Lyapunov-based Analysis on First Order Method for Composite Strong-Weak Convex Functions*

Milan Barik† Suvendu Ranjan Pattanaik‡

Abstract

The Nesterov's accelerated gradient (NAG) method generalizes the classical gradient descent algorithm by improving the convergence rate from $\mathcal{O}\left(\frac{1}{t}\right)$ to $\mathcal{O}\left(\frac{1}{t^2}\right)$ in convex optimization. This study examines the proximal gradient framework for additively separable composite functions with smooth and non-smooth components. We demonstrate that Nesterov's accelerated proximal gradient (NAPG_α) method attains a convergence rate of $o\left(\frac{1}{t^2}\right)$ when $\alpha > 3$ for strong-weak convex functions. We also present a Lyapunov analysis to establish the rapid convergence of the composite gradient operator when the smooth component is strongly convex and the non-smooth component is weakly convex. Also, we establish the equivalence between Nesterov's accelerated proximal gradient and Ravine accelerated proximal gradient method.

Keywords. *Nestrov accelerated gradient method, Ravine method, Proximal gradient method, Lyapunov analysis, convergence rates, weakly convex function*

MSC codes. *46N10, 90C25, 65B99, 49M37, 65K05, 90C30*

1 Introduction

Let us consider the following optimization problem

$$\min_{w\in\mathbb{R}^n} F(w) = f(w) + g(w), \tag{P}$$

where $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a lower semicontinuous (l.s.c.) proper weakly convex function, and $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable strongly convex function whose gradient satisfies Lipschitz continuity.

Since the 1960s, momentum methods have been used. In 1964, Polyak [28] suggested a heavy-ball (HB) approach that computes the subsequent iterate using the preceding two.

*Submitted to the editors DATE.

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

The initial goal of momentum methods was to accelerate the convergence rate of convex optimization problems. In 1983, Nesterov [26] proposed an accelerated method (NAG) that improves the rate of convergence of $f(\zeta_t) - f(\zeta^*)$ from $\mathcal{O}(\frac{1}{t})$ to $\mathcal{O}(\frac{1}{t^2})$ for smooth convex function f. The l_1-regularized problem: $\min\{f(y) + \lambda\|y\|_1 : y \in \mathbb{R}^n\}$ was solved by the Iterative Shrinkage-Thresholding Algorithm (ISTA) [14] for smooth convex f, which is a modification of the standard gradient procedure. The sequence $(\zeta_t)_{t\in\mathbb{N}}$ generated by ISTA is given by $\zeta_{t+1} = Prox_{\lambda g}(\zeta_t - \lambda\nabla f(\zeta_t))$ for some $\lambda > 0$, and the convergence rate of $F(\zeta_t) - F(\zeta^*)$ are $\mathcal{O}(\frac{1}{t})$ for convex f and $\mathcal{O}\Big((1 - \frac{\beta-\rho}{L})^t\Big)$ for strongly convex f. Beck and Teboulle [9] generalized the NAG method for the non-smooth convex composite functions (FISTA) and improved the rate of convergence from $\mathcal{O}(\frac{1}{t})$ to $\mathcal{O}(\frac{1}{t^2})$ by using Nesterov's acceleration [26] in ISTA. The sequence $(\zeta_t)_{t\in\mathbb{N}}$ generated by FISTA is given by $\zeta_t = Prox_{\frac{g}{L}}(v_t - \frac{1}{L}\nabla f(v_t))$ and then $v_{t+1} = \zeta_t + \frac{c_t - 1}{c_{t+1}}(\zeta_t - \zeta_{t-1})$ for the update of $c_{t+1} = \frac{1+\sqrt{1+4c_t^2}}{2}$.

The proximal gradient method (PGM) has its roots in the proximal point method (PPM), which was introduced by Martinet [25] and later expanded by Rockafellar [29]. Lions and Mercier [23] introduced the forward-backwards scheme (FB), which is written as: $\zeta_{t+1} = Prox_{\gamma g}(\zeta_t - \gamma\nabla f(\zeta_t))$ for $\gamma > 0$. By adding an extrapolation step: $v_t = \zeta_t + \alpha_t(\zeta_t - \zeta_{t-1})$ with $\alpha_t > 0$ in the FB method, it produces an inertial forward-backwards algorithm (IFB) [24] that enhances the convergence rate. The accelerated forward-backward method (AFB) [6] was introduced by taking $\alpha_t = \frac{t-1}{t+\alpha-1}$ in the extrapolation term, and the iterates of the AFB scheme retain their asymptotic characteristics if the $\alpha_t = \frac{t-1}{t+\alpha-1}$ is changed with an equivalent expression $\alpha_t = 1 - \frac{\alpha}{t}$. For $\alpha = 3$, the AFB procedure approximates the original FISTA algorithm.

