Convergence Analysis of an Inertial Dynamical System with Hessian-Driven Damping under θ-Parametrized Implicit–Explicit Discretization*

Milan Barik[†] Suvendu Ranjan Pattanaik[‡]

Abstract

In this paper, we consider an unconstrained composite convex optimization problem. We propose an inertial forward–backward algorithm derived from an implicit–explicit discretization of a second-order dynamical system with Hessian-driven damping. For $\alpha \geq 3$, we establish an $\mathcal{O}(1/d^2)$ convergence rate for the objective value gap. Furthermore, when $\alpha > 3$, we prove that the iterative sequence generated by the proposed method converges to a minimizer, and that the objective residual admits an improved rate of $o(1/d^2)$. Numerical experiments are provided to illustrate the effectiveness of the proposed approach.

Keywords. *Second-order differential dynamical system, Hessian-driven damping, Accelerated forward-backward method, Convergence analysis*

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

1 Introduction

We consider the following composite convex optimization problem:

$$\min_{u\in\mathbb{R}^n} \Phi(u) = \phi(u) + \psi(u), \tag{1}$$

where $\phi : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable convex function, and $\psi : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is a proper, lower semicontinuous, and convex function. A standard approach for solving problem 1 is the proximal gradient method (PGM). The iterates generated by PGM follow the update rule

$$\zeta_{d+1} = \mathrm{Prox}_{\lambda\psi}\big(\zeta_d - \lambda\nabla\phi(\zeta_d)\big), \tag{2}$$

*Submitted to the editors DATE.

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

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

where $\lambda > 0$ is the step size and Prox denotes the proximal operator is introduced in Section 2. If $\nabla\phi$ is L_ϕ-Lipschitz continuous, ϕ is convex, and ψ is convex on $\mathbb{R}^n$, then the sequence $\{\zeta_d\}_{d\geq 1}$ generated by 2 satisfies the convergence rate [12]

$$\Phi(\zeta_d) - \min_{\zeta} \Phi(\zeta) = \mathcal{O}\left(\frac{1}{d}\right).$$

Numerous efforts have been devoted to accelerating the PGM. Beck and Teboulle [12] proposed the fast iterative shrinkage-thresholding algorithm (FISTA), whose iterates satisfy

$$\Phi(\zeta_d) - \min_{\zeta} \Phi(\zeta) = \mathcal{O}\left(\frac{1}{d^2}\right),$$

with the same conditions. Many first-order methods for minimizing Φ can be interpreted as arising from suitable time discretizations of underlying differential equations. In particular, Su et al. [29] established a fundamental relation between Nesterov's accelerated gradient method (NAG) and second-order differential equation, providing a continuous-time perspective on acceleration; see also [4] for further developments. They proposed the inertial system characterized by a diminishing viscous damping coefficient $\frac{\alpha}{t}$ $(\text{AVD})_\alpha$:

$$\ddot{X} + \frac{\alpha}{t}\dot{X} + \nabla\phi(X(t)) = 0. \tag{3}$$

On building faster optimization algorithms, the vanishing damping $\frac{\alpha}{t}$ is an essential rule [[3], [16], [18]]. For $\alpha \geq 3$, each trajectory $X(\cdot)$ of AVD_α satisfies the asymptotic rate of convergence of

$$\phi(X(t)) - \min\phi = \mathcal{O}\Big(\frac{1}{t^2}\Big). \tag{4}$$

It is now well recognized that dynamical-system approaches provide powerful tools for solving optimization problems; see, for example, [6, 15, 19, 23, 28, 33]. In particular, suitable time discretizations of inertial dynamical systems with vanishing damping of the form $\frac{\alpha}{t}$ lead to accelerated forward–backward (AFB) algorithms [1, 5, 10, 23], accelerated higher-order gradient methods [22, 32, 33], and inertial primal–dual schemes [14, 18]. Later, Attouch and Peypouquet [3] introduced the AFB method by combining FISTA with a second-order differential inclusion with vanishing damping, and proved the improved rate $o\big(\frac{1}{d^2}\big)$ for function values, and (ζ_d) converges to a minimum point when $\alpha > 3$. In [8], accelerated proximal gradient algorithms were proposed for solving Problem 1, although no convergence guarantees were provided. Nonetheless, as stated in [7], the iterates of the inertial system $(\text{AVD})_\alpha$ for the Rosenbrook function will exhibit oscillations.

To counteract oscillations, Attouch et al. [[2], [7]] investigated the class of dynamical systems by adding an HDD term $\nabla^2\phi(X(t))\dot{X}(t)$ to the inertial system suggested in [29]. They studied the inertial system with viscous friction and Hessian-driven damping (HDD) $(\text{DIN-AVD})_{(\alpha,\beta,b)}$:

$$\ddot{X} + \frac{\alpha}{t}\dot{X} + \beta(t)\nabla^2\phi(X(t))\dot{X} + b(t)\nabla\phi(X(t)) = 0, \tag{5}$$

where $b(t)$ is a time scale parameter, and $\beta(t)$ is damping parameter. Furthermore, the inclusion of the Hessian term does not increase the computational burden. Indeed, in the differential equation, the Hessian appears through the expression $\nabla^2\phi(X(t))\dot{X}(t)$, which can be interpreted as the time derivative of the mapping $t \mapsto \nabla\phi(X(t))$. This HDD term establishes a close connection with Newton-type methods and plays an important role in mitigating transverse oscillations that may arise in the dynamical system 5; see [2, 6, 13]. Moreover, the system $(\text{DIN-AVD})_{\alpha,\beta,b}$ retains the convergence properties of the classical system $(\text{AVD})_\alpha$ while ensuring a rapid decay of the gradient norm. Furthermore, it has been demonstrated that the second-order differential equation with HDD can be equivalently reformulated as a first-order system in both time and space; see [17].

Subsequently, Shi et al. [28, 27] introduced a high-resolution dynamical system combining HDD with vanishing damping $\frac{\alpha}{t}$, obtained by refining NAG method. Using implicit Euler and symplectic Euler discretizations, they showed that for $\alpha \geq 3$ and smooth convex functions Φ, the convergence rates satisfy [28, 27]

$$\Phi(\zeta_d) - \min_{\zeta} \Phi(\zeta) = \mathcal{O}\left(\frac{1}{d^2}\right), \qquad \min_{1\leq i\leq d} \|\nabla\Phi(x_i)\|^2 = \mathcal{O}\left(\frac{1}{d^3}\right).$$

In addition, explicit–implicit discretizations of the continuous dynamic system $(\text{DIN-AVD})_{(\alpha,\beta,1)}$ [2] give rise to rapid forward–backward inertial algorithms closely related to FISTA, tailored for composite optimization problems. Furthermore, Attouch et al. [7] examined the system $(\text{DIN-AVD})_{(\alpha,\beta,1+\frac{\beta}{t})}$ and determined the rate $\mathcal{O}\big(\frac{1}{t^2}\big)$ for $\Phi(X(t)) - \min_x \Phi$ for $\alpha \geq 3$ and $\beta \geq 0$. More recently, Attouch et al. [9] investigated the convergence of discrete iterates generated by first-order methods through system 5, proving the rate $o\big(\frac{1}{d^2}\big)$ in the discrete case and the corresponding rate $o\big(\frac{1}{t^2}\big)$ for the continuous dynamics when $\alpha > 3$.

Based on the time discretization of $(\text{DIN-AVD})_{(\alpha,\beta,1+\frac{\beta}{t})}$, He and Fang [20] proposed the AFB algorithm with subgradient correction (AFBSC), given by

$$\begin{cases} \nu_d = \zeta_d + \dfrac{d-1}{d+\alpha-1}(\zeta_d - \zeta_{d-1}) - \dfrac{\beta\sqrt{s}(d+\alpha-2)}{d+\alpha-1}\big(\xi_d + \nabla\phi(\nu_{d-1})\big), \\ \zeta_{d+1} = \text{Prox}_{\gamma g}\big(\nu_d - \gamma\nabla\phi(\nu_d)\big), \\ \xi_{d+1} = -\nabla\phi(\nu_d) - \dfrac{1}{\gamma}(\zeta_{d+1} - \nu_d), \end{cases} \tag{6}$$

where $\alpha \geq 3$ and $\gamma = s + \beta\sqrt{s}$ for $s > 0$. They have obtained $\Phi(\zeta_d) - \min_\zeta \Phi(\zeta) = \mathcal{O}\big(\frac{1}{d^2}\big)$ and $\min_{1\leq i\leq d} \text{dist}^2(0, \partial\Phi(x_i)) = \mathcal{O}\big(\frac{1}{d^3}\big)$. When $\alpha > 3$, the convergence rate of the function values is improved to $o\big(\frac{1}{d^2}\big)$.

Wang et al. [30] observed that the velocity $\dot{X}(t)$ can be discretized using a convex combination of implicit and explicit schemes, leading to the AFB iteration. The implicit contribution is given by $\theta(\zeta_{d+1} - \zeta_d)$, while the explicit contribution is $(1-\theta)(\zeta_d - \zeta_{d-1})$, with $\theta = 1 - \frac{1}{\alpha}$. They improved the range of θ from $(\frac{2}{3}, 1)$ to $[0, +\infty)$, and thus θ controls the balance between numerical stability and efficiency. A small θ may lead to instability, whereas an excessively large θ may degrade performance. They proposed an AFB method by discretizing the ODE (DIN-AVD)$_{(\alpha,\beta,b)}$ with $\beta_d = 0$ and $b_d = \frac{d+\alpha\theta}{d}$, which is similar to (AVD)$_\alpha$ for non-smooth composite optimization.

In our work, we introduce an inertial proximal gradient method obtained by discretizing implicitly-explicitly an inertial dynamical system that incorporates **both a HDD term** $\nabla^2\Phi(X(t))\dot{X}(t)$ **and a vanishing damping term** $\dot{X}(t)$. We define a corresponding discrete energy sequence for Algorithm 1, inspired by the Lyapunov function of the associated continuous time dynamical system.

Main Contribution

(a) We discretize the ODE (DIN-AVD)$_{(\alpha,\beta_d,b_d)}$ in a implicit-explicit way where $\beta_d = \frac{(d+\alpha\theta)\beta}{d}$ and $b_d = \frac{d+\alpha\theta}{d} + \frac{\beta}{d\sqrt{s}}$ for solving the convex optimization problem 1. In this discretization HDD term $\nabla^2\Phi(X(t))\dot{X}(t)$ is there which is absent in the [30]. The proposed method also introduces the parameter θ to control the proportion between the implicit and explicit discretization of velocity term $\dot{X}(t)$, and θ lies in $[0, +\infty)$.

(b) The step size in [30] is less than $\frac{1}{L_\phi}$; we have also established that the step size γ is greater than $\frac{1}{L_\phi}$ for $\beta > 0$, which is an improvement compared to the result in [30].

(c) The proposed algorithm is the generalization of the algorithm, AFBSC [20], by introducing the parameter θ, which means when the value of $\theta = \frac{\alpha-1}{\alpha}$, the Algorithm 1 reduces to the algorithm AFBSC.

(d) We prove that, for $\alpha \geq 3$ and $\theta \geq 0$, the convergence rates satisfy $\Phi(\zeta_d) - \min_\zeta \Phi(\zeta) = \mathcal{O}\left(\frac{1}{d^2}\right)$ and improve to $o\left(\frac{1}{d^2}\right)$ when $\alpha > 3$. We have also obtained $\min_{1\leq i\leq d} \operatorname{dist}^2(0, \partial\Phi(x_i)) = \mathcal{O}\left(\frac{1}{d^3}\right)$ corresponding to $\alpha \geq 3$ and $\theta \geq 0$, which is the new one associated with θ.

(e) To demonstrate the effectiveness and practical usefulness of the proposed strategy, we have investigated two composite convex optimization problems: one is ℓ_1-regularized logistic regression, and the other is the Lasso problem. By choosing $\alpha = 90$ and $\theta = 10$ for the Lasso problem and $\alpha = 60$ and $\theta = 5$ for the ℓ_1-regularized logistic regression problem, it has been shown that the proposed method gives better results as compared to the AFBSC [20].

The remainder of this paper is organized as follows. Sect. 2 presents the necessary preliminaries. In Sect. 3, we discretize the ODE and propose the algorithm. Sect. 4 introduces a Lyapunov function for analyzing the algorithm. In Sect. 5, we study the convergence of the iterative sequence generated by the proposed method. Numerical experiments are reported in Sect. 6 to validate the theoretical results. Conclusion is provided in Sect. 7.

2 Preliminaries

To support the subsequent analysis, we first introduce some notation and basic concepts. For background on convex analysis, we refer to [11], [26]. The standard Euclidean norm and inner product on $\mathbb{R}^n$ are denoted by $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$, respectively, while $\|\cdot\|_1$ denotes the ℓ_1-norm. Let K be a proper closed subset of $\mathbb{R}^n$, we define $\mathrm{dist}^2(0,K) = \min_{w\in K}\|w\|^2$. For a function $\psi : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$, its domain is given by $\mathrm{dom}(\psi) = \{w \in \mathbb{R}^n \mid \psi(w) < +\infty\}$. The function ψ is called proper if $\mathrm{dom}(\psi) \neq \varnothing$, and it is lower semi-continuous at w if $\psi(w) \le \lim_{u\to w}\inf \psi(u)$.

