{t+1} + \nabla g(\omega_t)\|^2. \quad (128)
\end{aligned}$$

Also,

$$\begin{aligned}
\mathcal{E}^{\vartheta}_{t+1}(\alpha-1) - \mathcal{E}^{\vartheta}_{t}(\alpha-1) &\le -s(\alpha-3)(t-t'_{\alpha,\beta})(f(u_{t+1}) - f(u^*))\\
&- \frac{s^2(t-t'_{\alpha,\beta})^2}{4}[(2\beta+1) - L_{\nabla g}\lambda(\beta+1)]\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 + \Gamma_{t+1}\\
&\le \Gamma_{t+1}, \quad (129)
\end{aligned}$$

for all $t \ge t'_{\alpha,\beta} + 1$, and denote

$$\Gamma_{t+1} := s[(t+\alpha-\beta-1)(t-1)]\varepsilon_{t+1} + \frac{s^2\beta^2(t-t'_{\alpha,\beta})^2}{\delta}\|\vartheta_{t+1}\|^2. \tag{130}$$

Since the two sequences satisfy $\|\vartheta_t\| = \mathcal{O}\Big(\frac{1}{t^q}\Big)$ and $\varepsilon_t = \mathcal{O}\Big(\frac{1}{t^{q+1}}\Big)$ for $q > 2$, we obtain

$$\sum_{t=1}^{+\infty} t\|\vartheta_t\| < +\infty, \qquad \sum_{t=1}^{+\infty} t^2\varepsilon_t < +\infty. \tag{131}$$

From Equation (130) and (131), we get

$$\sum_{t=1}^{+\infty} \Gamma_t < +\infty. \tag{132}$$

Considering the definitions of $\mathcal{E}_t(\alpha-1)$ and $\mathcal{E}_t^{\vartheta}(\alpha-1)$, we derive

$$\begin{aligned}\frac{1}{2}\|\pi_t(\alpha-1)\|^2 &\leq \mathcal{E}_t(\alpha-1)\\ &= \mathcal{E}_t^{\vartheta}(\alpha-1) + \sum_{i=1}^{t} s(i+\alpha-\beta-2)\langle \vartheta_i, \pi_i(\alpha-1)\rangle. \end{aligned}\tag{133}$$

Applying Cauchy-Schwarz inequality and the result (129) in (133), we obtain

$$\begin{aligned}\frac{1}{2}\|\pi_t(\alpha-1)\|^2 \leq \mathcal{E}_t(\alpha-1) &\leq \mathcal{E}_1^{\vartheta}(\alpha-1) + \sum_{i=2}^{t}\Gamma_i + \sum_{i=1}^{t} s(i+\alpha-\beta-2)\|\vartheta_i\|\|\pi_i(\alpha-1)\|\\ &= \mathcal{E}_1(\alpha-1) + \sum_{i=2}^{t}\Gamma_i + \sum_{i=2}^{t} s(i+\alpha-\beta-2)\|\vartheta_i\|\|\pi_i(\alpha-1)\|\\ &\leq \mathcal{E}_1(\alpha-1) + \sum_{i=2}^{+\infty}\Gamma_i + \sum_{i=2}^{t} s(i+\alpha-\beta-2)\|\vartheta_i\|\|\pi_i(\alpha-1)\|. \end{aligned}\tag{134}$$

Let define $E := \mathcal{E}_1(\alpha-1) + \sum_{i=2}^{+\infty}\Gamma_i$ and from Equation (132), we have $E < +\infty$. Thus, Equation (134) becomes

$$\frac{1}{2}\|\pi_t(\alpha-1)\|^2 \leq E + \sum_{i=2}^{t} s(i+\alpha-\beta-2)\|\vartheta_i\|\|\pi_i(\alpha-1)\|. \tag{135}$$

From Equation (131), applying Lemma (6) in (135), we obtain

$$\|\pi_t(\alpha-1)\| \leq \sqrt{2E} + 2s\sum_{t=1}^{+\infty}(t+\alpha-\beta-2)\|\vartheta_t\| < +\infty. \tag{136}$$

Also,

$$\sup_{t\geq 1}\|\pi_t(\alpha-1)\| < +\infty. \tag{137}$$

From Equation (131), (133), (134) and (137), we obtain

$$\sup_{t\geq 1}\mathcal{E}_t(\alpha-1) \leq E + \sup_{t\geq 1}\|\pi_t(\alpha-1)\|\sum_{t=1}^{+\infty} s(t+\alpha-\beta-2)\|\vartheta_t\| < +\infty. \tag{138}$$

Hence, the sequence $\{\mathcal{E}_t(\alpha-1)\}_{t\geq 1}$ is bounded. From Equation (23), we get

$$f(u_t) - \min f = \mathcal{O}\Big(\frac{1}{t^2}\Big), \qquad \text{as } t \to +\infty. \tag{139}$$

(ii). As the sequence $\mathcal{E}_t(\alpha-1) \geq 1$, and the sequence $\{\|\pi_t(\alpha-1)\|\}_{t\geq 1}$ is bounded, we obtain

$$\inf_{t\geq 1}\mathcal{E}_t^{\vartheta}(\alpha-1) \geq -\sup_{t\geq 1}\|\pi_t(\alpha-1)\|\sum_{t=1}^{+\infty} s(t+\alpha-\beta-2)\|\vartheta_t\| > -\infty. \tag{140}$$

Therefore, $\{\mathcal{E}_t^{\vartheta}(\alpha-1)\}_{t\geq 1}$ is bounded from below. Since $2\beta+1-L_{\nabla g}\lambda(\beta+1) > 0$ and from Equation (129), we obtain

$$\mathcal{E}_{t+1}^{\vartheta}(\alpha-1) - \mathcal{E}_t^{\vartheta}(\alpha-1) \leq -\frac{s^2(t-t'_{\alpha,\beta})^2}{4}\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 + \Gamma_{t+1}. \tag{141}$$

From Equation (134), we have $\sum_{t=1}^{+\infty}\Gamma_t < +\infty$, applying Lemma (4) to Equation (141), we obtain $\lim_{t\to+\infty}\mathcal{E}_t^{\vartheta}(\alpha-1)$ exists and

$$\sum_{t=1}^{+\infty}(t-t'_{\alpha,\beta})^2\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 < +\infty. \tag{142}$$

We deduce

$$\begin{aligned}
&\sum_{t=1}^{+\infty}(t-t'_{\alpha,\beta})^2\mathrm{dist}^2(0,\partial_{\varepsilon_{t+1}}f(u_{t+1})) \leq \sum_{t=1}^{+\infty}(t-t'_{\alpha,\beta})^2\|\sigma_{t+1}+\vartheta_{t+1}+\nabla g(u_{t+1})\|^2\\
&\leq 2\sum_{t=1}^{+\infty}(t-t'_{\alpha,\beta})^2\|\sigma_{t+1}+\nabla g(u_{t+1})\|^2 + 2\sum_{t=1}^{+\infty}(t-t'_{\alpha,\beta})^2\|\vartheta_{t+1}\|^2\\
&\overset{(58)}{\leq} 4(1+L_{\nabla g}^2\lambda^2)\sum_{t=1}^{+\infty}(t-t'_{\alpha,\beta})^2\|\sigma_{t+1}+\nabla g(\sigma_{t+1})\|^2 + 2\sum_{t=1}^{+\infty}(t-t'_{\alpha,\beta})^2\|\vartheta_{t+1}\|^2\\
&\overset{(131),(142)}{<} +\infty.
\end{aligned} \tag{143}$$