The extension of the IFB method and the AFB method has more advantages than the FISTA algorithm. To begin with, proving the convergence of the iterates produced by the FISTA algorithm is still an unresolved challenge [12]. Chambolle and Dossal [12] proved the convergence of the iterates generated by a modified FISTA. Attouch and Peypouquet [6] provide a better convergence rate $o(\frac{1}{t^2})$ for AFB rather than $\mathcal{O}(\frac{1}{t^2})$ and have shown the weak convergence of the iterates when $\alpha > 3$. Attouch-Chbani-Peypouquet-Redont [3] generalized the results in [6] by adding a perturbation term. Attouch and Cobot [2] studied the IFB algorithm for a positive α_t and got a constructive rate $F(\zeta_t) - F(\zeta^*) = o\left(\frac{1}{k_t^2}\right)$ and $\zeta_t - \zeta_{t-1} = o\left(\frac{1}{k_t}\right)$ where $k_t = 1 + \sum_{j=t}^{+\infty}\prod_{i=t}^{j}\alpha_i$, and their results also satisfy $\alpha_t = 1 - \frac{\alpha}{t^r}$ for $r \in (0,1)$. Apidopoulos, Aujol, and Dossal [1], and Attouch, Chbani, and Riahi [4], studied the IFB algorithm with $\alpha_t = 1 - \frac{\alpha}{t}$ when $\alpha \in (0,3)$ and got a constructive rate of convergence $\mathcal{O}(\frac{1}{t^{\frac{2\alpha}{3}}})$.

Much work has been done towards the minimization problems involving weakly convex functions; among all, Hoheisel-Laborde-Oberman [20] studied the PPM for weakly convex functions. Bohm and J. Wright [10] develop a variable smoothing algorithm for weakly convex composite functions and prove a complexity $\mathcal{O}(\frac{1}{\epsilon^3})$ to achieve an ϵ-approximation solution. Liao and Zheng [22] propose a proximal descent method that combines the inexact PPM with classical convex bundle techniques and gets non-asymptotic convergence rates in terms of (n, ϵ)-inexact stationary.

In the literature, the accelerated methods for a strongly convex function [[13], [11], [16], [17], [15]] have been proposed, and they attain the rate $\mathcal{O}\big(\min\big((1-\sqrt{q})^t, \frac{1}{t^2}\big)\big)$ for $F(\zeta_t) - F(\zeta^*)$. Ushiyama [30] proposes an accelerated FISTA method for composite strongly convex functions and achieves $\mathcal{O}\big(\min\big((1+\sqrt{2q}+q)^{-t}, \frac{1}{t^2}\big)\big)$.

The primary motivation for this work follows as Attouch and Fadili [5] show the equivalent relation between the NAG algorithm and the Ravine procedure. Furthermore, the fast convergence of the gradient norms to zero is established for both the Ravine method and NAG procedure, assuming that f is convex and smooth. Khanh-Mordukhovich-Phat-Tran [21] develop the inexact PGM, which is applied to weakly convex functions and demonstrates global convergence of the iterates. He and Fang [19] propose an AFB scheme with subgradient correction in a convex setting. They obtain the convergence rate $F(\zeta_t) - F(\zeta^*) = \mathcal{O}(\frac{1}{t^2})$ and $\min_{1\leq p\leq t} \operatorname{dist}^2(0, \partial F(x_{p+1})) = \mathcal{O}(\frac{1}{t^3})$ for $\alpha \geq 3$, and the convergence rate of the objective value gap improved to $o(\frac{1}{t^2})$ for $\alpha > 3$.

Based on the preceding discussions, we conduct a Lyapunov-based analysis to demonstrate the rapid convergence of the gradient to zero for the accelerated method applied to the additive composite function structure with the smooth and non-smooth types, where the smooth component is strongly convex and the non-smooth component is weakly convex. To our understanding, this is the first study using Lyapunov function to provide this type of analysis in this specific setting.