The function ϕ on $\mathbb{R}^n$ is convex if $\phi(\lambda u + (1-\lambda)v) \le \lambda\phi(u) + (1-\lambda)\phi(v)$, for all $u, v \in \mathbb{R}^n$ and for any $\lambda \in [0,1]$ [26].

The proximal operator $\mathrm{Prox}_{\lambda\psi} : \mathbb{R}^n \to \mathbb{R}^n$ with respect to a proper lower semi-continuous function $\lambda\psi$ for $\lambda > 0$ is defined as

$$\mathrm{Prox}_{\lambda\psi}(w) := \arg\min_{u\in\mathbb{R}^n}\Big\{\psi(u) + \frac{1}{2\lambda}\|w-u\|^2\Big\}, \tag{7}$$

for all $w \in \mathbb{R}^n$ [11]. We make the following assumptions:

Assumption 1. *(a) $\phi : \mathbb{R}^n \to \mathbb{R}$ is a differentiable convex function whose gradient $\nabla\phi$ is Lipschitz continuous with constant L_ϕ.*

(b) $\psi : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is a proper, lower semi-continuous convex function.

(c) $\arg\min_{x\in\mathbb{R}^n}\Phi = S \neq \varnothing$.

The following outcomes are essential.

Lemma 1 ((Theorem 2.1.5, [25]), (Lemma 3, [21]))**.** *If $\phi : \mathbb{R}^n \to \mathbb{R}$ is a convex function with the gradient vector $\nabla\phi$ is Lipschitz continuous function with constant L_ϕ, then we have*

(i). $\langle\nabla\phi(w), u-v\rangle \ge \phi(u) - \phi(v) - \frac{L_\phi}{2}\|u-w\|^2$.

(ii). $\langle\nabla\phi(u) - \nabla\phi(v), u-v\rangle \ge \frac{L_\phi}{2}\|\nabla\phi(u) - \nabla\phi(v)\|^2$,

for all u, v and w belongs to $\mathbb{R}^n$.

Lemma 2 (Lemma 5.31, [11])**.** *Let $\{a_d\}_{d\ge1}$ be a sequence in $\mathbb{R}$ that is bounded from below, and let $\{b_d\}_{d\ge1}$ and $\{c_d\}_{d\ge1}$ be nonnegative sequences such that $\sum_{d=1}^{\infty} c_d < \infty$. Assume that, for all $d \ge 1$,*

$$a_{d+1} \le a_d - b_d + c_d.$$

Then the sequence $\{a_d\}_{d\ge1}$ converges, and moreover, $\sum_{d=1}^{\infty} b_d < \infty$.

Lemma 3 (Lemma 2.47, [11])**.** *Suppose that $\{\zeta_d\}_{d\ge1}$ is a sequence in $\mathbb{R}^n$ and S is a nonempty subset of $\mathbb{R}^n$. Assume*

(i) For every $\zeta^* \in S$, the limit $\lim_{d\to\infty}\|\zeta_d - \zeta^*\|$ exists.

(ii) Each cluster point of the sequence $\{\zeta_d\}$ belongs to S.

Then the sequence $\{\zeta_d\}_{d\geq 1}$ converges to some point in S.

Lemma 4. *[Theorem 2.1.5, [24]] If $\phi : \mathbb{R}^n \to \mathbb{R}$ is a convex function with the gradient vector $\nabla\phi$ is Lipschitz continuous function with constant L_ϕ, then we have*

$$\frac{1}{L_\phi}\|\nabla\phi(u) - \nabla\phi(v)\|^2 \leq \langle \nabla\phi(u) - \nabla\phi(v), u - v\rangle,$$

for all $u, v \in \mathbb{R}^n$.

Lemma 5. *[20] For any $u, v \in \mathbb{R}^n$ and $s > 0$, the following holds*

$$\langle u, u - v\rangle - \frac{1}{2}\|u - v\|^2 = \frac{1}{2}\|u\|^2 - \frac{1}{2}\|v\|^2, \tag{8}$$

$$-\frac{s}{2}\|u\|^2 - \frac{1}{2s}\|v\|^2 \leq \langle u, v\rangle \leq \frac{s}{2}\|u\|^2 + \frac{1}{2s}\|v\|^2. \tag{9}$$

Proof. The proof is obvious. □

3 Discretization of high-resolution of ODE

We discretize the ODE 5 by setting $\beta_d = \frac{(d+\alpha\theta)\beta}{k}, b_d = \frac{d+\alpha\theta}{k} + \frac{\beta}{d\sqrt{s}}$, where $\theta \geq 0, \beta > 0$, $s > 0$, and $\alpha > 3$. For the time step size is fixed $\sqrt{s}$ and $t_d = d\sqrt{s}$, the implicit-explicit finite difference scheme approach provides

$$\begin{aligned}
&\frac{\zeta_{d+1} - \zeta_d}{s} - \frac{\zeta_d - \zeta_{d-1}}{s} + \frac{\alpha\theta}{k}\Big(\frac{\zeta_{d+1} - \zeta_d}{s}\Big) + \frac{\alpha(1-\theta)}{k}\Big(\frac{\zeta_d - \zeta_{d-1}}{s}\Big)\\
&+ \frac{(d+\alpha\theta)\beta}{d\sqrt{s}}\Big(\nabla\psi(\zeta_{d+1})) + \nabla\phi(\zeta_{d+1}) - \big(\nabla\psi(\zeta_d) + \nabla\phi(\zeta_d)\big)\Big)\\
&+ \frac{d+\alpha\theta}{k}\big(\nabla\psi(\zeta_{d+1})) + \nabla\phi(\zeta_{d+1})\big) + \frac{\beta}{d\sqrt{s}}\big(\nabla\psi(\zeta_d) + \nabla\phi(\zeta_d)\big) = 0.
\end{aligned} \tag{10}$$

Multiplying s in Equation 10, we obtain

$$\begin{aligned}
&(\zeta_{d+1} - \zeta_d) - (\zeta_d - \zeta_{d-1}) + \frac{\alpha\theta}{k}(\zeta_{d+1} - \zeta_d) + \frac{\alpha(1-\theta)}{k}(\zeta_d - \zeta_{d-1})\\
&+ \frac{(d+\alpha\theta)\beta\sqrt{s}}{k}\Big(\nabla\psi(\zeta_{d+1}) + \nabla\phi(\zeta_{d+1}) - \big(\nabla\psi(\zeta_d) + \nabla\phi(\zeta_d)\big)\Big)\\
&+ \frac{d+\alpha\theta}{k}s\big(\nabla\psi(\zeta_{d+1}) + \nabla\phi(\zeta_{d+1})\big) + \frac{\beta\sqrt{s}}{k}\big(\nabla\psi(\zeta_d) + \nabla\phi(\zeta_d)\big) = 0.
\end{aligned}$$

Also,

$$\begin{aligned}&\frac{d+\alpha\theta}{k}(\zeta_{d+1}-\zeta_d)-\frac{d+\alpha(\theta-1)}{k}(\zeta_d-\zeta_{d-1})\\&+\frac{(d+\alpha\theta)\beta\sqrt{s}}{k}\Big(\nabla\psi(\zeta_{d+1})+\nabla\phi(\zeta_{d+1})-\big(\nabla\psi(\zeta_d)+\nabla\phi(\zeta_d)\big)\Big)\\&+\frac{d+\alpha\theta}{k}s\big(\nabla\psi(\zeta_{d+1})+\nabla\phi(\zeta_{d+1})\big)+\frac{\beta\sqrt{s}}{k}\big(\nabla\psi(\zeta_d)+\nabla\phi(\zeta_d)\big)=0.\end{aligned}$$

Now, we obtain

$$\begin{aligned}\zeta_{d+1}&=\zeta_d+\frac{d+\alpha(\theta-1)}{d+\alpha\theta}(\zeta_d-\zeta_{d-1})\\&-\beta\sqrt{s}\Big(\nabla\psi(\zeta_{d+1})+\nabla\phi(\zeta_{d+1})-\big(\nabla\psi(\zeta_d)+\nabla\phi(\zeta_d)\big)\Big)\\&-s\big(\nabla\psi(\zeta_{d+1})+\nabla\phi(\zeta_{d+1})\big)-\frac{\beta\sqrt{s}}{d+\alpha\theta}\big(\nabla\psi(\zeta_d)+\nabla\phi(\zeta_d)\big).\end{aligned}\tag{11}$$

However, the update of Equation 11 is implicit, we use $\nabla\phi(\nu_d)$ to approximate $\nabla\phi(\zeta_{d+1})$, where

$$\nu_d=\zeta_d+\frac{d+\alpha(\theta-1)}{d+\alpha\theta}(\zeta_d-\zeta_{d-1})+\frac{(d+\alpha\theta-1)\beta\sqrt{s}}{d+\alpha\theta}\big(\nabla\psi(\zeta_d)+\nabla\phi(\nu_{d-1})\big).$$

Now,

$$\zeta_{d+1}=\nu_d-(\beta\sqrt{s}+s)(\nabla\psi(\zeta_{d+1})+\nabla\phi(\nu_d)).$$

Also,

$$\zeta_{d+1}+\gamma\nabla\psi(\zeta_{d+1})=\nu_d-\gamma\nabla\phi(\nu_d),$$

where $\gamma=\beta\sqrt{s}+s$. Since we assume ψ is a convex lower semi-continuous (possibly non-smooth) function, we obtain

$$\zeta_{d+1}=Prox_{\gamma\psi}(\nu_d-\gamma\nabla\phi(\nu_d)).\tag{12}$$

From Equation 12, we have $-\frac{1}{\gamma}(\zeta_{d+1}-\nu_d)-\nabla\phi(\nu_d)\in\partial\psi(\zeta_{d+1})$. We can choose

$$\xi_{d+1}=-\frac{1}{\gamma}(\zeta_{d+1}-\nu_d)-\nabla\phi(\nu_d)\in\partial\psi(\zeta_{d+1}).\tag{13}$$

Based on the the above analysis, we now propose the corresponding algorithm.

Algorithm 1 Improved accelerated forward-backward algorithm with subgradient correction (IAFBSC)

Initialize: Let $\zeta_1 = \zeta_0 = \nu_0 \in \mathbb{R}^n$, $\xi_1 \in \partial g(x_1)$, $\alpha \geq 3, \theta \geq 0$, $s > 0$, $\beta \geq 0$, $\gamma = s + \beta\sqrt{s}$.

for $d = 1, 2, \dots$ **do**

Step i:

$$\nu_d = \zeta_d + \frac{d + \alpha(\theta - 1)}{d + \alpha\theta}(\zeta_d - \zeta_{d-1}) + \frac{(d + \alpha\theta - 1)\beta\sqrt{s}}{d + \alpha\theta}\big(\xi_d + \nabla\phi(\nu_{d-1})\big),$$

Step ii:

$$\zeta_{d+1} = \mathrm{Prox}_{\gamma\psi}\big(\nu_d - \gamma\nabla\phi(\nu_d)\big),$$

Step iii:

$$\xi_{d+1} = -\nabla\phi(\nu_d) - \frac{1}{\gamma}(\zeta_{d+1} - \nu_d).$$

Until a stopping criterion is met

end for

return (ζ_d, ν_d, ξ_d)

4 Construction of Lyapunov Function

We introduce the energy sequence $\{\mathcal{E}_d(\omega)\}_{d\geq 1}$ defined by

$$\begin{aligned}\mathcal{E}_d(\omega) :=& \Big(s(d + 1 + \alpha(\theta - 1)) + \eta(\beta\sqrt{s} + s)\Big)(d + \alpha\theta)\big(\Phi(\zeta_d) - \Phi(\zeta^*)\big) \\ &+ \frac{1}{2}\|u_d(\omega)\|^2 + \frac{\omega\eta}{2}\|\zeta_d - \zeta^*\|^2,\end{aligned} \tag{14}$$

where

$$\begin{aligned}u_d(\omega) =& \omega(\zeta_d - \zeta^*) + (d + \alpha(\theta - 1))(\zeta_d - \zeta_{d-1}) \\ &+ \beta\sqrt{s}(d + \alpha\theta - 1)\big(\xi_d + \nabla\phi(\nu_{d-1})\big),\end{aligned} \tag{15}$$

for any $\theta \geq 0$, $\omega \in [2, \alpha]$, and $\eta = \alpha - \omega - 1$. We are now in a position to state the following result.

Lemma 6. *Let Assumption 1 holds, and $\{(\zeta_d, \nu_d, \xi_d)\}_{d\geq 1}$ be the sequences produced by Algorithm 1, and let $\zeta^* \in S$. We define the associated Lyapunov function $\mathcal{E}_d(\omega)$ as in* (14).