Therefore, there exists a positive constant C such that [16, 26]

$$0 \leq \lim_{t\to+\infty}\left(t^3 \min_{1\leq i\leq t} \mathrm{dist}^2(0, \partial_{\varepsilon_{t+1}} f(u_{t+1}))\right) \leq C \lim_{t\to+\infty} \sum_{i=\lfloor \frac{t}{2}\rfloor}^{t} t^2 \mathrm{dist}^2(0, \partial_{\varepsilon_{t+1}} f(u_{t+1})) = 0. \tag{144}$$

Hence,

$$\min_{1\leq i\leq t} \mathrm{dist}^2(0, \partial_{\varepsilon_{t+1}} f(u_{t+1})) = o\Big(\frac{1}{t^3}\Big). \tag{145}$$

(iii). Whenever $\alpha > 3$, for $\gamma \in (2, \alpha - 1)$, we have $\eta = \alpha - \gamma - 1 > 0$. Choose $T'_{\alpha,\beta} = \max\big\{T_{\alpha,\beta}, \lfloor \frac{\beta}{\alpha} \rfloor + 1\big\}$, from Equation (65), (66), (67) and the result from Lemma (12) becomes

$$\begin{aligned}
&\mathcal{E}^{\vartheta}_{t+1}(\gamma) - \mathcal{E}^{\vartheta}_{t}(\gamma) \\
&\leq -s(\gamma-2)(t - t'_{\alpha,\beta})(f(u_{t+1}) - f(u^*)) - \big[\eta(t - t'_{\alpha,\beta}) + \frac{\gamma\eta}{2} + \frac{\eta^2}{2}\big]\|u_{t+1} - u_t\|^2 \\
&- \frac{s^2(t - t'_{\alpha,\beta})^2}{2}((2\beta+1) - L_{\nabla g}\lambda(\beta+1))\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 \\
&+ s[(t+\alpha-\beta-1)(t-1) + \eta(t\beta - \alpha)]\varepsilon_{t+1} \\
&+ \eta s(\beta+1)(t-1)\langle \vartheta_{t+1}, u_{t+1} - u_t\rangle - s^2\beta(t - t'_{\alpha,\beta})^2\langle \vartheta_{t+1}, \sigma_{t+1} + \nabla g(\omega_t)\rangle.
\end{aligned} \tag{146}$$

Putting $s = s(\beta+1)$, $u = \vartheta_{t+1}$ and $v = u_{t+1} - u_t$ in Equation (14), we obtain

$$\langle \vartheta_{t+1}, u_{t+1} - u_t\rangle \leq \frac{s(\beta+1)}{2}\|\vartheta_{t+1}\|^2 + \frac{1}{2s(\beta+1)}\|u_{t+1} - u_t\|^2. \tag{147}$$

Applying the results (127) and 147 in Equation (146), we obtain

$$\begin{aligned}
&\mathcal{E}^{\vartheta}_{t+1}(\gamma) - \mathcal{E}^{\vartheta}_{t}(\gamma) \\
&\leq -s(\gamma-2)(t - t'_{\alpha,\beta})(f(u_{t+1}) - f(u^*)) - \big[\eta(t - t'_{\alpha,\beta}) + \frac{\gamma\eta}{2} + \frac{\eta^2}{2}\big]\|u_{t+1} - u_t\|^2 \\
&- \frac{s^2(t - t'_{\alpha,\beta})^2}{2}((2\beta+1) - L_{\nabla g}\lambda(\beta+1))\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 \\
&+ s[(t+\alpha-\beta-1)(t-1) + \eta(t\beta - \alpha)]\varepsilon_{t+1} + \frac{\eta(t-1)}{2}\|u_{t+1} - u_t\|^2 \\
&+ \Big[\frac{\eta s^2(\beta-1)^2(t-1)}{2} + \frac{s^2\beta^2(t - t'_{\alpha,\beta})^2}{\delta}\Big]\|\vartheta_{t+1}\|^2 \\
&+ \frac{s^2(t - t'_{\alpha,\beta})^2}{4}((2\beta+1) - L_{\nabla g}\lambda(\beta+1))\|\sigma_{t+1} + \nabla g(\omega_t)\|^2.
\end{aligned} \tag{148}$$

Also, since $(2\beta+1)-L_{\nabla g}\lambda(\beta+1)>0$, we have

$$\begin{aligned}
&\mathcal{E}^{\vartheta}_{t+1}(\gamma)-\mathcal{E}^{\vartheta}_{t}(\gamma)\\
&\le -s(\gamma-2)(t-t'_{\alpha,\beta})(f(u_{t+1})-f(u^*))\\
&-\Big[\frac{\eta(t-2T'_{\alpha,\beta}+1)}{2}+\frac{\gamma\eta}{2}+\frac{\eta^2}{2}\Big]\|u_{t+1}-u_t\|^2\\
&-\frac{s^2(t-t'_{\alpha,\beta})^2}{4}((2\beta+1)-L_{\nabla g}\lambda(\beta+1))\|\sigma_{t+1}+\nabla g(\omega_t)\|^2+\Gamma'_{t+1}\\
&\le \Gamma'_{t+1}, \qquad (149)
\end{aligned}$$

for $t\ge T'_{\alpha,\beta}+1$ and where

$$\begin{aligned}
\Gamma'_{t+1}:=\Big[&\frac{\eta s^2(\beta-1)^2(t-1)}{2}+\frac{s^2\beta^2(t-t'_{\alpha,\beta})^2}{(2\beta+1)-L_{\nabla g}\lambda(\beta+1)}\Big]\|\vartheta_{t+1}\|^2\\
&+s[(t+\alpha-\beta-1)(t-1)+\eta(t\beta-\alpha)]\varepsilon_{t+1}. \qquad (150)
\end{aligned}$$

From Equation (131), we get

$$\sum_{t=1}^{+\infty}\Gamma'_{t+1}<+\infty. \qquad (151)$$

Applying Cauchy-Schwarz inequality and the result (149) in Equation (133), we have

$$\begin{aligned}
\frac{1}{2}\|\pi_t(\gamma)\|^2 &\le \mathcal{E}^{\vartheta}_1(\gamma)+\sum_{i=2}^{t}\Gamma'_i+\sum_{i=1}^{t}s(i+\alpha-\beta-2)\|\vartheta_i\|\|\pi_i(\gamma)\|\\
&=\mathcal{E}_1(\gamma)+\sum_{i=2}^{t}\Gamma'_i+\sum_{i=2}^{t}s(i+\alpha-\beta-2)\|\vartheta_i\|\|\pi_t(\gamma)\|\\
&\le \mathcal{E}_1(\gamma)+\sum_{i=2}^{+\infty}\Gamma'_i+\sum_{i=2}^{t}s(i+\alpha-\beta-2)\|\vartheta_i\|\|\pi_i(\gamma)\|. \qquad (152)
\end{aligned}$$

Let define $E':=\mathcal{E}_1(\gamma)+\sum_{i=2}^{+\infty}\Gamma'_i$ and from Equation (151), we have $E'<+\infty$. Thus, Equation (152) becomes

$$\frac{1}{2}\|\pi_t(\gamma)\|^2\le E'+\sum_{i=2}^{t}s(i+\alpha-\beta-2)\|\vartheta_i\|\|\pi_i(\gamma)\|. \qquad (153)$$

From Equation (151), applying Lemma (6) in (153), we obtain

$$\|\pi_t(\gamma)\|\le\sqrt{2E'}+2s\sum_{t=1}^{+\infty}(t+\alpha-\beta-2)\|\vartheta_t\|<+\infty. \qquad (154)$$