We introduce the Nesterov accelerated proximal gradient method

$$(\mathrm{NAPG}_\alpha) \begin{cases} v_t = \zeta_t + \alpha_t(\zeta_t - \zeta_{t-1}) \\ \zeta_{t+1} = Prox_{\gamma g}(v_t - \gamma \nabla f(v_t)), \end{cases}$$

where $\alpha_t = 1 - \frac{\alpha}{t}$ with $\alpha > 0$ and $\gamma > 0$, when $\alpha = 3$ NAPG$_\alpha$ reduce to FISTA [9].

1.1 Contribution

The primary contributions are as follows, where g is l.s.c. proper weakly convex and f is a $\mathcal{C}^1$ strongly convex function with a Lipschitz continuous gradient function of f:

- We generalize the descent lemma of composite function in convex setting to the said class of functions and study the convergence analysis of NAPG$_\alpha$ method when $\alpha \geq 3$ through Lyapunov analysis, and obtain the convergence rate $F(\zeta_t) - F(\zeta^*) = o(\frac{1}{t^2})$ when $\alpha > 3$.
- We prove that the iterates (ζ_t) induced by NAPG$_\alpha$ converge to some minimum point when $\alpha > 3$.

- We extend the descent lemma for our class of functions and develop a Lyapunov-based analysis to demonstrate that the convergence rate of the gradient type operator produced by the NAPG_α is $o\left(\frac{1}{t}\right)$, particularly in the case where $\alpha \geq 3$.
- We show an equivalence relation between NAPG_α and the Ravine accelerated proximal gradient (RAPG_α), and prove similar kinds of results for RAPG_α as NAPG_α for our class of functions.

1.2 Paper organization

In Section 2, we first generalize the descent lemma for F and then develop a Lyapunov function to show the convergence efficiency of $F(\zeta_t) - F(\zeta^*)$ as $o(\frac{1}{t^2})$. In Section 3, the convergence of the iterates of NAPG_α is proved. In Section 4, we extend the descent lemma and develop a Lyapunov analysis to show that the gradient equivalent converges to zero. Section 5 introduces the Ravine method for F and shows the equivalence relation between NAPG_α and RAPG_α. The sequence produced by RAPG_α converges and its convergence rate of $F(v_t) - F(\zeta^*) = o(\frac{1}{t^2})$. Also, we demonstrate that the rate of convergence of the gradient map of RAPG_α is equivalent to that of NAPG_α.

1.3 Basic definitions and preliminaries

This section goes through the fundamental definitions and notations needed for the analysis that follows, as well as the important characteristics that will be used in the advancements that follow.

Definition 1. *Let f be a $\mathcal{C}^1$ function on $\mathbb{R}^n$, then f is β-strongly convex for some $\beta > 0$ if the following holds*

$$f(v) + \langle \nabla f(v), u - v \rangle + \frac{\beta}{2}\|u - v\|^2 \leq f(u),$$

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

The definition above implies that if $f(\cdot) - \frac{\beta}{2}\|\cdot\|^2$ is convex, then f is β-strongly convex mapping.

Definition 2. *Suppose g is a l.s.c. proper function on $\mathbb{R}^n$ is said be a ρ-weakly convex for some $\rho \geq 0$ if the function $g(\cdot) + \frac{\rho}{2}\|\cdot\|^2$ is convex.*

Definition 3. *Let g be a l.s.c. proper ρ-weakly convex function on $\mathbb{R}^n$. Let $\gamma \in (0, \frac{1}{\rho})$. Then the proximal operator with regard to γg is defined as*

$$Prox_{\gamma g}(w) = \arg\min_{z \in \mathbb{R}^n} \left\{ g(z) + \frac{\gamma}{2}\|z - w\|^2 \right\},$$

for $w \in \mathbb{R}^n$.

Assume the following conditions:

$$(K) \quad \begin{cases} g \text{ is l.s.c. proper } \rho\text{-weakly convex mapping;} \\ f \text{ is differentiable, } \beta\text{-strongly convex function, } \nabla f \text{ is } L\text{-Lipschitz continuous;} \\ \arg\min\limits_{\mathbb{R}^n} F = S \neq \phi; \\ \beta > \rho. \end{cases}$$

Next, we define the gradient mapping for F.

Definition 4. *Suppose f, g are two function on $\mathbb{R}^n$ with f is differentiable convex and g is non smooth ρ-weakly convex for some $\rho > 0$. Let $\gamma \in (0, \frac{1}{\rho})$ then the composite gradient map $P_\gamma : \mathbb{R}^n \to \mathbb{R}^n$ is defined by*

$$P_\gamma(w) = Prox_{\gamma g}(w - \gamma \nabla f(w)), \tag{1}$$

$w \in \mathbb{R}^n$.