Then, we have

$$
\begin{aligned}
&\mathcal{E}_{d+1}(\omega) - \mathcal{E}_d(\omega) \\
&\leq \big(s(d+\alpha\theta)(3-\alpha+\eta) + s(2-\alpha+\eta) + \eta\beta\sqrt{s}\big)(\Phi(\zeta_{d+1}) - \Phi(\zeta^*)) \\
&- \frac{s(d+\alpha\theta)^2}{2}\Big[s + 2\beta\sqrt{s} - L_\phi(s+\beta\sqrt{s})^2\Big(1 + \frac{\eta\beta}{\sqrt{s}(d+\alpha\theta)}\Big)\Big]\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2 \\
&- \Big[\eta(d+1+\alpha(\theta-1)) + \frac{\eta^2}{2} + \frac{\omega\eta}{2}\Big]\|\zeta_{d+1} - \zeta_d\|^2.
\end{aligned}
$$

Proof. From Equation 15, we get

$$
\begin{aligned}
&u_{d+1}(\omega) - u_d(\omega) \\
&= \omega(\zeta_{d+1} - \zeta^*) + (d+1+\alpha(\theta-1))(\zeta_{d+1} - \zeta_d) \\
&+ \beta\sqrt{s}(d+\alpha\theta)(\xi_{d+1} + \nabla\phi(\nu_d)) - \omega(\zeta_d - \zeta^*) - (d+\alpha(\theta-1))(\zeta_d - \zeta_{d-1}) \\
&- \beta\sqrt{s}(d+\alpha\theta-1)(\xi_d + \nabla\phi(\nu_{d-1})) \\
&= \omega(\zeta_{d+1} - \zeta_d) + (d+1+\alpha(\theta-1))(\zeta_{d+1} - \zeta_d) \\
&+ \beta\sqrt{s}(d+\alpha\theta)(\xi_{d+1} + \nabla\phi(\nu_d)) - (d+\alpha(\theta-1))(\zeta_d - \zeta_{d-1}) \\
&- \beta\sqrt{s}(d+\alpha\theta-1)(\xi_d + \nabla\phi(\nu_{d-1})).
\end{aligned} \tag{16}
$$

Putting $\omega = \alpha - \eta - 1$ in Equation 16, we obtain

$$
\begin{aligned}
&u_{d+1}(\omega) - u_d(\omega) \\
&= -\eta(\zeta_{d+1} - \zeta_d) + (d+\alpha\theta)(\zeta_{d+1} - \zeta_d) + \beta\sqrt{s}(d+\alpha\theta)(\xi_{d+1} + \nabla\phi(\nu_d)) \\
&- (d+\alpha(\theta-1))(\zeta_d - \zeta_{d-1}) - \beta\sqrt{s}(d+\alpha\theta-1)(\xi_d + \nabla\phi(\nu_{d-1})).
\end{aligned} \tag{17}
$$

Using stage i of IAFBSC, we get

$$
\begin{aligned}
&u_{d+1}(\omega) - u_d(\omega) \\
&= -\eta(\zeta_{d+1} - \zeta_d) + (d+\alpha\theta)(\zeta_{d+1} - \zeta_d) + \beta\sqrt{s}(d+\alpha\theta)(\xi_{d+1} + \nabla\phi(\nu_d)) \\
&- (d+\alpha\theta)(\nu_d - \zeta_d) \\
&= -\eta(\zeta_{d+1} - \zeta_d) + (d+\alpha\theta)(\zeta_{d+1} - \nu_d) + \beta\sqrt{s}(d+\alpha\theta)(\xi_{d+1} + \nabla\phi(\nu_d)).
\end{aligned} \tag{18}
$$

From step iii of IAFBSC and using $\gamma = s + \beta\sqrt{s}$, we deduce

$$
\begin{aligned}
&u_{d+1}(\omega) - u_d(\omega) \\
&= -\eta(\zeta_{d+1} - \zeta_d) - \gamma(d+\alpha\theta)(\xi_{d+1} + \nabla\phi(\nu_d)) + \beta\sqrt{s}(d+\alpha\theta)(\xi_{d+1} + \nabla\phi(\nu_d)) \\
&= -\eta(\zeta_{d+1} - \zeta_d) - (\beta\sqrt{s} + s)(d+\alpha\theta)(\xi_{d+1} + \nabla\phi(\nu_d)) \\
&+ \beta\sqrt{s}(d+\alpha\theta)(\xi_{d+1} + \nabla\phi(\nu_d)) \\
&= -\eta(\zeta_{d+1} - \zeta_d) - s(d+\alpha\theta)(\xi_{d+1} + \nabla\phi(\nu_d)).
\end{aligned} \tag{19}
$$

Using Equation 15 and 19, we obtain

$$
\begin{aligned}
&\langle u_{d+1}(\omega), u_{d+1}(\omega) - u_d(\omega)\rangle \\
&= \langle \omega(\zeta_{d+1} - \zeta^*) + (d+1+\alpha(\theta-1))(\zeta_{d+1} - \zeta_d) \\
&+ \beta\sqrt{s}(d+\alpha\theta)(\xi_{d+1} + \nabla\phi(\nu_d)), -\eta(\zeta_{d+1} - \zeta_d) - s(d+\alpha\theta)(\xi_{d+1} + \nabla\phi(\nu_d))\rangle \\
&= -\omega\eta\langle \zeta_{d+1} - \zeta^*, \zeta_{d+1} - \zeta_d\rangle - \eta(d+1+\alpha(\theta-1))\|\zeta_{d+1} - \zeta_d\|^2 \\
&- \eta\beta\sqrt{s}(d+\alpha\theta)\langle \zeta_{d+1} - \zeta_d, \xi_{d+1} + \nabla\phi(\nu_d)\rangle - s\beta\sqrt{s}(d+\alpha\theta)^2\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2 \\
&- \omega s(d+\alpha\theta)\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \\
&- s(d+\alpha\theta)(d+1+\alpha(\theta-1))\langle \zeta_{d+1} - \zeta_d, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \\
&= -\omega\eta\langle \zeta_{d+1} - \zeta^*, \zeta_{d+1} - \zeta_d\rangle - \eta(d+1+\alpha(\theta-1))\|\zeta_{d+1} - \zeta_d\|^2 \\
&- (\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)(d+1+\alpha(\theta-1)))\langle \zeta_{d+1} - \zeta_d, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \\
&- \omega s(d+\alpha\theta)\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d)\rangle - s\beta\sqrt{s}(d+\alpha\theta)^2\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2. \qquad (20)
\end{aligned}
$$

Next,

$$
\begin{aligned}
\|u_{d+1}(\omega) - u_d(\omega)\|^2 &= \|\eta(\zeta_{d+1} - \zeta_d) + s(d+\alpha\theta)(\xi_{d+1} + \nabla\phi(\nu_d))\|^2 \\
&= \eta^2\|\zeta_{d+1} - \zeta_d\|^2 + s^2(d+\alpha\theta)^2\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2 \\
&+ 2\eta s(d+\alpha\theta)\langle \zeta_{d+1} - \zeta_d, \xi_{d+1} + \nabla\phi(\nu_d)\rangle. \qquad (21)
\end{aligned}
$$

Using Equation 8, we get

$$
\begin{aligned}
&\frac{1}{2}\|u_{d+1}(\omega)\|^2 + \frac{\omega\eta}{2}\|\zeta_{d+1} - \zeta^*\|^2 - \frac{1}{2}\|u_d(\omega)\|^2 - \frac{\omega\eta}{2}\|\zeta_d - \zeta^*\|^2 \\
&= \langle u_{d+1}(\omega), u_{d+1}(\omega) - u_d(\omega)\rangle - \frac{1}{2}\|u_{d+1}(\omega) - u_d(\omega)\|^2 \\
&+ \omega\eta\langle \zeta_{d+1} - \zeta^*, \zeta_{d+1} - \zeta_d\rangle - \frac{\omega\eta}{2}\|\zeta_{d+1} - \zeta_d\|^2.
\end{aligned}
$$

From Equation 20 and 21, we obtain

$$
\begin{aligned}
&\frac{1}{2}\|u_{d+1}(\omega)\|^2 + \frac{\omega\eta}{2}\|\zeta_{d+1} - \zeta^*\|^2 - \frac{1}{2}\|u_d(\omega)\|^2 - \frac{\omega\eta}{2}\|\zeta_d - \zeta^*\|^2 \\
&= -\omega\eta\langle \zeta_{d+1} - \zeta^*, \zeta_{d+1} - \zeta_d\rangle - \eta(d+1+\alpha(\theta-1))\|\zeta_{d+1} - \zeta_d\|^2 \\
&- (\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)(d+1+\alpha(\theta-1)))\langle \zeta_{d+1} - \zeta_d, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \\
&- \omega s(d+\alpha\theta)\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d)\rangle - s\beta\sqrt{s}(d+\alpha\theta)^2\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2 \\
&- \frac{\eta^2}{2}\|\zeta_{d+1} - \zeta_d\|^2 - \frac{s^2(d+\alpha\theta)^2}{2}\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2 - \frac{\omega\eta}{2}\|\zeta_{d+1} - \zeta_d\|^2 \\
&- \eta s(d+\alpha\theta)\langle \zeta_{d+1} - \zeta_d, \xi_{d+1} + \nabla\phi(\nu_d)\rangle + \omega\eta\langle \zeta_{d+1} - \zeta^*, \zeta_{d+1} - \zeta_d\rangle. \qquad (22)
\end{aligned}
$$

Also, we have

$$
\begin{aligned}
&\frac{1}{2}\|u_{d+1}(\omega)\|^2 + \frac{\omega\eta}{2}\|\zeta_{d+1} - \zeta^*\|^2 - \frac{1}{2}\|u_d(\omega)\|^2 - \frac{\omega\eta}{2}\|\zeta_d - \zeta^*\|^2 \\
&= -\Big[\eta(d+1+\alpha(\theta-1)) + \frac{\eta^2}{2} + \frac{\omega\eta}{2}\Big]\|\zeta_{d+1} - \zeta_d\|^2 \\
&- \omega s(d+\alpha\theta)\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \\
&- \eta s(d+\alpha\theta)\langle \zeta_{d+1} - \zeta_d, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \\
&- \Big[s\beta\sqrt{s}(d+\alpha\theta)^2 + \frac{s^2(d+\alpha\theta)^2}{2}\Big]\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2 \\
&- (\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)(d+1+\alpha(\theta-1)))\langle \zeta_{d+1} - \zeta_d, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \\
&= -\Big[\eta(d+1+\alpha(\theta-1)) + \frac{\eta^2}{2} + \frac{\omega\eta}{2}\Big]\|\zeta_{d+1} - \zeta_d\|^2 \\
&- \omega s(d+\alpha\theta)\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \\
&- \frac{s(d+\alpha\theta)^2}{2}(s + 2\beta\sqrt{s})\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2 \\
&- \Big[\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)(d+1+\alpha(\theta-1)+\eta)\Big] \\
&\times \langle \zeta_{d+1} - \zeta_d, \xi_{d+1} + \nabla\phi(\nu_d)\rangle.
\end{aligned} \tag{23}
$$

Using step ii and step iii of IAFBSC, we obtain

$$\xi_{d+1} \in \partial\psi(\zeta_{d+1}).$$

Therefore, taking into account the convexity of ψ for any $w \in \mathbb{R}^n$, we obtain

$$\langle \xi_{d+1}, \zeta_{d+1} - w\rangle \geq \psi(\zeta_{d+1}) - \psi(w). \tag{24}$$

Lemma 1 (i), step iii of IAFBSC implies that

$$
\begin{aligned}
\langle \nabla\phi(\nu_d), \zeta_{d+1} - w\rangle &\geq \phi(\zeta_{d+1}) - \phi(w) - \frac{L_\phi}{2}\|\zeta_{d+1} - \nu_d\|^2 \\
&= \phi(\zeta_{d+1}) - \phi(w) - \frac{L_\phi\gamma^2}{2}\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2.
\end{aligned} \tag{25}
$$

Adding Equation 24 and 25, we obtain

$$
\begin{aligned}
&\langle \zeta_{d+1} - w, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \\
&\geq \phi(\zeta_{d+1}) + \psi(\zeta_{d+1}) - \phi(w) - \psi(w) - \frac{L_\phi\gamma^2}{2}\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2 \\
&= \Phi(\zeta_{d+1}) - \Phi(w) - \frac{L_\phi\gamma^2}{2}\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2.
\end{aligned} \tag{26}
$$

Replacing w in Equation 26 with ζ^* and ζ_d, respectively, we arrive at

$$\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d) \rangle \geq \Phi(\zeta_{d+1}) - \Phi(\zeta^*) - \frac{L_\phi \gamma^2}{2} \|\xi_{d+1} + \nabla\phi(\nu_d)\|^2, \tag{27}$$

and

$$\langle \zeta_{d+1} - \zeta_d, \xi_{d+1} + \nabla\phi(\nu_d) \rangle \geq \Phi(\zeta_{d+1}) - \Phi(\zeta_d) - \frac{L_\phi \gamma^2}{2} \|\xi_{d+1} + \nabla\phi(\nu_d)\|^2. \tag{28}$$