Also,

$$\sup_{t\geq 1}\|\pi_t(\gamma)\| < +\infty. \tag{155}$$

From Equation (131), (133), (134) and (155), we obtain

$$\sup_{t\geq 1}\mathcal{E}_t(\gamma) \leq E' + \sup_{t\geq 1}\|\pi_t(\gamma)\|\sum_{t=1}^{+\infty} s(t+\alpha-\beta-2)\|\vartheta_t\| < +\infty. \tag{156}$$

Hence, the sequence $\{\mathcal{E}_t(\gamma)\}_{t\geq 1}$ is bounded for $\gamma \in (2, \alpha-1)$. Since $\mathcal{E}_t(\gamma) \geq 1$, the boundedness of $\{\|\pi_t(\gamma)\|\}_{t\geq 1}$ implies,

$$\inf_{t\geq 1}\mathcal{E}_t^{\vartheta}(\gamma) \geq -\sup_{t\geq 1}\|\pi_t(\gamma)\|\sum_{t=1}^{+\infty} s(t+\alpha-\beta-2)\|\vartheta_t\| > -\infty. \tag{157}$$

Therefore, $\{\mathcal{E}_t^{\vartheta}(\gamma)\}_{t\geq 1}$ is bounded from below for $\gamma \in (2, \alpha-1)$. From Equation (149), we obtain

$$\mathcal{E}_{t+1}^{\vartheta}(\gamma) - \mathcal{E}_t^{\vartheta}(\gamma) \leq -s(\gamma-2)(t-t'_{\alpha,\beta})(f(u_{t+1})-f(u^*)) + \Gamma_{t+1}. \tag{158}$$

Applying Lemma (4) to Equation (158), we obtain $\lim_{t\to+\infty}\mathcal{E}_t^{\vartheta}(\gamma)$ exists and

$$\sum_{t=T'_{\alpha,\beta}+1}^{+\infty}(t-t'_{\alpha,\beta})(f(u_{t+1})-f(u^*)) < +\infty. \tag{159}$$

Also, from Equation (149), we obtain

$$\mathcal{E}_{t+1}^{\vartheta}(\gamma) - \mathcal{E}_t^{\vartheta}(\gamma) \leq -\frac{\eta(t-2T'_{\alpha,\beta}+1)}{2}\|u_{t+1}-u_t\|^2 + \Gamma_{t+1} \tag{160}$$

Applying Lemma (4) to Equation (160), we obtain

$$\sum_{t=2T'_{\alpha,\beta}+1}^{+\infty}(t-2T'_{\alpha,\beta}+1)\|u_{t+1}-u_t\|^2 < +\infty. \tag{161}$$

(iv). We observe that the sequence $\{\mathcal{E}_t^{\vartheta}(\gamma)\}_{t\geq 1}$ is bounded for any $\gamma \in (2, \alpha-1)$. Observe that Equation (114) and (23) imply that $\{u_t\}_{t\geq 1}$ is bounded. Since $\{\mathcal{E}_t^{\vartheta}(\alpha-1)\}_{t\geq 1}$ is bounded, from Equation (24), we imply that the sequence

$$\Big\{(\alpha-1)(u_{t+1}-u^*) + (t-\alpha)(u_{t+1}-u_t) + s(\beta t-\alpha)(\sigma_{t+1}+\nabla g(\omega_t))\Big\}_{t\geq 1} \tag{162}$$

is bounded. From Equation (142), the sequence $\big\{t(\sigma_{t+1} + \nabla g(\omega_t))\big\}_{t\geq 1}$ is bounded, along with $\{u_t\}_{t\geq 1}$ is bounded, we get

$$\|u_{t+1} - u_t\| = \mathcal{O}\Big(\frac{1}{t}\Big), \tag{163}$$

and each cluster point of $\{u_t\}_{t\geq 1}$ belongs to S. Furthermore, $\sigma_{t+1} + \vartheta_{t+1} \in \partial_{\varepsilon_{t+1}} h(u_{t+1})$ and $0 \in \partial f(u^*) = \nabla g(u^*) + \partial h(u^*)$ implies that $-\nabla g(u^*) \in \partial h(u^*)$, from Equation (113) (when $\varepsilon = 0$) we have

$$\langle \nabla g(u^*), y - u^* \rangle \geq h(u^*) - h(y), \qquad \forall y \in \mathbb{R}^d. \tag{164}$$

Using Equation (113) and (164)

$$\begin{aligned}
&\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(u^*)\rangle \\
&= \langle u_{t+1} - u^*, \sigma_{t+1} + \vartheta_{t+1}\rangle + \langle u_{t+1} - u^*, \nabla g(u^*)\rangle - \langle u_{t+1} - u^*, \vartheta_{t+1}\rangle \\
&\geq \langle u_{t+1} - u^*, \sigma_{t+1} + \vartheta_{t+1}\rangle + h(u^*) - h(u_{t+1}) - \langle u_{t+1} - u^*, \vartheta_{t+1}\rangle \\
&\geq h(u_{t+1}) - h(u^*) - \varepsilon_{t+1} + h(u^*) - h(u_{t+1}) - \langle u_{t+1} - u^*, \vartheta_{t+1}\rangle \\
&\geq -\varepsilon_{t+1} - \|u_{t+1} - u^*\|\|\vartheta_{t+1}\|.
\end{aligned} \tag{165}$$

Combining results (92) and (94), and utilizing Equation (165) we get

$$\begin{aligned}
&\langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(\omega_t)\rangle \\
&= \langle u_{t+1} - u^*, \sigma_{t+1} + \nabla g(u^*)\rangle + \langle u_{t+1} - u^*, \nabla g(\omega_t) - \nabla g(u^*)\rangle \\
&\geq -\varepsilon_{t+1} - \|u_{t+1} - u^*\|\|\vartheta_{t+1}\| - \frac{L_{\nabla g}}{4}\|u_{t+1} - \omega_t\|^2.
\end{aligned} \tag{166}$$

Putting Equation (166) in (90), and from stage 5 of Algorithm 2 we obtain

$$\Delta_{t+1} - \frac{t-1-\alpha}{t-1}\Delta_t \leq e'_t, \tag{167}$$

where

$$\begin{aligned}
e'_t = {}& \frac{L_{\nabla g}\lambda^2\beta s}{4}\|\sigma_{t+1} + \nabla g(\omega_t)\|^2 + \frac{1+\beta s}{2}\|u_{t+1} - u_t\|^2 + \beta s\varepsilon_{t+1} + \beta s\|u_{t+1} - u^*\|\|\vartheta_{t+1}\| \\
&+ \|u_t - u_{t-1}\|^2 + \frac{\beta s}{2}\|\sigma_t + \nabla g(\sigma_{t-1})\|^2 + s(\beta - 1)\|u_t - u^*\|\|\sigma_t + \nabla g(\sigma_{t-1})\| \\
&\overset{(97)}{=} e_t + \beta s\|u_{t+1} - u^*\|\|\vartheta_{t+1}\| + \beta s\varepsilon_{t+1}.
\end{aligned} \tag{168}$$