Also, the mapping P_γ is written as

$$P_\gamma(w) = \arg\min_{z \in \mathbb{R}^n} \Psi_{\gamma, w - \gamma \nabla f(w)}(z),$$

where $\Psi_{\gamma,w}(z) = g(z) + \frac{1}{2\gamma}\|z - w\|^2 = g(z) + \frac{\rho}{2}\|z - w\|^2 + \frac{\gamma^{-1}-\rho}{2}\|z - w\|^2$ which is a $(\gamma^{-1} - \rho)$-strongly convex function since $g(\cdot) + \frac{\rho}{2}\|\cdot\|^2$ is a convex function. Observe the quadratic estimation of the map Ψ,

$$\begin{aligned} P_\gamma(w) &= \arg\min_{z \in \mathbb{R}^n} \Big\{ g(z) + \frac{1}{2\gamma}\|z - (w - \gamma \nabla f(w))\|^2 \Big\} \\ &= \arg\min_{z \in \mathbb{R}^n} \Big\{ \frac{\gamma}{2}\|\nabla f(w)\|^2 + \langle \nabla f(w), z - w \rangle + \frac{1}{2\gamma}\|z - w\|^2 + g(z) \Big\} \\ &= \arg\min_{z \in \mathbb{R}^n} \Big\{ f(w) + \langle \nabla f(w), z - w \rangle + \frac{1}{2\gamma}\|z - w\|^2 + g(z) \Big\}. \end{aligned}$$

Definition 5. *Suppose f, g are two function on $\mathbb{R}^n$ with f is differentiable convex function and g is non smooth ρ-weakly convex function for some $\rho > 0$. Let $\gamma \in (0, \frac{1}{\rho})$ and P_γ be the composite gradient mapping, then the gradient mapping $G_\gamma : \mathbb{R}^n \to \mathbb{R}^n$ defined by*

$$G_\gamma(w) = \frac{1}{\gamma}(w - P_\gamma(w)), \tag{2}$$

for $w \in \mathbb{R}^n$.

From the above definition of gradient mapping, we can rewrite the NAPG$_\alpha$ method in the following form

$$\begin{cases} v_t = \zeta_t + \left(1 - \dfrac{\alpha}{t}\right)(\zeta_t - \zeta_{t-1}); \\ \zeta_{t+1} = v_t - \gamma G_\gamma(v_t). \end{cases}$$

Remark 1. *Here, we analyze the gradient mapping value at $\zeta^* \in S$, i.e., $G_\gamma(\zeta^*)$.*

$$\begin{aligned} P_\gamma(\zeta^*) &= \arg\min_{w\in\mathbb{R}^n} \left\{ f(\zeta^*) + \langle \nabla f(\zeta^*), w - \zeta^* \rangle + \frac{1}{2\gamma}\|w - \zeta^*\|^2 + g(w) \right\} \\ &= \arg\min_{w\in\mathbb{R}^n} \left\{ f(\zeta^*) + \frac{1}{2\gamma}\|w - \zeta^*\|^2 + g(w) \right\} \\ &= \zeta^*, \end{aligned}$$

as ζ^* is the minimizer of $F = f + g$. Hence $G_\gamma(\zeta^*) = 0$.

Lemma 1. *[7] Let (ζ_t) be a sequence in $\mathbb{R}^n$, and $\phi \neq S \subseteq \mathbb{R}^n$. Assume that*

(i) $\lim_{t\to+\infty} \|\zeta_t - \zeta^*\|$ *exists, for all $\zeta^* \in S$.*

(ii) Every cluster point of ζ_t in S.

Then, as $t \to \infty$, (ζ_t) converges to a point in S.

Lemma 2. *[7] Suppose $(a_t)_{t\in\mathbb{N}}$, $(\beta_t)_{t\in\mathbb{N}}$, $(\gamma_t)_{t\in\mathbb{N}}$ and $(\eta_t)_{t\in\mathbb{N}}$ are positive sequences of real numbers and the series $\sum_{t\in\mathbb{N}} \gamma_t$ and $\sum_{t\in\mathbb{N}} \eta_t$ are convergent. If*

$$a_{t+1} \leq (1 + \gamma_t)a_t - \beta_t + \eta_t,$$

for all $t \in \mathbb{N}$. Then the sequence $(a_t)_{t\in\mathbb{N}}$ and the series $\sum_{t\in\mathbb{N}} \beta_t$ converge.