From Equation 27 and 28, we get

$$\begin{aligned}
&\omega s(d+\alpha\theta) \langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d) \rangle \\
&+ \Big[\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)(d+1+\alpha(\theta-1)+\eta)\Big] \langle \zeta_{d+1} - \zeta_d, \xi_{d+1} + \nabla\phi(\nu_d) \rangle \\
&\geq \omega s(d+\alpha\theta)\Big[\Phi(\zeta_{d+1}) - \Phi(\zeta^*) - \frac{L_\phi \gamma^2}{2} \|\xi_{d+1} + \nabla\phi(\nu_d)\|^2\Big] \\
&+ \Big[\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)(d+1+\alpha(\theta-1)+\eta)\Big] \\
&\times \Big[\Phi(\zeta_{d+1}) - \Phi(\zeta_d) - \frac{L_\phi \gamma^2}{2} \|\xi_{d+1} + \nabla\phi(\nu_d)\|^2\Big] \\
&= \omega s(d+\alpha\theta)(\Phi(\zeta_{d+1}) - \Phi(\zeta^*)) \\
&+ \Big[\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)(d+1+\alpha(\theta-1)+\eta)\Big](\Phi(\zeta_{d+1}) - \Phi(\zeta_d)) \\
&- \frac{L_\phi \gamma^2}{2}\Big[\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)(d+1+\alpha(\theta-1)+\eta) + \omega s(d+\alpha\theta)\Big] \\
&\times \|\xi_{d+1} + \nabla\phi(\nu_d)\|^2 \\
&= \omega s(d+\alpha\theta)(\Phi(\zeta_{d+1}) - \Phi(\zeta^*)) \\
&+ \Big[\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)(d+1+\alpha(\theta-1)+\eta)\Big](\Phi(\zeta_{d+1}) - \Phi(\zeta^*)) \\
&- \Big[\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)(d+1+\alpha(\theta-1)+\eta)\Big](\Phi(\zeta_d) - \Phi(\zeta^*)) \\
&- \frac{L_\phi \gamma^2}{2}\Big[\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)(d+1+\alpha(\theta-1)+\eta+\omega)\Big] \\
&\times \|\xi_{d+1} + \nabla\phi(\nu_d)\|^2.
\end{aligned} \tag{29}$$

Using $\eta = \alpha - \omega - 1$, Equation 29 becomes

$$
\begin{aligned}
&\omega s(d+\alpha\theta)\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \\
&+ \Big[\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)(d+1+\alpha(\theta-1)+\eta)\Big] \\
&\langle \zeta_{d+1} - \zeta_d, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \\
&\geq \big(\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)^2\big)(\Phi(\zeta_{d+1}) - \Phi(\zeta^*)) \\
&- \Big[\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)(d+1+\alpha(\theta-1)+\eta)\Big](\Phi(\zeta_d) - \Phi(\zeta^*)) \\
&- \frac{L_\phi\gamma^2}{2}\Big[\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)^2\Big]\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2.
\end{aligned} \tag{30}
$$

Then, from Equation 14 and 23, we have

$$
\begin{aligned}
&\mathcal{E}_{d+1}(\omega) - \mathcal{E}_d(\omega) \\
&= \big(s(k+2+\alpha(\theta-1)) + \eta(\beta\sqrt{s}+s)\big)(d+1+\alpha\theta)(\Phi(\zeta_{d+1}) - \Phi(\zeta^*)) \\
&- \big(s(d+1+\alpha(\theta-1)) + \eta(\beta\sqrt{s}+s)\big)(d+\alpha\theta)(\Phi(\zeta_d) - \Phi(\zeta^*)) \\
&+ \frac{1}{2}\|u_{d+1}(\omega)\|^2 + \frac{\omega\eta}{2}\|\zeta_{d+1} - \zeta^*\|^2 - \frac{1}{2}\|u_d(\omega)\|^2 - \frac{\omega\eta}{2}\|\zeta_d - \zeta^*\|^2 \\
&= \big(s(k+2+\alpha(\theta-1)) + \eta(\beta\sqrt{s}+s)\big)(d+1+\alpha\theta)(\Phi(\zeta_{d+1}) - \Phi(\zeta^*)) \\
&- \big(s(d+1+\alpha(\theta-1)) + \eta(\beta\sqrt{s}+s)\big)(d+\alpha\theta)(\Phi(\zeta_d) - \Phi(\zeta^*)) \\
&- \Big[\eta(d+1+\alpha(\theta-1)) + \frac{\eta^2}{2} + \frac{\omega\eta}{2}\Big]\|\zeta_{d+1} - \zeta_d\|^2 \\
&- \omega s(d+\alpha\theta)\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \\
&- \Big[\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)(d+1+\alpha(\theta-1)+\eta)\Big] \\
&\times \langle \zeta_{d+1} - \zeta_d, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \\
&- \frac{s(d+\alpha\theta)^2}{2}(s + 2\beta\sqrt{s})\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2.
\end{aligned} \tag{31}
$$

From Equation 30, we deduce

$$
\begin{aligned}
&\mathcal{E}_{d+1}(\omega) - \mathcal{E}_d(\omega) \\
&\leq \big(s(k+2+\alpha(\theta-1)) + \eta(\beta\sqrt{s}+s)\big)(d+1+\alpha\theta)(\Phi(\zeta_{d+1}) - \Phi(\zeta^*)) \\
&- \big(s(d+1+\alpha(\theta-1)) + \eta(\beta\sqrt{s}+s)\big)(d+\alpha\theta)(\Phi(\zeta_d) - \Phi(\zeta^*)) \\
&- \left[\eta(d+1+\alpha(\theta-1)) + \frac{\eta^2}{2} + \frac{\omega\eta}{2}\right]\|\zeta_{d+1} - \zeta_d\|^2 \\
&- \big(\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)^2\big)(\Phi(\zeta_{d+1}) - \Phi(\zeta^*)) \\
&+ \frac{L_\phi\gamma^2}{2}\left[\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)^2\right]\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2 \\
&+ \left[\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)(d+1+\alpha(\theta-1)+\eta)\right](\Phi(\zeta_d) - \Phi(\zeta^*)) \\
&- \frac{s(d+\alpha\theta)^2}{2}(s+2\beta\sqrt{s})\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2 \\
&= \left[\Big(\big(s(k+2+\alpha(\theta-1)+\eta(\beta\sqrt{s}+s))\big)(d+1+\alpha\theta)\Big)\right. \\
&\left. - \Big(\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)^2\Big)\right](\Phi(\zeta_{d+1}) - \Phi(\zeta^*)) \\
&- \left[\eta(d+1+\alpha(\theta-1)) + \frac{\omega\eta}{2} + \frac{\eta^2}{2}\right]\|\zeta_{d+1} - \zeta_d\|^2 \\
&- \frac{s(d+\alpha\theta)^2}{2}\left[s + 2\beta\sqrt{s} - L_\phi\gamma^2\Big(1 + \frac{\eta\beta}{\sqrt{s}(d+\alpha\theta)}\Big)\right]\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2. \qquad (32)
\end{aligned}
$$

Now,

$$
\begin{aligned}
&\Big(\big(s(k+2+\alpha(\theta-1)+\eta(\beta\sqrt{s}+s))\big)(d+1+\alpha\theta)\Big) - \Big(\eta\beta\sqrt{s}(d+\alpha\theta) + s(d+\alpha\theta)^2\Big) \\
&= s(d+\alpha\theta+2-\alpha)(d+\alpha\theta+1) + \eta\beta\sqrt{s}(d+\alpha\theta+1) + \eta s(d+\alpha\theta+1) \\
&- \eta\beta\sqrt{s}(d+\alpha\theta) - s(d+\alpha\theta)^2 \\
&= s(d+\alpha\theta)^2 + s(2-\alpha)(d+\alpha\theta+1) + s(d+\alpha\theta) + \eta s(d+\alpha\theta+1) \\
&+ \eta\beta\sqrt{s}(d+\alpha\theta) + \eta\beta\sqrt{s} - \eta\beta\sqrt{s}(d+\alpha\theta) - s(d+\alpha\theta)^2 \\
&= s(2-\alpha)(d+1+\alpha\theta) + s(d+\alpha\theta) + \eta\beta\sqrt{s} + \eta s(d+1+\alpha\theta) \\
&= s(d+1+\alpha\theta)(2-\alpha+\eta) + s(d+\alpha\theta) + \eta\beta\sqrt{s} \\
&= s(d+\alpha\theta)(3-\alpha+\eta) + s(2-\alpha+\eta) + \eta\beta\sqrt{s}. \qquad (33)
\end{aligned}
$$

Putting Equation 33 in Equation 32, and since $\gamma = \beta\sqrt{s} + s$, we obtain

$$\begin{aligned}
&\mathcal{E}_{d+1}(\omega) - \mathcal{E}_d(\omega) \\
&\leq \big(s(d+\alpha\theta)(3-\alpha+\eta) + s(2-\alpha+\eta) + \eta\beta\sqrt{s}\big)(\Phi(\zeta_{d+1}) - \Phi(\zeta^*)) \\
&- \frac{s(d+\alpha\theta)^2}{2}\Big[s + 2\beta\sqrt{s} - L_\phi(s+\beta\sqrt{s})^2\Big(1 + \frac{\eta\beta}{\sqrt{s}(d+\alpha\theta)}\Big)\Big]\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2 \\
&- \Big[\eta(d+1+\alpha(\theta-1)) + \frac{\eta^2}{2} + \frac{\omega\eta}{2}\Big]\|\zeta_{d+1} - \zeta_d\|^2.
\end{aligned} \tag{34}$$

□

Theorem 7. *Suppose Assumption 1 holds. Let $\{(\zeta_d, \nu_d, \xi_d)\}_{d\geq 1}$ be the sequences generated by IAFBSC and $\zeta^* \in S$. Assuming $s + 2\beta\sqrt{s} \geq L_\phi(s+\beta\sqrt{s})^2$, we have*

(i). $\Phi(\zeta_d) - \Phi(\zeta^*) = \mathcal{O}\Big(\frac{1}{d^2}\Big)$.

(ii). $(\alpha-3)\sum_{d=1}^{+\infty}(d+\alpha\theta)(\Phi(\zeta_{d+1}) - \Phi(\zeta^*)) < +\infty$.

Moreover, when $s + 2\beta\sqrt{s} > L_\phi(s+\beta\sqrt{s})^2$, we have

(iii).

$$\begin{aligned}
&\frac{s}{2}\sum_{d=1}^{+\infty}(d+\alpha\theta)^2\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2 \leq \mathcal{E}_1(\alpha-1) < +\infty, \\
&\frac{s}{2}\sum_{d=1}^{+\infty}(d+\alpha\theta)^2\|\xi_{d+1} + \nabla\phi(\zeta_{d+1})\|^2 < +\infty.
\end{aligned}$$

(iv).

$$\begin{aligned}
&\sum_{d=1}^{+\infty}(d+\alpha\theta)^2\mathrm{dist}^2(0, \partial\Phi(\zeta_{d+1})) < +\infty, \\
&\min_{1\leq j\leq d}\mathrm{dist}(0, \partial F(x_{j+1})) = \mathcal{O}\Big(\frac{1}{d^3}\Big).
\end{aligned}$$

(v). If $\alpha > 3$, $\sum_{d=1}^{+\infty}(d+1+\alpha(\theta-1))\|\zeta_{d+1} - \zeta_d\|^2 < +\infty$.

Proof. (i). Put $\omega = \alpha - 1$ in Equation 34, we obtain

$$\begin{aligned}
\mathcal{E}_{d+1}(\alpha-1) - \mathcal{E}_d(\alpha-1) &\leq \big(s(d+\alpha\theta)(3-\alpha) + s(2-\alpha)\big)(\Phi(\zeta_{d+1}) - \Phi(\zeta^*)) \\
&- \frac{s(d+\alpha\theta)^2}{2}[s + 2\beta\sqrt{s} - L_\phi(s+\beta\sqrt{s})^2]\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2 \\
&\leq 0,
\end{aligned} \tag{35}$$

as $\eta = 0$, $\alpha \geq 3$ and $s + 2\beta\sqrt{s} \geq L_\phi(s + \beta\sqrt{s})^2$. Next, we obtain that $\{\mathcal{E}_d(\alpha-1)\}_{d\geq 1}$ is a non-increasing positive sequence and thus is bounded. As $\mathcal{E}_d(\alpha-1) \leq \mathcal{E}_1(\alpha-1)$, and from Equation 14, we get

$$\begin{aligned}&\big(s(d+1+\alpha(\theta-1)) + \eta(\beta\sqrt{s}+s)\big)(d+\alpha\theta)(\Phi(\zeta_d) - \Phi(\zeta^*))\\&+\frac{1}{2}\|u_d(\omega)\|^2 + \frac{(\alpha-1)\eta}{2}\|\zeta_d - \zeta^*\|^2 \leq \mathcal{E}_1(\alpha-1).\end{aligned}$$

Also,

$$s(d+1+\alpha(\theta-1))(d+\alpha\theta)(\Phi(\zeta_d) - \Phi(\zeta^*)) \leq \mathcal{E}_1(\alpha-1).$$

And then,

$$(\Phi(\zeta_d) - \Phi(\zeta^*)) \leq \frac{\mathcal{E}_1(\alpha-1)}{s(d+1+\alpha(\theta-1))(d+\alpha\theta)}. \tag{36}$$

(ii). From Equation 35, we have

$$\mathcal{E}_{d+1}(\alpha-1) - \mathcal{E}_d(\alpha-1) + (s(d+\alpha\theta)(\alpha-3)(\Phi(\zeta_{d+1}) - \Phi(\zeta^*)) \leq 0. \tag{37}$$