Since the two sequences satisfy $\|\vartheta_t\| = \mathcal{O}\Big(\frac{1}{t^q}\Big)$ and $\varepsilon_t = \mathcal{O}\Big(\frac{1}{t^{q+1}}\Big)$ for $q > 2$, combining with the boundedness of $\{u_t\}_{t\geq 1}$ and Equation (100), we obtain

$$\sum_{t=1}^{+\infty} t e'_t < +\infty. \tag{169}$$

We can further derive

$$[\Delta_{t+1}]_+ \leq \frac{t-1-\alpha}{t-1}[\Delta_t]_+ + e'_t, \tag{170}$$

applying Lemma (9), we obtain

$$\sum_{t=1}^{+\infty}[\Delta_t]_+ < +\infty. \tag{171}$$

From Equation (171) and (104), Equation (103) becomes

$$\sum_{t=1}^{+\infty}[h_{t+1} - h_t]_+ < +\infty. \tag{172}$$

Applying Lemma (4),

$$h_{t+1} \leq h_t + [h_{t+1} - h_t]_+, \tag{173}$$

we conclude $\lim_{t\to+\infty} h_t = \frac{1}{2}\lim_{t\to+\infty}\|u_t - u^*\|^2$ exists. Since each sequential cluster point of $\{u_t\}_{t\geq 1}$ belongs to S, by Lemma (5), we conclude that $\lim_{t\to+\infty} u_t = u^* \in S$.
(v). From Equation (114) and (129), we imply that

$$\begin{aligned}\mathcal{E}_{t+1}(\alpha-1) - \mathcal{E}_t(\alpha-1) &\leq \Gamma_{t+1} + s(t+\alpha-\beta-1)\langle \vartheta_{t+1}, \pi_{t+1}(\alpha-1)\rangle \\ &\leq \Gamma_{t+1} + s(t+\alpha-\beta-1)\|\vartheta_{t+1}\|\|\pi_{t+1}(\alpha-1)\|.\end{aligned} \tag{174}$$

Since $\|\vartheta_{t+1}\| = \mathcal{O}\Big(\frac{1}{t^q}\Big)$ and $\varepsilon_{t+1} = \mathcal{O}\Big(\frac{1}{t^{q+1}}\Big)$for $q > 2$, we have

$$\begin{aligned}&\sum_{t=1}^{+\infty}(\Gamma_{t+1} + s(t+\alpha-\beta-1)\|\vartheta_{t+1}\|\|\pi_t(\alpha-1)\|) \\ &\leq \sum_{t=1}^{+\infty}\Gamma_{t+1} + \sup_{t\geq 1}\|\pi_{t+1}(\alpha-1)\|\sum_{t=1}^{+\infty}s(t+\alpha-\beta-1)\|\vartheta_{t+1}\| < +\infty.\end{aligned} \tag{175}$$

Using Equation (175), applying Lemma (4) in Equation (174), we obtain $\lim_{t\to+\infty}\mathcal{E}_t(\alpha-1)$ exists. Analogously, using (110), (111) and (112), we conclude that

$$f(u_t) - \min f = o\Big(\frac{1}{t^2}\Big), \qquad \|u_{t+1} - u_t\| = o\Big(\frac{1}{t}\Big). \tag{176}$$

□

7 Numerical experiments

In this section, we provide three numerical experiments designed to validate the effectiveness of IFBASC algorithm and its inexact version I-IFBASC. For each algorithm, the parameters are chosen to meet the assumptions necessary for the theoretical convergence analysis. All experiments are conducted using MATLAB R2025a on a desktop computer with an Intel Core i7 processor running at 3.40 GHz and 10 GB of RAM.

7.1 Lasso Problem

The following represent the LASSO problem:

$$\min_{y\in\mathbb{R}^d} f(y) = \frac{1}{2}\|My - x\|^2 + \rho\|y\|_1. \tag{177}$$

Here, $M \in \mathbb{R}^{m\times d}$, $x \in \mathbb{R}^m$, and $m \leq d$. We set $\rho = 1$, $m = 300$, and $d = 800$ for this experiment. The matrix M are taken from a standard Normal distribution. We take $x = My$, where the original vector y is chosen independently from a standard Normal distribution.

Equation (177) is modified as follows:

$$g(y) = \frac{1}{2}\|My - x\|^2, \qquad h(y) = \rho\|y\|_1.$$

Consequently, the gradient vector of function g is defined by $\nabla g(y) = M^T(My - x)$, and the proximal operator related with the function h is in the form

$$\text{Prox}_{\gamma h}(y) = \max\{y - \gamma\rho, 0\} - \max\{-y - \gamma\rho, 0\}.$$

For the IFBASC, the parameters are chosen as

$$\alpha = 6, \qquad \beta = 1.15, \qquad s = \frac{2\beta + 1}{(\beta+1)^2\|M^TM\|}, \qquad \gamma = \frac{2\beta+1}{(\beta+1)\|M^TM\|}.$$

In case of FISTA, we set

$$\gamma = \frac{1}{\|M^TM\|},$$

as commonly adopted in [5, 15].

We compare our Algorithm (1) (IFBASC) with FISTA [5, 15] in the first 1000 iterations in Fig. (1). The comparison of IFBASC and FISTA in terms of iterations and CPU time in seconds under various stopping conditions is shown in Table (1). Our theoretical convergence results are validated by Fig. (1) and Table (1). We have discussed a subgradient correction method, called AFBA proposed in [21] in the Sect. (**??**). Table (2) is obtained by taking $\beta = 0.5$ for IFBASC and $\rho = 0.7$ for AFBA and $\alpha = 20$ for all algorithms. Table (2) shows that our Algorithm (1) gives better result as compare to AFBA and also FISTA. Here, α is chosen in the range $[12, 25]$.

Table 1: Comparison of FISTA and IFBASC: Iterations and CPU Time

Criterion	**FISTA**		**IFBASC**	
2-5	**Iterations**	**CPU (sec)**	**Iterations**	**CPU (sec)**
$\|u_t - u^*\| \leq 10^{-12}$	915	1.4706	735	1.6958
$f(u_t) - \min f \leq 10^{-8}$	532	1.3957	446	1.4819
$\text{dist}(0, \partial f(u_t)) \leq 10^{-8}$	753	1.3939	627	1.4943
$t^2\big(f(u_t) - \min f\big) \leq 10^{-6}$	871	1.4255	663	1.4663

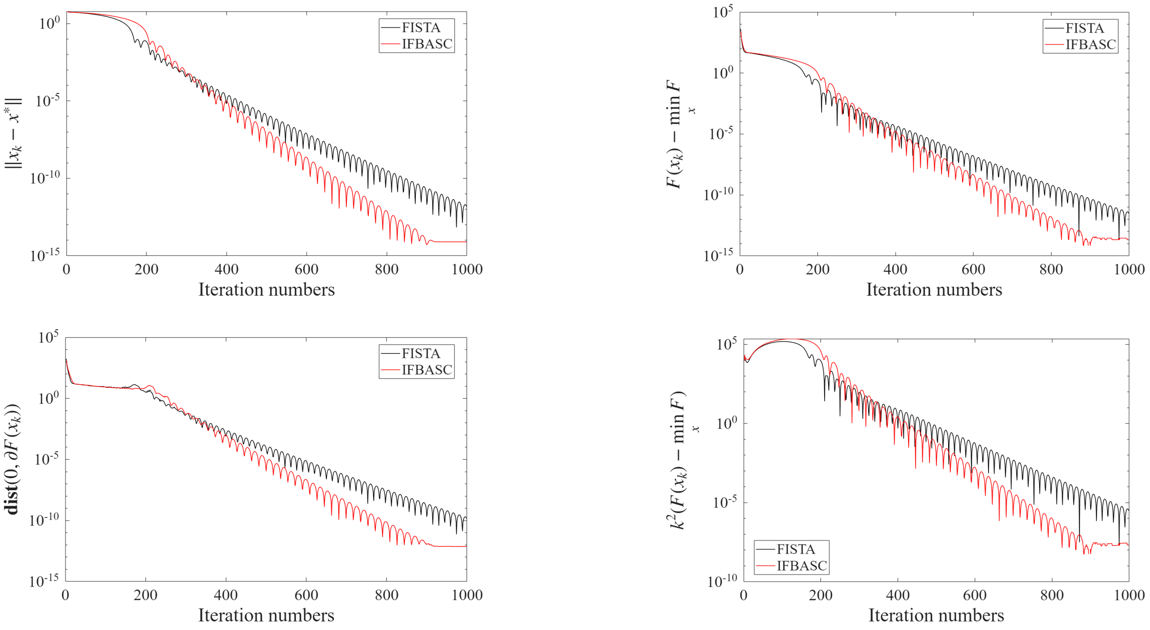

Figure 1: Lasso Problem: Evaluations of $\|u_t - u^*\|$, $\|f(u_t) - f(u^*)\|$, $\text{dist}(0, \partial f(u_t))$ and $t^2\|f(u_t) - f(u^*)\|$ with respect to the number of iterations