Summing Equation 37 for $d = 1, 2, \ldots, N$, we obtain

$$\mathcal{E}_{N+1}(\alpha-1) + s(\alpha-3)\sum_{d=1}^{N}(d+\alpha\theta)(\Phi(\zeta_{d+1}) - \Phi(\zeta^*)) \leq \mathcal{E}_1(\alpha-1).$$

Also, we have

$$s(\alpha-3)\sum_{d=1}^{N}(d+\alpha\theta)(\Phi(\zeta_{d+1}) - \Phi(\zeta^*)) \leq \mathcal{E}_1(\alpha-1).$$

As $N \to +\infty$, we obtain

$$(\alpha-3)\sum_{d=1}^{+\infty}(d+\alpha\theta)(\Phi(\zeta_{d+1}) - \Phi(\zeta^*)) \leq \frac{\mathcal{E}_1(\alpha-1)}{s} < +\infty. \tag{38}$$

(iii). Whenever $s + 2\beta\sqrt{s} > L_\phi(s+\beta\sqrt{s})^2$, Equation 35 yields

$$\begin{aligned}&\mathcal{E}_{d+1}(\alpha-1) - \mathcal{E}_d(\alpha-1)\\&+\frac{s(d+\alpha\theta)^2}{2}[s+2\beta\sqrt{s} - L_\phi(s+\beta\sqrt{s})^2]\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2 \leq 0.\end{aligned} \tag{39}$$

Summing Equation 39 for $d = 1, 2, \ldots, N$, we obtain

$$\begin{aligned}&\mathcal{E}_{N+1}(\alpha-1) + \frac{s}{2}[s+2\beta\sqrt{s} - L_\phi(s+\beta\sqrt{s})^2]\sum_{d=1}^{N}(d+\alpha\theta)^2\|\xi_{d+1}+\nabla\phi(\nu_d)\|^2\\&\leq \mathcal{E}_1(\alpha-1).\end{aligned}$$

Also, we have

$$\frac{s}{2}[s+2\beta\sqrt{s}-L_\phi(s+\beta\sqrt{s})^2]\sum_{d=1}^{N}(d+\alpha\theta)^2\|\xi_{d+1}+\nabla\phi(\nu_d)\|^2 \leq \mathcal{E}_1(\alpha-1).$$

As, $N \to +\infty$

$$\sum_{d=1}^{+\infty}(d+\alpha\theta)^2\|\xi_{d+1}+\nabla\phi(\nu_d)\|^2 \leq \frac{2\mathcal{E}_1(\alpha-1)}{s(s+2\beta\sqrt{s}-L_\phi(s+\beta\sqrt{s})^2)} < +\infty. \tag{40}$$

As $\nabla\phi(x)$ is L_ϕ-Lipschitz continuous, and Step iii of IAFBSC implies that

$$\begin{aligned}\|\xi_{d+1}+\nabla\phi(\zeta_{d+1})\|^2 &= \|\xi_{d+1}+\nabla\phi(\nu_d)+\nabla\phi(\zeta_{d+1})-\nabla\phi(\nu_d)\|^2\\ &\leq 2(\|\xi_{d+1}+\nabla\phi(\nu_d)\|^2+\|\nabla\phi(\zeta_{d+1})-\nabla\phi(\nu_d)\|^2)\\ &\leq 2(\|\xi_{d+1}+\nabla\phi(\nu_d)\|^2+L_\phi^2\|\zeta_{d+1}-\nu_d\|^2).\end{aligned} \tag{41}$$

Using step iii of IAFBSC, and as $\gamma = s+\beta\sqrt{s}$, Equation 41 becomes

$$\|\xi_{d+1}+\nabla\phi(\zeta_{d+1})\|^2 \leq 2(1+(s+\beta\sqrt{s})^2L_\phi^2)\|\xi_{d+1}+\nabla\phi(\nu_d)\|^2. \tag{42}$$

From Equation 40 and 42, we deduce

$$\sum_{d=1}^{+\infty}(d+\alpha\theta)^2\|\xi_{d+1}+\nabla\phi(\zeta_{d+1})\|^2 < +\infty. \tag{43}$$

(iv). Since $\xi_{d+1} \in \partial\psi(\zeta_{d+1})$, we have

$$\sum_{d=1}^{+\infty}(d+\alpha\theta)^2\text{dist}^2(0,\partial\Phi(\zeta_{d+1})) \leq \sum_{d=1}^{+\infty}(d+\alpha\theta)^2\|\xi_{d+1}+\nabla\phi(\zeta_{d+1})\|^2 < +\infty. \tag{44}$$

Then,

$$\min_{1\leq j\leq d}\text{dist}(0,\partial F(x_{i+1}))\sum_{j=1}^{d}j^2 \leq \sum_{d=1}^{+\infty}(d+\alpha\theta)^2\text{dist}^2(0,\partial\Phi(\zeta_{d+1})) < +\infty.$$

This yields,

$$\min_{1\leq j\leq d}\text{dist}(0,\partial F(x_{j+1})) = \mathcal{O}\Big(\frac{1}{d^3}\Big). \tag{45}$$

(v). Set $\omega \in (2,\alpha-1)$ if $\alpha > 3$, thus $\eta = \alpha-\omega-1 > 0$. As $3-\alpha+\eta = -\omega+2 < 0$, $2-\alpha+\eta = -\omega+1 < 0$, then there exists $d_0 \geq 1$ such that for any $d \geq d_0$, from Equation 34, we get

$$s(d+\alpha\theta)(3-\alpha+\eta)+s(2-\alpha+\eta)+\eta\beta\sqrt{s} \leq 0, \tag{46}$$

and

$$\frac{s(d+\alpha\theta)^2}{2}\Big[s+2\beta\sqrt{s}-L_\phi(s+\beta\sqrt{s})^2\Big(1+\frac{\eta\beta}{\sqrt{s}(d+\alpha\theta)}\Big)\Big]\geq 0, \tag{47}$$

using $s+2\beta\sqrt{s}>L_\phi(\beta\sqrt{s}+s)^2$. From Equation 46 and 47 and Equation 34, we get

$$\mathcal{E}_{d+1}(\omega)-\mathcal{E}_d(\omega)\leq -\big(\eta(d+1+\alpha(\theta-1))\big)\|\zeta_{d+1}-\zeta_d\|^2, \tag{48}$$

for all $d\geq d_0$. Next, we obtain that $\mathcal{E}_d(\omega)\}_{d\geq 1}$ is bounded for every $\omega\in(2,\alpha-1)$. Summing Equation 48 for $d=1,2,\dots,N$, we obtain

$$\mathcal{E}_{N+1}(\omega)+\eta\sum_{d=1}^{N}(d+1+\alpha(\theta-1))\|\zeta_{d+1}-\zeta_d\|^2\leq\mathcal{E}_1(\omega).$$

Also, we have

$$\eta\sum_{d=1}^{N}(d+1+\alpha(\theta-1))\|\zeta_{d+1}-\zeta_d\|^2\leq\mathcal{E}_1(\omega).$$

As $N\to\infty$, we obtain

$$\sum_{d=1}^{+\infty}(d+1+\alpha(\theta-1))\|\zeta_{d+1}-\zeta_d\|^2\leq\frac{\mathcal{E}_1(\omega)}{\eta}<+\infty. \tag{49}$$

□

Remark 1. *The parameter condition $2\beta\sqrt{s}+s\geq L_\phi(\beta\sqrt{s}+s)^2$ can be rewritten equivalently as*

$$\gamma L_\phi\leq 1+\frac{\beta\sqrt{s}}{\gamma}.$$

As $\gamma=s+\beta\sqrt{s}$, this implies $\beta\sqrt{s}\leq\gamma$, it follows that $\gamma\leq\frac{2}{L_\phi}$. Moreover, when $\beta>0$, we can choose $\gamma\geq\frac{1}{L_\phi}$. In a particular case $\beta=0$, the parameter condition reduces to $\gamma\leq\frac{1}{L_\phi}$, which coincides with the classical step-size requirement for the AFB algorithm (see [5], [12]).

5 Convergence analysis of generated sequence

Lemma 8. *Let $\alpha\geq 3$ and $\theta\in[0,+\infty)$, and let $\{a_d\}_{d\geq 1}$ and $\{b_d\}_{d\geq 1}$ be two sequences in $[0,+\infty)$ such that*

$$a_{d+1}\leq\frac{d+\alpha(\theta-1)}{d+\alpha\theta}a_d+b_d,$$

for all $d\geq 1$. If $\sum_{d=1}^{+\infty}db_d<+\infty$ then $\sum_{d=1}^{+\infty}a_d<+\infty$.

Proof. Since $\alpha \geq 3, \theta \geq 0$, we have $\alpha(\theta-1) \leq \alpha\theta - 2$. Therefore, we can write

$$a_{d+1} \leq \frac{d+\alpha\theta-2}{d+\alpha\theta} a_d + b_d. \tag{50}$$

Multiplying $(d+1+\alpha\theta)^2$ in Equation 50, we obtain

$$\begin{aligned}(d+1+\alpha\theta)^2 a_{d+1} &\leq \frac{(d+1+\alpha\theta)^2(d+\alpha\theta-2)}{d+\alpha\theta} a_d + (d+1+\alpha\theta)^2 b_d \\ &\leq (d+\alpha\theta)^2 a_d + (d+1+\alpha\theta)^2 b_d. \end{aligned} \tag{51}$$

Summing Equation 51 for $j = 1, 2 \ldots, d-1$, we obtain

$$(d+\alpha\theta)^2 a_d \leq (1+\alpha\theta)^2 a_1 + \sum_{j=1}^{d-1} (j+1+\alpha\theta)^2 b_j.$$

Dividing by $(d+\alpha\theta)^2$, and summing for $d = 2, \ldots, m$, we get

$$\sum_{d=2}^{m} a_d \leq a_1 \sum_{d=2}^{m} \frac{(1+\alpha\theta)^2}{(d+\alpha\theta)^2} + \sum_{d=2}^{m} \frac{1}{(d+\alpha\theta)^2} \sum_{j=1}^{d-1} (j+1+\alpha\theta)^2 b_j. \tag{52}$$

Observing that

$$\sum_{d=j+1}^{+\infty} \frac{1}{(d+\alpha\theta)^2} \leq \sum_{d=j+1}^{+\infty} \frac{1}{d^2} \leq \int_{d=j+1}^{+\infty} \frac{1}{d^2} dd = \frac{1}{j+1}. \tag{53}$$

Applying Fubini's Theorem to this last sum of Equation 52, we obtain

$$\sum_{d=2}^{m} a_d \leq a_1 \sum_{d=2}^{m} \frac{(1+\alpha\theta)^2}{(d+\alpha\theta)^2} + \sum_{j=1}^{m-1} \Big(\sum_{d=j+1}^{+\infty} \frac{1}{(d+\alpha\theta)^2} \Big) (j+1+\alpha\theta)^2 b_j.$$

Since $\sum_{d=1}^{+\infty} d b_d < +\infty$, and from Equation 53, we get

$$\begin{aligned}\sum_{d=2}^{m} a_d &\leq a_1 \sum_{d=2}^{m} \frac{(1+\alpha\theta)^2}{(d+\alpha\theta)^2} + \sum_{j=1}^{m-1} \frac{(j+1+\alpha\theta)^2}{j+1} b_j \\ &\leq a_1 \sum_{d=2}^{m} \frac{(1+\alpha\theta)^2}{(d+\alpha\theta)^2} + \sum_{j=1}^{m-1} (j+1+\alpha\theta(2+\alpha\theta)) b_j < +\infty. \end{aligned} \tag{54}$$

Hence, proved. ∎

Theorem 9. *Suppose that Assumption 1 is satisfied, $\alpha > 3$, and $s+2\beta\sqrt{s} \geq L_\phi(\beta\sqrt{s}+s)^2$. Let $\{(\zeta_d, \nu_d, \xi_d)\}_{d\geq 1}$ be produced by Algorithm 1. Then the successive differences satisfy*

$$\|\zeta_{d+1} - \zeta_d\| = \mathcal{O}\left(\frac{1}{d}\right).$$

Moreover, the sequence $\{\zeta_d\}_{d\geq 1}$ converges, and its limit belongs to the solution set S; that is, there exists $\zeta^* \in S$ such that

$$\zeta_d \to \zeta^* \quad \textit{as } d \to \infty.$$

Proof. From (48), we see that the energy sequence $\{\mathcal{E}_d(\omega)\}_{d\geq 1}$ is bounded for any $\omega \in (2, \alpha-1)$. Together with (14) and the fact that $\eta = \alpha - \omega - 1 > 0$, this implies that the sequence $\{\zeta_d\}_{d\geq 1}$ is bounded. Furthermore, by invoking Theorem 7(iii), we obtain

$$\lim_{d\to\infty} (d+\alpha\theta)\|\xi_{d+1} + \nabla\phi(\nu_d)\| = 0,$$

signifying that $\{(d+\alpha\theta)\beta\sqrt{s}(\xi_{d+1} + \nabla\phi(\nu_d))\}_{d\geq 1}$ is also bounded. Moreover, as $\{\mathcal{E}_d(\alpha - 1)\}_{d\geq 1}$ is bounded. From Equation 14 and 15, we obtain

$$\{\omega(\zeta_{d+1} - \zeta^*) + (d+1+\alpha(\theta-1))(\zeta_{d+1} - \zeta_d) + (d+\alpha\theta)\beta\sqrt{s}(\xi_{d+1} + \nabla\phi(\nu_d))\}_{d\geq 1}$$

is bounded. Since the sequences $\{\zeta_d\}_{d\geq 1}$ and $\{(d+\alpha\theta)\beta\sqrt{s}(\xi_{d+1}+\nabla\phi(\nu_d))\}_{d\geq 1}$ are bounded, we obtain the sequence $\{d+1+\alpha(\theta-1))(\zeta_{d+1} - \zeta_d)\}_{d\geq 1}$ is bounded. Hence,

$$\|\zeta_{d+1} - \zeta_d\| = \mathcal{O}\left(\frac{1}{d}\right). \tag{55}$$

By the lower semicontinuity of Φ, for any sequential cluster point ζ^∞ of the sequence $\{\zeta_d\}_{d\geq 1}$, that is, for any subsequence $\{\zeta_{d_j}\}$ satisfying $\zeta_{d_j} \to \zeta^\infty$ as $d_j \to \infty$, we have

$$\Phi(\zeta^\infty) \leq \liminf_{d_j\to\infty} \Phi(\zeta_{d_j}) = \min_w \Phi,$$

where Theorem 7(i) yields the last equality. Consequently, every cluster point of $\{\zeta_d\}_{d\geq 1}$ belongs to the solution set S. Next, fix an arbitrary $\zeta^* \in S$. We now show that the sequence $\{\|\zeta_d - \zeta^*\|\}_{d\geq 1}$ converges. Denote the another sequence

$$h_d := \frac{1}{2}\|\zeta_d - \zeta^*\|^2. \tag{56}$$

Utilizing Equation 8 and step i of IAFBSC, we get

$$\begin{aligned} h_{d+1} - h_d &= \langle \zeta_{d+1} - \zeta^*, \zeta_{d+1} - \zeta_d\rangle - \frac{1}{2}\|\zeta_{d+1} - \zeta_d\|^2 \\ &= \langle \zeta_{d+1} - \zeta^*, \zeta_{d+1} - \nu_d, \rangle + \frac{d+\alpha(\theta-1)}{d+\alpha\theta}\langle \zeta_{d+1} - \zeta^*, \zeta_d - \zeta_{d-1}\rangle \\ &\quad + \frac{(d+\alpha\theta-1)\beta\sqrt{s}}{d+\alpha\theta}\langle \zeta_{d+1} - \zeta^*, \xi_d + \nabla\phi(\nu_{d-1})\rangle - \frac{1}{2}\|\zeta_{d+1} - \zeta_d\|^2, \end{aligned} \tag{57}$$

and

$$h_d - h_{d-1} = \langle \zeta_d - \zeta^*, \zeta_d - \zeta_{d-1} \rangle - \frac{1}{2}\|\zeta_d - \zeta_{d-1}\|^2. \tag{58}$$

Then, we have

$$\begin{aligned}
&(h_{d+1} - h_d) - \frac{d+\alpha(\theta-1)}{d+\alpha\theta}(h_d - h_{d-1}) \\
&= \langle \zeta_{d+1} - \zeta^*, \zeta_{d+1} - \nu_d, \rangle + \frac{d+\alpha(\theta-1)}{d+\alpha\theta}\langle \zeta_{d+1} - \zeta^*, \zeta_d - \zeta_{d-1} \rangle \\
&+ \frac{(d+\alpha\theta-1)\beta\sqrt{s}}{d+\alpha\theta}\langle \zeta_{d+1} - \zeta^*, \xi_d + \nabla\phi(\nu_{d-1}) \rangle - \frac{1}{2}\|\zeta_{d+1} - \zeta_d\|^2 \\
&- \frac{d+\alpha(\theta-1)}{d+\alpha\theta}\langle \zeta_d - \zeta^*, \zeta_d - \zeta_{d-1} \rangle + \frac{d+\alpha(\theta-1)}{2(d+\alpha\theta)}\|\zeta_d - \zeta_{d-1}\|^2.
\end{aligned} \tag{59}$$

From the step iii of IAFBSC, Equation 59 becomes

$$\begin{aligned}
&(h_{d+1} - h_d) - \frac{d+\alpha(\theta-1)}{d+\alpha\theta}(h_d - h_{d-1}) \\
&= -\gamma\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d) \rangle + \frac{d+\alpha(\theta-1)}{d+\alpha\theta}\langle \zeta_{d+1} - \zeta_d, \zeta_d - \zeta_{d-1} \rangle \\
&+ \frac{(d+\alpha\theta-1)\beta\sqrt{s}}{d+\alpha\theta}\langle \zeta_{d+1} - \zeta^*, \xi_d + \nabla\phi(\nu_{d-1}) \rangle - \frac{1}{2}\|\zeta_{d+1} - \zeta_d\|^2 \\
&+ \frac{d+\alpha(\theta-1)}{2(d+\alpha\theta)}\|\zeta_d - \zeta_{d-1}\|^2 \\
&\leq -\gamma\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d) \rangle + \frac{d+\alpha(\theta-1)}{d+\alpha\theta}\langle \zeta_{d+1} - \zeta_d, \zeta_d - \zeta_{d-1} \rangle \\
&+ \frac{(d+\alpha\theta-1)\beta\sqrt{s}}{d+\alpha\theta}\langle \zeta_{d+1} - \zeta^*, \xi_d + \nabla\phi(\nu_{d-1}) \rangle + \frac{d+\alpha(\theta-1)}{2(d+\alpha\theta)}\|\zeta_d - \zeta_{d-1}\|^2.
\end{aligned} \tag{60}$$

Denote

$$\Delta_{d+1} := h_{d+1} - h_d + \beta\sqrt{s}\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d) \rangle. \tag{61}$$

Using Equation 60 and $\gamma = \beta\sqrt{s} + s$, we have

$$\begin{aligned}
&\Delta_{d+1} - \frac{d+\alpha(\theta-1)}{d+\alpha\theta}\Delta_d \\
&= h_{d+1} - h_d + \beta\sqrt{s}\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \\
&- \frac{d+\alpha(\theta-1)}{d+\alpha\theta}(h_d - h_{d-1}) - \frac{d+\alpha(\theta-1)}{d+\alpha\theta}\beta\sqrt{s}\langle \zeta_d - \zeta^*, \xi_d + \nabla\phi(\nu_{d-1})\rangle \\
&\leq -\gamma\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d)\rangle + \frac{d+\alpha(\theta-1)}{d+\alpha\theta}\langle \zeta_{d+1} - \zeta_d, \zeta_d - \zeta_{d-1}\rangle \\
&+ \frac{(d+\alpha\theta-1)\beta\sqrt{s}}{d+\alpha\theta}\langle \zeta_{d+1} - \zeta^*, \xi_d + \nabla\phi(\nu_{d-1})\rangle + \frac{(d+\alpha(\theta-1))}{2(d+\alpha\theta)}\|\zeta_d - \zeta_{d-1}\|^2 \\
&+ \beta\sqrt{s}\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \\
&- \frac{(d+\alpha(\theta-1))\beta\sqrt{s}}{d+\alpha\theta}\langle \zeta_d - \zeta^*, \xi_d + \nabla\phi(\nu_{d-1})\rangle \\
&= -s\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d)\rangle + \frac{d+\alpha(\theta-1)}{d+\alpha\theta}\langle \zeta_{d+1} - \zeta_d, \zeta_d - \zeta_{d-1}\rangle \\
&+ \frac{(d+\alpha\theta-1)\beta\sqrt{s}}{d+\alpha\theta}\langle \zeta_{d+1} - \zeta_d, \xi_d + \nabla\phi(\nu_{d-1})\rangle + \frac{d+\alpha(\theta-1)}{2(d+\alpha\theta)}\|\zeta_d - \zeta_{d-1}\|^2 \\
&+ \frac{\beta\sqrt{s}}{d+\alpha\theta}\big((d+\alpha\theta-1) - (d+\alpha(\theta-1))\big)\langle \zeta_d - \zeta^*, \xi_d + \nabla\phi(\nu_{d-1})\rangle \\
&\leq -s\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d)\rangle + \|\zeta_{d+1} - \zeta_d\|\|\zeta_d - \zeta_{d-1}\| + \frac{1}{2}\|\zeta_d - \zeta_{d-1}\|^2 \\
&+ \beta\sqrt{s}\|\zeta_{d+1} - \zeta_d\|\|\xi_d + \nabla\phi(\nu_{d-1})\| + \frac{(\alpha-1)\beta\sqrt{s}}{k}\|\zeta_d - \zeta^*\|\|\xi_d + \nabla\phi(\nu_{d-1})\|. \qquad (62)
\end{aligned}$$

The last inequality comes from the Cauchy–Schwarz inequality. Since ψ is convex, its subdifferential mapping $\partial\psi$ is monotone. Furthermore, because

$$0 \in \partial\Phi(\zeta^*) = \nabla\phi(\zeta^*) + \partial\psi(\zeta^*),$$

it follows that

$$-\nabla\phi(\zeta^*) \in \partial\psi(\zeta^*).$$

Using this and fact $\xi_{d+1} \in \partial\psi(\zeta_{d+1})$, we obtain

$$\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\zeta^*)\rangle \geq 0. \qquad (63)$$

Using Equation 63, we can establish,

$$\begin{aligned}
&\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d)\rangle = \langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\zeta^*) + \nabla\phi(\nu_d) - \nabla\phi(x*)\rangle \\
&\geq \langle \zeta_{d+1} - \zeta^*, \nabla\phi(\nu_d) - \nabla\phi(x*)\rangle \\
&= \langle \zeta_{d+1} - \nu_d, \nabla\phi(\nu_d) - \nabla\phi(x*)\rangle + \langle \nu_d - \zeta^*, \nabla\phi(\nu_d) - \nabla\phi(x*)\rangle. \qquad (64)
\end{aligned}$$

Using Lemma 4 and Cauchy-Schwarz inequality, Equation 64 becomes

$$\begin{aligned}
&\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \\
&\geq -\|\zeta_{d+1} - \nu_d\| \|\nabla\phi(\nu_d) - \nabla\phi(\zeta^*)\| + \frac{1}{L_\phi}\|\nabla\phi(\nu_d) - \nabla\phi(\zeta^*)\|^2.
\end{aligned} \tag{65}$$

We have

$$\|x\|\|y\| \leq \frac{1}{2}(\|x\|^2 + \|y\|^2). \tag{66}$$

Put $x = \frac{\|\nabla\phi(\nu_d) - \nabla\phi(\zeta^*)\|}{\sqrt{\frac{L_\phi}{2}}}$ and $y = \sqrt{\frac{L_\phi}{2}}\|\zeta_{d+1} - \nu_d\|$ in Equation 66, we obtain

$$\begin{aligned}
&\frac{1}{L_\phi}\|\nabla\phi(\nu_d) - \nabla\phi(\zeta^*)\|^2 - \|\zeta_{d+1} - \nu_d\|\|\nabla\phi(\nu_d) - \nabla\phi(\zeta^*)\| \\
&\geq \frac{-L_\phi}{4}\|\zeta_{d+1} - \nu_d\|^2.
\end{aligned} \tag{67}$$

Equation 65 becomes

$$\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \geq \frac{-L_\phi}{4}\|\zeta_{d+1} - \nu_d\|^2.$$

Since $\zeta_{d+1} - \nu_d = \gamma(\xi_{d+1} + \nabla\phi(\nu_d))$, from above equation we obtain

$$\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \geq \frac{-L_\phi\gamma^2}{4}\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2. \tag{68}$$

As $\gamma = s + \beta\sqrt{s}$, Equation 68 becomes

$$\langle \zeta_{d+1} - \zeta^*, \xi_{d+1} + \nabla\phi(\nu_d)\rangle \geq \frac{-L_\phi(s + \beta\sqrt{s})^2}{4}\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2. \tag{69}$$

From Equation 69 and 66, Equation 62 becomes

$$\Delta_{d+1} - \frac{d + \alpha(\theta - 1)}{d + \alpha\theta}\Delta_d \leq e_d, \tag{70}$$

where

$$\begin{aligned}
e_d = &\frac{sL_\phi(s + \beta\sqrt{s})^2}{4}\|\xi_{d+1} + \nabla\phi(\nu_d)\|^2 + \frac{1 + \beta\sqrt{s}}{2}\|\zeta_{d+1} - \zeta_d\|^2 \\
&+ \|\zeta_d - \zeta_{d-1}\|^2 + \frac{\beta\sqrt{s}}{2}\|\xi_d + \nabla\phi(\nu_{d-1})\|^2 \\
&+ \frac{(\alpha - 1)\beta\sqrt{s}}{k}\|\zeta_d - \zeta^*\|\|\xi_d + \nabla\phi(\nu_{d-1})\|.
\end{aligned}$$