Table 2: Performance comparison of FISTA, AFBA, and IFBASC

Criterion	**FISTA**		**AFBA**		**IFBASC**	
	CPU (sec)	**Iter**	**CPU (sec)**	**Iter**	**CPU (sec)**	**Iter**
$\|u_t - u^*\| \leq 10^{-12}$	1.2513	788	1.2327	794	1.3481	706
$f(u_t) - \min f \leq 10^{-8}$	1.2513	552	1.2327	622	1.3481	505
$\text{dist}(0, \partial f(u_t)) \leq 10^{-8}$	1.2513	670	1.2327	708	1.3481	629
$t^2(f(u_t) - \min f) \leq 10^{-5}$	1.2513	685	1.2327	751	1.3481	644

7.2 Image deblurring via wavelet transform

In this subsection, we compare the effectiveness of IFBASC and C-FISTA which is introduced in [10], and FISTA is proposed in [5, 15] to address the following image deblurring via wavelet transform problem:

$$\min_{W} \|M(W) - N\|^2 + \rho\|X(W)\|_1, \tag{178}$$

where $M : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ is a linear operator associated with a spatially uniform point spread function modeling the blur process. The matrix $N \in \mathbb{R}^{m\times n}$ represents the observed image corrupted by blur and noise. The operator $X : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ denotes a three stages Haar wavelet transformation, and $\rho > 0$ is a regularization factor.

First, every pixel value from the original photos utilized in the experiments is normalized to the interval $[0, 1]$. Next, zero-mean white Normal noise via standard deviation $\sigma > 0$ is added after a Normal blur is applied using a 9×9 kernel with standard deviation 4. The algorithmic parameters are chosen: for C-FISTA, $t_0 = 1$ and $\gamma = \frac{1}{2\|M\|^2}$; for FISTA, $\alpha = 4$ and $\gamma = \frac{1}{2\|M\|^2}$; and for IFBASC, $\alpha = 4$, $\beta = 0.1$, $s = \frac{2\beta+1}{(\beta+1)^2\|M\|^2}$, and $\gamma = \frac{2\beta+1}{(\beta+1)\|M\|^2}$.

In case of the *cameraman* image, the addition of Normally distributed noise with a standard deviation of $\sigma = 1\text{e}{-3}$, and the regularization factor in Problem (178) is set to $\rho = 1\text{e}{-4}$. The original image and its blurry observation are shown in Fig. (2). Fig. (3) reports the value of function f and the rebuilt pictures employing C-FISTA, FISTA, and IFBASC after 200 iterations.

For the *woman* image, the addition of Normally distributed noise with a standard deviation of $\sigma = 1\text{e}{-2}$, and the regularization factor in Problem (178) is chosen as $\rho = 1\text{e}{-3}$. The original and blurry pictures are shown in Fig. (4). Fig. (5) reports the value of function f and the rebuilt pictures employing C-FISTA, FISTA, and IFBASC after 200 iterations.

Table (3) reports a comparison in terms of iterations (Iter), CPU time in seconds (Cpu), and the objective function value f for IFBASC, C-FISTA and FISTA under various termination conditions determined by the relative error $\frac{\|u_{t+1}-u_t\|_f}{\|u_t\|_f}$. As illustrated in Figs. (3) and (5), as well as in Table (3), IFBASC exhibits faster performance than both C-FISTA and FISTA.

Table 3: Performance comparison of C-FISTA, FISTA, and IFBASC methods for deblurring

Image	**Termination Criteria**	**Metric**	**C-FISTA**	**FISTA**	**IFBASC**
Cameraman 2-6	$\frac{\|u_{t+1}-u_t\|_f}{\|u_t\|_f} \leq 10^{-4}$	Iter	218	187	180
		Cpu	5.91	5.25	4.83
		f-value	3.15115×10^{-1}	3.15054×10^{-1}	3.15050×10^{-1}
	$\frac{\|u_{t+1}-u_t\|_f}{\|u_t\|_f} \leq 10^{-5}$	Iter	781	525	477
		Cpu	19.73	12.60	11.18
		f-value	3.14914×10^{-1}	3.14914×10^{-1}	3.14915×10^{-1}
Woman 2-6	$\frac{\|u_{t+1}-u_t\|_f}{\|u_t\|_f} \leq 10^{-4}$	Iter	215	176	168
		Cpu	31.59	24.29	25.39
		f-value	2.97494×10^{1}	2.97493×10^{1}	2.97492×10^{1}
	$\frac{\|u_{t+1}-u_t\|_f}{\|u_t\|_f} \leq 10^{-5}$	Iter	729	487	464
		Cpu	102.60	70.72	68.78
		f-value	2.97487×10^{1}	2.97487×10^{1}	2.97487×10^{1}

7.3 Least squares with regularization

Consider the least squares with regularization problem given as:

$$\min_{u \in \mathbb{R}^d} f(u) = \frac{1}{2}\|Mu - x\|^2 + \rho\|Nu\|_1, \tag{179}$$

where M and $N \in \mathbb{R}^{m \times d}$, and $x \in \mathbb{R}^m$. Matrices M and N are produced from standard Normal distribution. We take $u^* \in \mathbb{R}^d$ such that $Nu^* = 0$ and $x = Mu^*$. We provide

Figure 2: Deblurring cameraman

Figure 3: C-FISTA, FISTA and IFBASC for deblurring cameraman

$f(u) = g(u) + h(u)$ with $g(u) = \frac{1}{2}\|Aw - b\|^2$ and $h(u) = \rho\|Bw\|_1$, and we have $\min_u f = 0$. The problem of the proximal operator in the t-th iteration of IFBASC becomes:

$$\min_{u \in \mathbb{R}^d} P_z(u) = \rho\|Nu\|_1 + \frac{1}{2\lambda}\|u - z\|^2, \tag{180}$$

and its dual problem is represented as:

$$\max_{\omega \in \mathbb{R}^m} Q_z(\omega) = -\frac{\lambda}{2}\|N^T\omega\|^2 + \langle Nz, \omega\rangle - \delta_{\rho\|\cdot\|_\infty}(\omega), \tag{181}$$

where $\delta_{\rho\|\cdot\|_\infty}(\cdot)$ denotes the indicator function of the set $\{v \in \mathbb{R}^m : \|v\|_\infty \le \rho\}$, and $z = \omega_t - \lambda\nabla g(\omega_t)$.

Figure 4: Deblurring women

Figure 5: C-FISTA, FISTA and IFBASC for deblurring women

In this case, we can not find the closed solution of the proximal operator of $\|Nu\|_1$, therefore we can not get the exact solution of the Problem (180). Rather, we can only use some optimization strategies to determine an estimated solution. In this section, we will solve Problem (179) using our I-IFBASC (Algorithm (2)). In order to validate the algorithm's performance, we will also examine the convergence effectiveness of the I-IFBASC under various termination criteria and compare it with existing inexact inertial methods.