From Theorem 7 (iii) and (v), we obtain

$$\sum_{d=1}^{+\infty} d^2\|\xi_{d+1}+\nabla\phi(\nu_d)\|^2 < +\infty \quad \text{and} \quad \sum_{d=1}^{+\infty} d\|\zeta_{d+1}-\zeta_d\|^2 < +\infty,$$

and combined with the boundedness of $\{\zeta_d\}_{d\geq 1}$, we obtain

$$\begin{aligned}\sum_{d=1}^{+\infty}\|\zeta_d-\zeta^*\|\|\xi_d+\nabla\phi(\nu_{d-1})\| &\leq \sum_{d=1}^{+\infty}\Big(\frac{\|\zeta_d-\zeta^*\|^2}{2d^2}+\frac{d^2}{2}\|\xi_d+\nabla\phi(\nu_{d-1})\|^2\Big)\\ &< +\infty,\end{aligned} \tag{71}$$

as $\{\|\zeta_d-\zeta^*\|^2\}_{d\geq 1}$ is a bounded sequence. Since

$$\sum_{d=1}^{+\infty} d\|\xi_{d+1}+\nabla\phi(\nu_d)\|^2 \leq \sum_{d=1}^{+\infty} d^2\|\xi_{d+1}+\nabla\phi(\nu_d)\|^2 < +\infty,$$

we get $\sum_{d=1}^{+\infty} de_d < +\infty$. From Equation 70, we can further derive

$$[\Delta_{d+1}]_+ \leq \frac{d+\alpha(\theta-1)}{d+\alpha\theta}[\Delta_d]_+ + e_d. \tag{72}$$

Applying Lemma 8, we get

$$\sum_{d=1}^{+\infty}[\Delta_d]_+ < +\infty.$$

Along with Equation 61 and 71, this implies

$$\begin{aligned}\sum_{d=1}^{+\infty}[h_{d+1}-h_d]_+ &\leq \sum_{d=1}^{+\infty}[\Delta_{d+1}]_+ + \sum_{d=1}^{+\infty}\beta\sqrt{s}\|\zeta_{d+1}-\zeta^*\|\|\xi_{d+1}+\nabla\phi(\nu_d)\|\\ &< +\infty.\end{aligned} \tag{73}$$

Applying Lemma 2, with

$$h_{d+1} \leq h_d + [h_{d+1}-h_d]_+,$$

we conclude that $\lim_{d\to+\infty} h_d = \frac{1}{2}\lim_{d\to+\infty}\|\zeta_d-\zeta^*\|^2$ exists. Since each sequential cluster point of $\{\zeta_d\}_{d\geq 1} \in S$, by Lemma 3, we establish that $\lim_{d\to+\infty}\zeta_d = \zeta^* \in S$. □

Theorem 10. *Suppose Assumption 1 hold. For $\alpha > 3$, and $s+2\beta\sqrt{s} > L_\phi(s+\beta\sqrt{s})^2$. Let $\{(\zeta_d,\nu_d,\xi_d)\}_{d\geq 1}$ be the sequence generated by IAFBSC. Then, the following hold:*

(i). $\Phi(\zeta_d)-\Phi(\zeta^*) = o\big(\frac{1}{d^2}\big)$.

(ii). $\|\zeta_{d+1}-\zeta_d\| = o\big(\frac{1}{d}\big)$.

Proof. By Theorem 9, the sequence $\{\zeta_d\}_{d\geq 1}$ converges to an element of S. Hence, we denote its limit by

$$\lim_{d\to+\infty} \zeta_d = \zeta^* \in S.$$

Moreover, combining Equation 14 with the proof of Theorem 7 yields

$$\mathcal{E}_d(\alpha-1) \geq 0,$$

and

$$\mathcal{E}_{d+1}(\alpha-1) \leq \mathcal{E}_d(\alpha-1),\ \forall d \geq 1.$$

This yields $\lim_{d\to+\infty} \mathcal{E}_d(\alpha-1)$ exists. From Theorem 7 (iii) and Theorem 9, we obtain

$$\lim_{d\to+\infty} (d+\alpha\theta)(\xi_{d+1} + \nabla\phi(\nu_d)) = 0 \ \text{ and } \ \sup_{d\geq 1} d\|\zeta_{d+1} - \zeta_d\| < +\infty.$$

This together with $\zeta_d \to \zeta^*$ and Equation 14 implies

$$\begin{aligned}\lim_{d\to+\infty} \mathcal{E}_d(\alpha-1) = \lim_{d\to+\infty} \Big[& s(d+1+\alpha(\theta-1))(d+\alpha\theta)(\Phi(\zeta_d) - \Phi(\zeta^*)) \\ & + \frac{(d+\alpha(\theta-1))^2}{2}\|\zeta_d - \zeta_{d-1}\|^2 \Big],\end{aligned}$$

exits. We set

$$\begin{aligned}&\lim_{d\to+\infty} \Big[s(d+\alpha(\theta-1))(d+\alpha\theta)(\Phi(\zeta_d) - \Phi(\zeta^*)) + \frac{(d+\alpha(\theta-1))^2}{2}\|\zeta_d - \zeta_{d-1}\|^2 \Big] \\ &= k \geq 0. \end{aligned} \tag{74}$$

Thus, our goal is to show that $k = 0$. If the contrary is true, if $k > 0$, there exists $d_0 \geq 1$ such that

$$\begin{aligned}&s(d+\alpha(\theta-1))(d+\alpha\theta)(\Phi(\zeta_d) - \Phi(\zeta^*)) + \frac{(d+\alpha(\theta-1))^2}{2}\|\zeta_d - \zeta_{d-1}\|^2 \\ &\geq \frac{k}{2} > 0,\end{aligned}$$

for all $d \geq d_0$. This conditions yields

$$\begin{aligned}&\sum_{d=d_0}^{+\infty} \Big[s(d+\alpha\theta)(\Phi(\zeta_d) - \Phi(\zeta^*)) + \frac{(d+1+\alpha(\theta-1))}{2}\|\zeta_d - \zeta_{d-1}\|^2 \Big] \\ &= \sum_{d=d_0}^{+\infty} \frac{1}{(d+\alpha(\theta-1))} \Big[s(d+1+\alpha(\theta-1))(d+\alpha\theta)(\Phi(\zeta_d) - \Phi(\zeta^*)) \\ &+ \frac{(d+1+\alpha(\theta-1))^2}{2}\|\zeta_d - \zeta_{d-1}\|^2 \Big] \\ &\geq \frac{k}{2} \sum_{d=d_0}^{+\infty} \frac{1}{(d+\alpha(\theta-1))} = +\infty.\end{aligned}$$

Subsequently, using Theorem 7 (ii) and (v), we yield

$$\sum_{d=1}^{+\infty}\left[s(d+\alpha\theta)(\Phi(\zeta_d)-\Phi(\zeta^*))+\frac{(d+1+\alpha(\theta-1))}{2}\|\zeta_d-\zeta_{d-1}\|^2\right]<+\infty.$$

This provides an obvious contradiction. As a result, $k=0$, and Equation 74 produces the required results. $\square$

6 Numerical experiments

In this segment, we present two numerical experiments to assess the effectiveness of the AFB method (Algorithm 1). For all algorithms, the parameters are selected to satisfy the assumptions required by the theoretical convergence analysis. All experiments are implemented in MATLAB R2025a and executed on a desktop computer equipped with an Intel Core i7 processor operating at 3.40 GHz and 10 GB of RAM.

6.1 Lasso problem

The LASSO problem is present in the following

$$\min_{w\in\mathbb{R}^d}\Phi(w)=\frac{1}{2}\|Aw-b\|^2+\mu\|w\|_1, \tag{75}$$

where $A\in\mathbb{R}^{m\times d}$, $b\in\mathbb{R}^m$, and $m\le d$. In our numerical experiments, we fix $\lambda=1$, $m=300$, and $d=800$. The sensing matrix A is generated with i.i.d. entries drawn from a standard Gaussian distribution. The ground-truth signal w is constructed to be sparse by randomly selecting a prescribed number of nonzero entries, whose values are sampled independently from a standard Gaussian distribution. The observation vector is then generated as $b=Aw$.

Rewriting Equation 75 in the form of Problem 1, we define

$$\phi(w)=\frac{1}{2}\|Aw-b\|^2,\qquad \psi(w)=\lambda\|w\|_1.$$

Accordingly, the gradient of the smooth term is given by $\nabla\phi(w)=A^T(Aw-b)$, and the proximal operator associated with ψ admits the explicit soft-thresholding form

$$\mathrm{Prox}_{\gamma\psi}(w)=\max\{w-\gamma\lambda,0\}-\max\{-w-\gamma\lambda,0\}.$$

For the IAFBSC, the parameters are chosen as $\gamma=\frac{1+\rho}{\|A^TA\|}$, $\rho=0.2$. When $\beta=0$, the proposed method reduces to the FISTA; in this case, we set $\gamma=\frac{1}{\|A^TA\|}$, as commonly adopted in [5, 16].

Table 1: Iterations and CPU time for different values of α and θ

α	θ											
	0		5		10		20		30		40	
	iteration	time	iteration	time	iteration	time	iteration	time	iteration	time	iteration	time
5	800	1.3449	800	1.4873	800	1.2866	800	1.2455	800	1.2908	800	1.4397
8	800	1.3729	800	1.2658	799	1.2068	800	1.2289	800	1.2261	800	1.1789
10	787	1.1806	789	1.2756	760	1.2739	793	1.2208	800	1.2131	800	1.1682
20	797	1.1519	715	1.1048	678	1.0453	677	1.0583	727	1.1857	759	1.2376
30	791	1.2245	686	1.0471	647	1.0163	664	1.0189	712	1.0650	758	1.1331
40	790	1.1624	650	0.9912	626	0.9525	651	0.9970	712	1.0605	743	1.0861
50	798	1.2079	638	1.0005	603	0.9348	637	0.9745	697	1.0284	758	1.1207
60	800	1.2152	638	0.9789	593	0.9181	637	0.9708	682	1.0331	743	1.1239
70	800	1.1440	635	0.9563	567	0.8918	623	0.9689	697	1.0290	728	1.0824
80	800	1.1443	631	0.9537	570	0.8756	624	0.9588	668	1.0184	728	1.0806
90	800	1.1937	625	0.9563	**558**	0.8831	609	0.9092	682	1.0621	728	1.0878
100	800	1.1303	619	0.9442	560	0.8615	609	0.9281	668	1.0221	728	1.0866

To investigate the influence of algorithmic parameters on the performance of IAFBSC, we consider an example with $m = 300$ and $d = 800$, and examine IAFBSC under various choices of α and θ. Specifically, we test $\alpha \in \{5, 8, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$, $\theta \in \{0, 5, 10, 20, 30, 40\}$. The results reported in Table 1 are averaged over 60 randomly generated instances. As shown in Table 1, Algorithm 1 with $\theta > 0$ consistently outperforms when $\theta = 0$, indicating that the parameter θ has a positive impact on the algorithm's performance in many scenarios. Moreover, among all tested parameter combinations, the choice $\alpha = 90$ and $\theta = 10$ achieves the best overall performance.

To demonstrate the efficiency of IAFBSC, we analyze it with FISTA [12] and AFBSC [20]. In these experiments, we set $\alpha = 90$ and $\theta = 10$ for IAFBSC, and $\alpha = 90$ for AFBSC. The numerical results are shown in Fig. 1. In the first row of Fig. 1, we plot $\|\zeta_d - \zeta^*\|$ and $|\Phi(\zeta_d) - \Phi(\zeta^*)|$ versus the iteration number, where ζ^* denotes the approximate solution obtained upon termination of each method. In the second row, we present the evolution of $\mathrm{dist}(0, \partial\Phi(\zeta_d))$ and $d^2|\Phi(\zeta_d) - \Phi(\zeta^*)|$ with respect to the iteration index. As can be observed from Fig. 1, IAFBSC exhibits superior performance compared with the other algorithms, achieving faster convergence in both solution accuracy and objective value reduction.

6.2 ℓ_1 regularized logistic regression problem

We consider the ℓ_1-regularized logistic regression problem

$$\Phi_{\min} := \min_{\tilde{w}\in\mathbb{R}^d,\, w_0\in\mathbb{R}} \sum_{j=1}^{m} \log\big(1 + \exp\big(-b_i(a_i^T\tilde{w} + w_0)\big)\big) + \mu\|\tilde{w}\|_1, \tag{76}$$

where $a_j \in \mathbb{R}^n$, $b_j \in \{-1, 1\}$ for $j = 1, 2, \ldots, m$, with the labels b_j not all identical, $m < d$, and $\mu > 0$ is the regularization parameter.

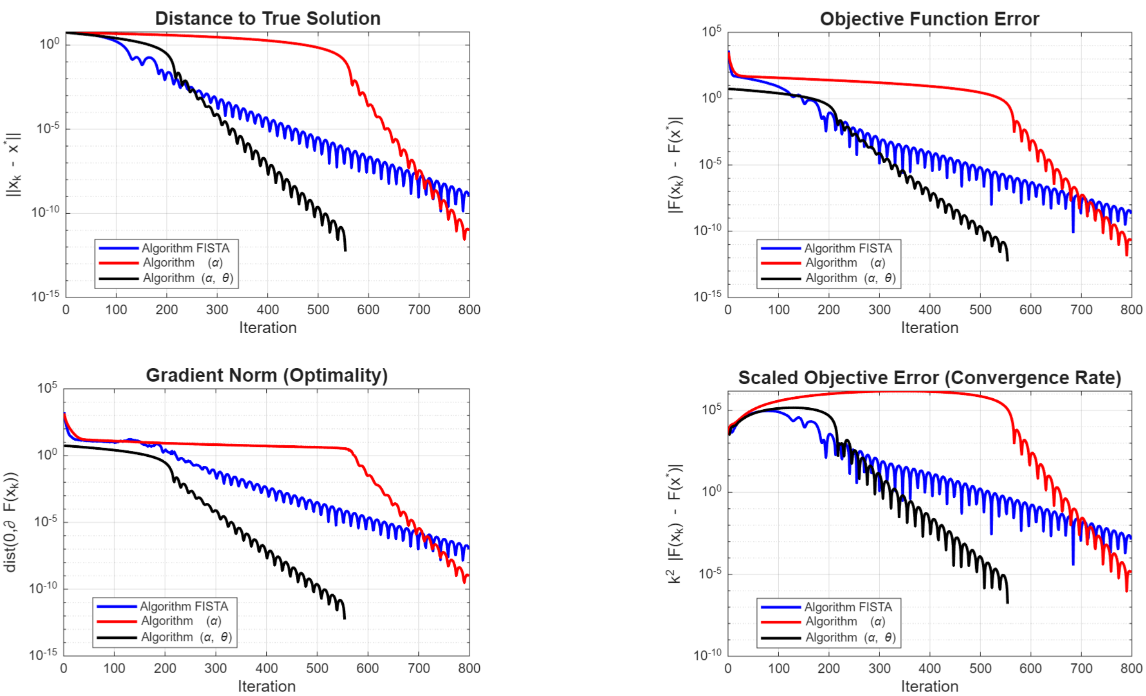

Figure 1: Lasso problem: Evaluations of $\|\zeta_d - \zeta^*\|$, $\|\Phi(\zeta_d) - \Phi(\zeta^*)\|$, $\operatorname{dist}(0, \partial\Phi(\zeta_d))$ and $d^2\|\Phi(\zeta_d) - \Phi(\zeta^*)\|$ with respect to the number of iterations

To cast this problem into the form of Problem 1, we define

$$\phi(w) = \sum_{j=1}^{m} \log\big(1 + \exp\big(- b_j (Dw)_j\big)\big) \,, \qquad \psi(w) = \mu \|\tilde{w}\|_1,$$

where $w = (\tilde{w}, w_0) \in \mathbb{R}^{d+1}$, and $D \in \mathbb{R}^{m\times(d+1)}$ is the matrix whose jth row is $(a_j^T, 1)$. It is well known that the gradient $\nabla\phi$ is Lipschitz continuous with Lipschitz constant $L_\phi = 0.25\, \lambda_{\max}(D^T D)$. In our experiments, we set $\mu = 5$ and initialize all algorithms at the zero vector. The stopping criterion for all methods follows that in [31].

We generate synthetic data with parameters $(m, d, c) = (800, 8000, 80)$. The matrix $A \in \mathbb{R}^{m\times d}$ is drawn with independent and identically distributed standard Gaussian entries. A support set C of size c is selected uniformly at random, and a c-sparse vector $\hat{w}$ supported on C is constructed, whose nonzero components are sampled independently from a standard Gaussian distribution. The measurement vector is then generated as $b = \mathrm{sign}(A\hat{w} + te)$, where t is drawn uniformly from $[0, 1]$.

The numerical results are reported in Fig. 2. In the first row, we plot the error $\|\zeta_d - \zeta^*\|$ and the objective gap $|\Phi(\zeta_d) - \Phi(\zeta^*)|$ versus the iteration index, where ζ^* denotes the approximate solution obtained at termination. In the second row, we present the evolution of $\mathrm{dist}(0, \partial\Phi(\zeta_d))$ and $d^2|\Phi(\zeta_d) - \Phi(\zeta^*)|$. For this experiment, we set $\alpha = 60$ and $\theta = 5$ for IAFBSC, and $\alpha = 60$ for AFBSC. As shown in Fig. 2, IAFBSC consistently outperforms the other methods, exhibiting a noticeably faster convergence in terms of both solution accuracy and objective value reduction.

7 Conclusions

In this paper, we develop a fast gradient-based algorithm incorporating both HDD and vanishing damping for composite convex optimization Problem 1. We demonstrate that the suggested approach achieves an inverse cubic convergence rate for the squared subdifferential norm and a $\mathcal{O}\big(\frac{1}{d^2}\big)$ convergence rate for the objective value for $\alpha \geq 3$. Additionally, for $\alpha > 3$, we prove that the generated iterates converge and further increase the objective value's convergence rate to $o\left(\frac{1}{d^2}\right)$. Numerical experiments on two representative problems confirm the effectiveness of the proposed algorithm.

References

[1] Apidopoulos, V., Aujol, J.-F., Dossal, C.: Convergence rate of inertial forward–backward algorithm beyond Nesterov's rule. Mathematical Programming 180(1), 137–156 (2020)

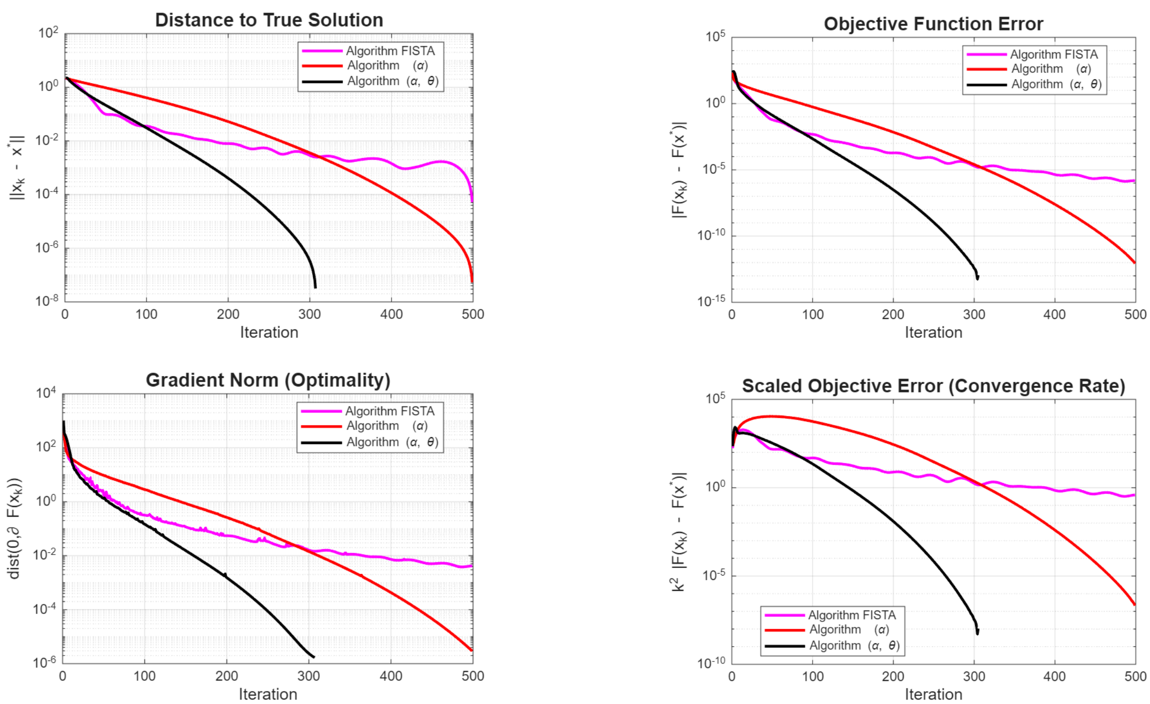

Figure 2: ℓ_1 regularized logistic regression problem: Evaluations of $\|\zeta_d - \zeta^*\|$, $\|\Phi(\zeta_d) - \Phi(\zeta^*)\|$, $\mathrm{dist}(0, \partial\Phi(\zeta_d))$ and $d^2\|\Phi(\zeta_d) - \Phi(\zeta^*)\|$ with respect to the number of iterations

[2] Attouch, H., Peypouquet, J., Redont, P.: Fast convex optimization via inertial dynamics with Hessian driven damping. Journal of Differential Equations 261(10), 5734–5783 (2016)

[3] Attouch, H., Peypouquet, J.: The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$. SIAM Journal on Optimization 26(3), 1824–1834 (2016)

[4] Attouch, H., Cabot, A.: Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity. Journal of Differential Equations 263(9), 5412–5458 (2017)

[5] Attouch, H., Chbani, Z., Peypouquet, J., Redont, P.: Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. Mathematical Programming 168(1), 123–175 (2018)

[6] Attouch, H., Boţ, R.I., Csetnek, E.R.: Fast optimization via inertial dynamics with closed-loop damping. Journal of the European Mathematical Society 25(5), 1985–2056 (2022)

[7] Attouch, H., Chbani, Z., Fadili, J., Riahi, H.: First-order optimization algorithms via inertial systems with Hessian driven damping. Mathematical Programming 193(1), 113–155 (2022)

[8] Attouch, H., Balhag, A., Chbani, Z., Riahi, H.: Accelerated gradient methods combining Tikhonov regularization with geometric damping driven by the Hessian. Applied Mathematics & Optimization 88(2), 29 (2023)

[9] Attouch, H., Chbani, Z., Fadili, J., Riahi, H.: Convergence of iterates for first-order optimization algorithms with inertia and Hessian driven damping. Optimization 72(5), 1199–1238 (2023)

[10] Aujol, J.-F., Dossal, C., Rondepierre, A.: FISTA is an automatic geometrically optimized algorithm for strongly convex functions. Mathematical Programming 204(1), 449–491 (2024)

[11] Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer (2017)

[12] Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences 2(1), 183–202 (2009)

[13] Boţ, R.I., Csetnek, E.R., László, S.C.: Tikhonov regularization of a second order dynamical system with Hessian driven damping. Mathematical Programming 189(1), 151–186 (2021)

[14] Boţ, R.I., Csetnek, E.R., Nguyen, D.-K.: Fast augmented Lagrangian method in the convex regime. Mathematical Programming 200(1), 147–197 (2023)

[15] Boţ, R.I., Csetnek, E.R., Nguyen, D.-K.: Fast optimistic gradient descent ascent method. Foundations of Computational Mathematics 25(1), 163–222 (2025)

[16] Chambolle, A., Dossal, C.: On the convergence of FISTA. Journal of Optimization Theory and Applications 166(3), 968–982 (2015)

[17] Chen, L., Luo, H.: First order optimization methods based on Hessian-driven Nesterov accelerated gradient flow. arXiv:1912.09276 (2019)

[18] He, X., Hu, R., Fang, Y.-P.: Inertial accelerated primal-dual methods. Numerical Algorithms 90(4), 1669–1690 (2022)

[19] He, X., Hu, R., Fang, Y.: A second order primal–dual dynamical system. Applied Mathematics & Optimization 89(2), 30 (2024)

[20] He, X., Fang, Y.: Accelerated forward-backward algorithms with subgradient corrections. Computational Optimization and Applications (2025)

[21] He, X., Huang, N.-J., Fang, Y.-P.: Non-ergodic convergence rate of an inertial accelerated primal–dual algorithm. Communications in Nonlinear Science and Numerical Simulation 140, 108289 (2025)

[22] Jordan, M.I.: Dynamical, symplectic and stochastic perspectives on gradient-based optimization. ICM 2018, 523–549 (2018)

[23] Luo, H., Chen, L.: From differential equation solvers to accelerated first-order methods. Mathematical Programming 195(1), 735–781 (2022)

[24] Nesterov, Y.: Introductory Lectures on Convex Optimization. Springer (2013)

[25] Nesterov, Y.: Lectures on Convex Optimization. Springer (2018)

[26] Rockafellar, R.T.: Convex Analysis. Princeton University Press (1970)

[27] Shi, B., Du, S., Su, W., Jordan, M.I.: Acceleration via symplectic discretization. NeurIPS 32 (2019)

[28] Shi, B., Du, S., Jordan, M.I., Su, W.: Understanding acceleration. Mathematical Programming 195(1), 79–148 (2022)

[29] Su, W., Boyd, S., Candès, E.J.: A differential equation for modeling Nesterov's accelerated gradient method. JMLR 17(153), 1–43 (2016)

[30] Wang, T., Cai, X., Li, T.: On the $o(1/k^2)$ convergence rate analysis. Journal of Global Optimization (2025)

[31] Wen, B., Chen, X., Pong, T.K.: Linear convergence of proximal gradient algorithm. SIAM Journal on Optimization 27(1), 124–145 (2017)

[32] Wilson, A.C., Recht, B., Jordan, M.I.: A Lyapunov analysis of accelerated methods. Journal of Machine Learning Research 22(113), 1–34 (2021)

[33] Wibisono, A., Wilson, A.C., Jordan, M.I.: A variational perspective on accelerated methods. PNAS 113(47), E7351–E7358 (2016)