In this experiments, an inexact solution to problem (180) is obtained by resolving its dual formulation given in (181). In particular, the dual sequence ω^n, corresponding to (181), is generated using the IFBASC method with parameters $\alpha_{\text{dual}} = 5$, $s_{\text{dual}} = s + 0.35$, and $\gamma_{\text{dual}} = \frac{2\beta+1}{\gamma(\beta+1)\|BB^T\|}$. The maximum number of iterations is set to 500. The associated primal sequence u^n is then recovered via

$$u^n = -\gamma N^T \omega^n + z.$$

During the initial experiment, we fix $m = 30$, $n = 200$, and $\rho = 2$. We study the convergence behavior of I-IFBASC under various stopping criteria for the Problem (180). The parameters of I-IFBASC are chosen as $\alpha = 5$, $\beta = 1.5$, $s = \frac{(2\beta+1)\times 0.95}{(\beta+1)^2\|M^T M\|}, \gamma = \frac{(2\beta+1)\times 0.95}{(\beta+1)\|M^T M\|}$. In the first inexact setting, we take $\vartheta_t = 0$ and $\varepsilon_t = \frac{m}{t^q}$ with $q \in \{3.5, 4, 4.5, 5\}$. The termination criterion for the Problem (180) in I-IFBASC is given by

$$P_z(u^n) - Q_z(\omega^n) \leq \varepsilon_{t+1}. \tag{182}$$

We then set $u_{t+1} = u^n$ such that $0 \in \partial_{\varepsilon_{t+1}} h(u_{t+1}) + \nabla g(\omega_t) + \frac{1}{\gamma}(u_{t+1} - \omega_t)$, (see Proposition 2.3 in [38]). This yields an inexact solution to the problem in (180) that satisfies the assumptions of Theorem (13). Figure (6) shows the numerical performance of I-IFBASC over the first 1000 iterations for different choices of ε_t.

In the next inexact setting, the termination conditions for I-IFBASC is defined by $\vartheta_{t+1} \in \partial h(u_{t+1}) + \nabla g(\omega_t) + \frac{1}{\gamma}(u_{t+1} - \omega_t)$, where $\|\vartheta_{t+1}\| = \frac{mn}{t^q}$ with $q \in \{2.5, 3, 3.5, 4\}$. This condition implies that $\vartheta_{t+1} + \sigma_{t+1} \in \partial h(u_{t+1})$.

Figure (7) illustrates the numerical behavior of I-IFBASC over the first 2000 iterations for different choices of the error sequence ϑ_t. As observed from Figs. (6) and (7), whenever the error in solving the Problem (180) satisfies the conditions stated in Theorem (13), the I-IFBASC method maintains a fast rate of convergence. In addition, improving the accuracy of the Problem (180) solution results in enhanced numerical effectiveness of the inexact scheme. These observations are in agreement with the theoretical convergence results.

In the second experiment, we fix $m = 60$, $n = 200$ and $\rho = 2$. We make a comparison our I-IFBASC with the fast forward-backward algorithm (FFBA) [Equation 55, [5]] and inexact FISTA (I-FISTA) [11]. The termination criterion is set in Equation (182) with $\varepsilon_t = \{\frac{m}{t^4}, \frac{m}{t^5}\}$ for all each problems. The parameter configurations for every algorithm are as follows:

- I-FISTA: $\tau = 0.9$, $\pi_t = \frac{m}{t^4} I_n$, $L_{\nabla g} = \|M^T M\|$.
- FFBA: $\alpha = 5$, $g_t = \frac{m}{t^4} I_n$, $s = \frac{0.95}{\|M^T M\|}$.
- I-IFBASC: $\vartheta_t = \frac{m}{t^4} I_n$, $\alpha = 10$, $\beta = 2.5$, $s = \frac{(2\beta+1)\times 0.95}{(\beta+1)^2\|M^T M\|}$, $\gamma = \frac{(2\beta+1)\times 0.95}{(\beta+1)\|M^T M\|}$.

Figures (8) and (9) display the numerical performance of I-IFBASC, I-FISTA, and FFBA. The results show that, for various inexact solutions of the problems, I-IFBASC consistently achieves better performance than both I-FISTA and FFBA. It is also observed that increasing the parameters α can improve the convergence behavior to some degree. However, further enlarging α has only a limited effect on reducing the objective function gap.

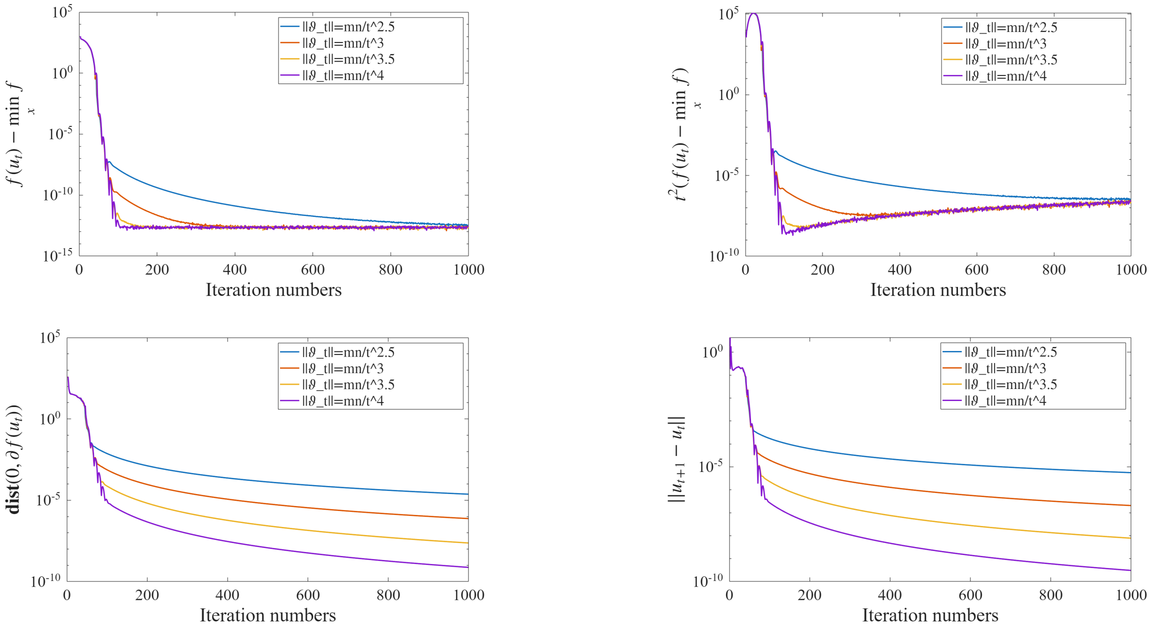

Figure 6: Least squares with regularization: Evaluations of $\|u_t - u^*\|$, $\|f(u_t) - f(u^*)\|$, $\mathrm{dist}(0, \partial f(u_t))$ and $t^2\|f(u_t) - f(u^*)\|$ with respect to the number of iterations, ϑ_t

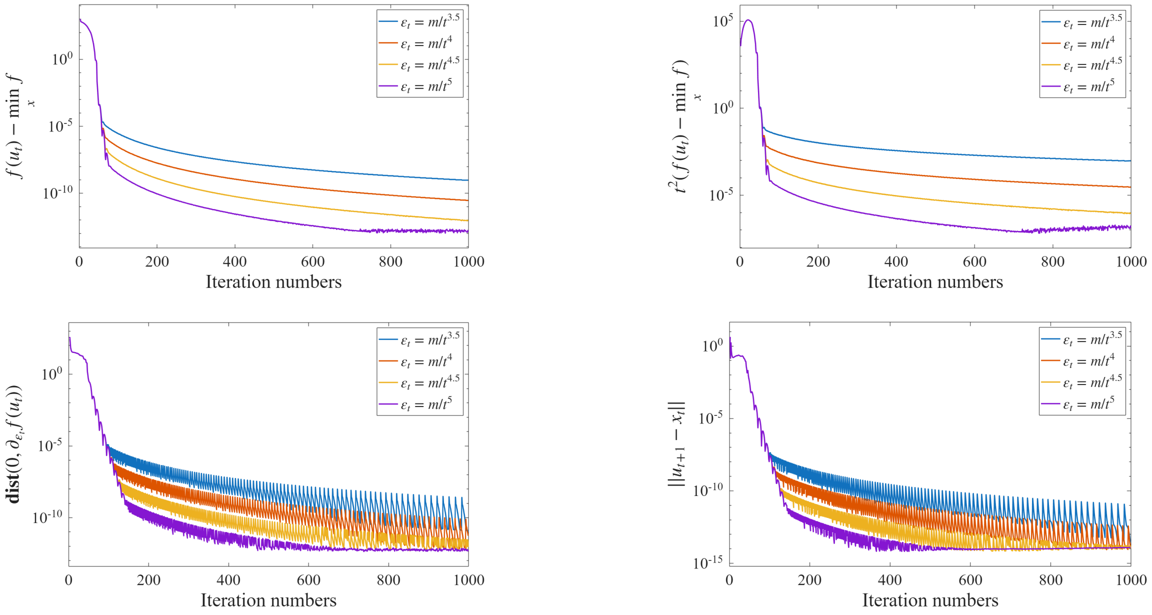

Figure 7: Least squares with regularization: Evaluations of $\|u_t - u^*\|$, $\|f(u_t) - f(u^*)\|$, $\mathrm{dist}(0, \partial f(u_t))$ and $t^2\|f(u_t) - f(u^*)\|$ with respect to the number of iterations, ε_t

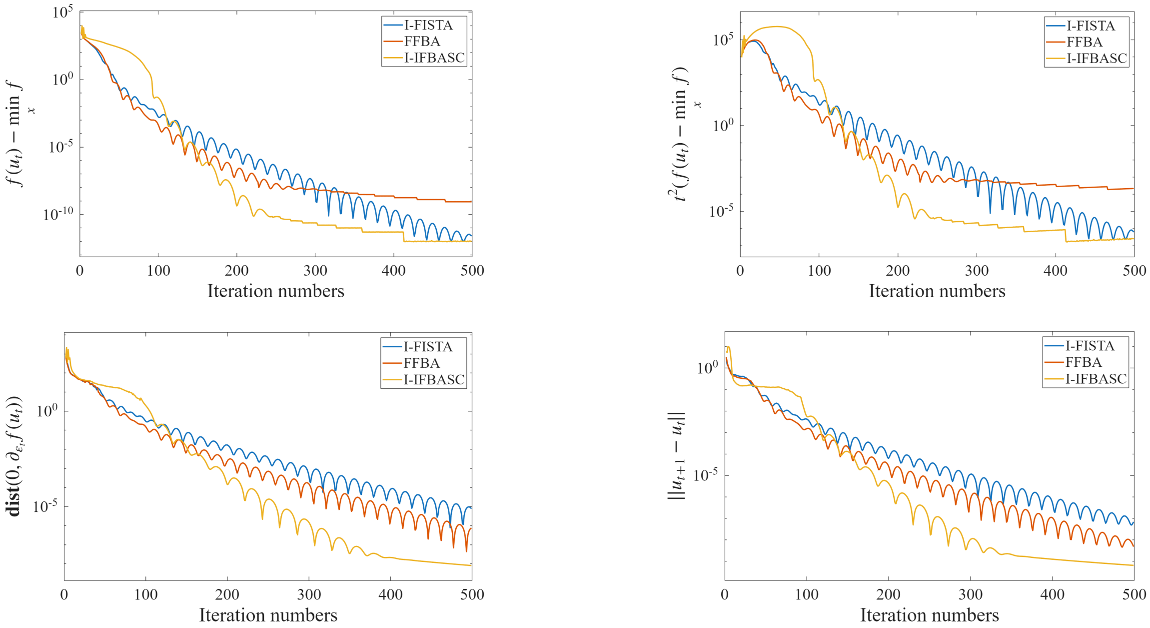

Figure 8: Least squares with regularization: Evaluations of $\|u_t - u^*\|$, $\|f(u_t) - f(u^*)\|$, $\text{dist}(0, \partial f(u_t))$ and $t^2\|f(u_t) - f(u^*)\|$ with respect to the number of iterations, $\varepsilon_t = \frac{m}{t^4}$

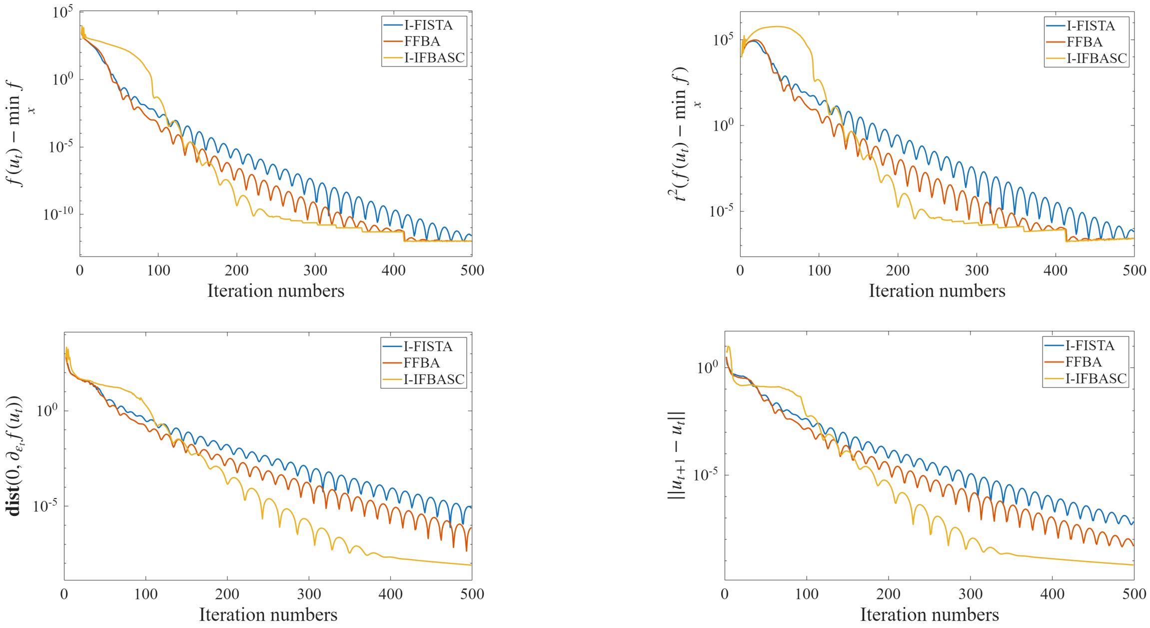

Figure 9: Least squares with regularization: Evaluations of $\|u_t - u^*\|$, $\|f(u_t) - f(u^*)\|$, $\text{dist}(0, \partial f(u_t))$ and $t^2\|f(u_t) - f(u^*)\|$ with respect to the number of iterations, $\varepsilon_t = \frac{m}{t^5}$

8 Conclusion

We conclude from our results that by discretizing the extended ODE given in [34], we obtain an inertial forward-backward algorithm with subgradient correction (Algorithm (1)) and its inexact version (Algorithm (2)) to solve the convex composite optimization problem. We arrive at the rate of convergence of the objective gap being $\mathcal{O}\left(\frac{1}{t^2}\right)$ and $\mathcal{O}\left(\frac{1}{t^3}\right)$ rate of convergence of squared subdifferential norm for $\alpha \geq 3$. When $\alpha > 3$, the rate of objective gap is improved to $o\left(\frac{1}{t^2}\right)$, and also the iterative sequence generated by the algorithm converges to the minimal point. In FISTA [10, 7], the required step size is $\lambda \leq \frac{1}{L_{\nabla g}}$; we improve the step size to $\lambda > \frac{1}{L_{\nabla g}}$ for our algorithms.

References

[1] Vassilis Apidopoulos, Jean-François Aujol, and Charles Dossal. Convergence rate of inertial forward–backward algorithm beyond nesterov's rule. *Mathematical Programming*, 180(1):137–156, 2020.

[2] Hedy Attouch, Radu Ioan Boţ, and Ernö Robert Csetnek. Fast optimization via inertial dynamics with closed-loop damping. *Journal of the European Mathematical Society*, 25(5):1985–2056, 2022.

[3] Hedy Attouch, Zaki Chbani, Jalal Fadili, and Hassan Riahi. First-order optimization algorithms via inertial systems with hessian driven damping. *Mathematical Programming*, 193(1):113–155, 2022.

[4] Hedy Attouch, Zaki Chbani, Jalal Fadili, and Hassan Riahi. Convergence of iterates for first-order optimization algorithms with inertia and hessian driven damping. *Optimization*, 72(5):1199–1238, 2023.

[5] Hedy Attouch, Zaki Chbani, Juan Peypouquet, and Patrick Redont. Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Mathematical Programming*, 168(1):123–175, 2018.

[6] Hedy Attouch and Juan Peypouquet. The rate of convergence of nesterov's accelerated forward-backward method is actually faster than 1/k2. *SIAM Journal on Optimization*, 26(3):1824–1834, 2016.

[7] Hedy Attouch, Juan Peypouquet, and Patrick Redont. Fast convex optimization via inertial dynamics with hessian driven damping. *Journal of Differential Equations*, 261(10):5734–5783, 2016.

[8] J-F Aujol, Ch Dossal, and Aude Rondepierre. Fista is an automatic geometrically optimized algorithm for strongly convex functions. *Mathematical Programming*, 204(1):449–491, 2024.

[9] HH Bauschke and PL Combettes. Convex analysis and monotone operator theory in hilbert spaces. *Springer, Cham*, 2016.

[10] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[11] Max LN Gonçalves Bello-Cruz, Yunier and Nathan Krislock. On fista with a relative error rule. *Computational Optimization and Applications*, 84(2):295–318, 2023.

[12] Radu Ioan Boţ, Ernö Robert Csetnek, and Dang-Khoa Nguyen. Fast optimistic gradient descent ascent (ogda) method in continuous and discrete time. *Foundations of Computational Mathematics*, 25(1):163–222, 2025.

[13] Stephen Boyd and Lieven Vandenberghe. *Convex optimization.* Cambridge university press, New York, 2004.

[14] Antonin Chambolle, Ronald A De Vore, Nam-Yong Lee, and Bradley J Lucier. Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Transactions on image processing*, 7(3):319–335, 1998.

[15] Antonin Chambolle and Ch Dossal. On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm". *Journal of Optimization theory and Applications*, 166(3):968–982, 2015.

[16] Shuo Chen, Bin Shi, and Ya-xiang Yuan. Gradient norm minimization of nesterov acceleration: $o(1/k^3)$. *arXiv preprint arXiv:2209.08862*, 2022.

[17] Patrick L Combettes and Jean-Christophe Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, New York, 2011.

[18] David L Donoho, Michael Elad, and Vladimir N Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on information theory*, 52(1):6–18, 2005.

[19] Mário AT Figueiredo and Robert D Nowak. An em algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 12(8):906–916, 2003.

[20] Osman Güler. New proximal point algorithms for convex minimization. *SIAM Journal on Optimization*, 2(4):649–664, 1992.

[21] Xin He and Yaping Fang. Accelerated forward-backward algorithms with subgradient corrections. *Computational Optimization and Applications*, pages 1–36, 2025.

[22] Xin He, Rong Hu, and Ya-Ping Fang. Inertial accelerated primal-dual methods for linear equality constrained convex optimization problems. *Numerical Algorithms*, 90(4):1669–1690, 2022.

[23] Xin He, Rong Hu, and Yaping Fang. A second order primal–dual dynamical system for a convex–concave bilinear saddle point problem. *Applied Mathematics & Optimization*, 89(2):30, 2024.

[24] Xin He, Nan-Jing Huang, and Ya-Ping Fang. Non-ergodic convergence rate of an inertial accelerated primal–dual algorithm for saddle point problems. *Communications in Nonlinear Science and Numerical Simulation*, 140:108289, 2025.

[25] Michael I. Jordan. Dynamical, symplectic and stochastic perspectives on gradient-based optimization. In *Proceedings of the International Congress of Mathematicians 2018*, pages 523–549, Rio de Janeiro, 2018. World Scientific.

[26] Bowen Li, Bin Shi, and Ya-xiang Yuan. Proximal subgradient norm minimization of ista and fista. *Applied and Computational Harmonic Analysis*, page 101848, 2025.

[27] Hao Luo and Long Chen. From differential equation solvers to accelerated first-order methods for convex optimization. *Mathematical Programming*, 195(1):735–781, 2022.

[28] Yurii Nesterov. A method for solving the convex programming problem with convergence rate o (1/k2). In *Dokl akad nauk Sssr*, volume 269, page 543.

[29] Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science and Business Media, Boston, MA, 2013.

[30] Yurii Nesterov et al. *Lectures on convex optimization*, volume 137. Springer, Cham, Switzerland, 2018.

[31] Daniel P Palomar and Yonina C Eldar. *Convex optimization in signal processing and communications*. Cambridge university press, New York, 2010.

[32] R Tyrrell Rockafellar. *Convex analysis*. Princeton university press, Princeton, 1970.

[33] R Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. *SIAM journal on control and optimization*, 14(5):877–898, 1976.

[34] Bin Shi, Simon S Du, Michael I Jordan, and Weijie J Su. Understanding the acceleration phenomenon via high-resolution differential equations. *Mathematical Programming*, 195(1):79–148, 2022.

[35] Bin Shi, Simon S Du, Weijie Su, and Michael I Jordan. Acceleration via symplectic discretization of high-resolution differential equations. *Advances in Neural Information Processing Systems*, 32, 2019.

[36] Suvrit Sra, Sebastian Nowozin, and Stephen J Wright. *Optimization for machine learning*. MIT press, Cambridge, MA, 2011.

[37] Weijie Su, Stephen Boyd, and Emmanuel J Candes. A differential equation for modeling nesterov's accelerated gradient method: Theory and insights. *Journal of Machine Learning Research*, 17(153):1–43, 2016.

[38] Silvia Villa, Saverio Salzo, Luca Baldassarre, and Alessandro Verri. Accelerated and inexact forward-backward algorithms. *SIAM Journal on Optimization*, 23(3):1607–1633, 2013.

[39] Lixia Wang and Hao Luo. A unified high-resolution ode framework for first-order methods. *arXiv preprint arXiv:2603.07075*, 2026.

[40] Tanxing Wang, Xingju Cai, and Ting Li. On the o (1/k 2) convergence rate analysis for general inertial gradient type method. *Journal of Global Optimization*, pages 1–23, 2025.

[41] Andre Wibisono, Ashia C Wilson, and Michael I Jordan. A variational perspective on accelerated methods in optimization. *proceedings of the National Academy of Sciences*, 113(47):E7351–E7358, 2016.

[42] Ashia C Wilson, Ben Recht, and Michael I Jordan. A lyapunov analysis of accelerated methods in optimization. *Journal of Machine Learning Research*, 22(113):1–34, 2021.