Theorem 3 (Theorem 2.1.8, [27])**.** *Suppose f is $\mathcal{C}^1$ β-strongly convex function on $\mathbb{R}^n$ and $\nabla f(u^*) = 0$. For all $v \in \mathbb{R}^n$,*

$$f(u^*) + \frac{\beta}{2}\|v - u^*\|^2 \leq f(v) \text{ holds.}$$

2 Lyapunov analysis

Now, we construct an energy (Lyapunov) function to analyze the stability and the convergence of the algorithm. Here, we extended the descent lemma for F.

Lemma 4. *Let assume (K) and $\gamma \in (0, \frac{1}{L+\rho})$. Suppose $\check{x},\ \check{y} \in \mathbb{R}^n$ with $\check{x} = P_\gamma(\check{y})$. Then*

$$F(\check{x}) \leq F(w) + (1 - \rho\gamma)\langle G_\gamma(\check{y}), \check{y} - w \rangle - \frac{\gamma(1 - \rho\gamma)}{2}\|G_\gamma(\check{y})\|^2, \tag{3}$$

holds for all $w \in \mathbb{R}^n$.

Proof. Consider the mapping $w \longmapsto g(w) + f(\check{y}) + \langle \nabla f(\check{y}), w - \check{y}\rangle + \frac{1}{2\gamma}\|w - \check{y}\|^2$, which is a $(\gamma^{-1} - \rho)$ -strongly convex function. Since $\check{x}$ is the minimizer of P_γ, from Theorem 3 we get

$$\begin{aligned} g(\check{x}) + f(\check{y}) + \langle \nabla f(\check{y}), \check{x} - \check{y}\rangle + \frac{1}{2\gamma}\|\check{x} - \check{y}\|^2 + \frac{\gamma^{-1} - \rho}{2}\|w - \check{x}\|^2 \\ \leq g(w) + f(\check{y}) + \langle \nabla f(\check{y}), z - \check{y}\rangle + \frac{1}{2\gamma}\|w - \check{y}\|^2, \end{aligned} \tag{4}$$

for all $w \in \mathbb{R}^n$. From [Lemma 5.7, [8]], we have

$$f(a) \leq f(b) + \langle \nabla f(b), a - b\rangle + \frac{L}{2}\|a - b\|^2, \ \forall a, \ b \in \mathbb{R}^n. \tag{5}$$

Since $L < \frac{1}{\gamma}$, putting $a = \check{x}$ and $b = \check{y}$ in Equation 5, we get

$$f(\check{x}) \leq f(\check{y}) + \langle \nabla f(\check{y}), \check{x} - \check{y}\rangle + \frac{1}{2\gamma}\|\check{x} - \check{y}\|^2.$$

Now applying the above result in Equation 4, we get

$$\begin{aligned} f(\check{x}) + g(\check{x}) + \frac{\gamma^{-1} - \rho}{2}\|w - \check{x}\|^2 \leq g(w) + f(\check{y}) + \langle \nabla f(\check{y}), w - \check{y}\rangle \\ + \frac{\gamma^{-1} - \rho}{2}\|w - \check{y}\|^2 + \frac{\rho}{2}\|w - \check{y}\|^2. \end{aligned} \tag{6}$$

Since $\rho < \beta$, we obtain

$$\begin{aligned} g(\check{x}) + f(\check{x}) + \frac{\gamma^{-1} - \rho}{2}\|w - \check{x}\|^2 \leq g(w) + f(\check{y}) + \langle \nabla f(\check{y}), w - \check{y}\rangle \\ + \frac{\gamma^{-1} - \rho}{2}\|w - \check{y}\|^2 + \frac{\beta}{2}\|w - \check{y}\|^2. \end{aligned}$$

From Definition 1, we get $f(\check{y})) + \langle \nabla f(\check{y}), w - \check{y}\rangle \leq f(w) - \frac{\beta}{2}\|w - \check{y}\|^2$ holds. Then

$$g(\check{x}) + f(\check{x}) + \frac{\gamma^{-1} - \rho}{2}\|w - \check{x}\|^2 \leq g(w) + f(w) + \frac{\gamma^{-1} - \rho}{2}\|w - \check{y}\|^2.$$

As $f + g = F$, we have

$$F(\check{x}) + \frac{\gamma^{-1} - \rho}{2}\|w - \check{x}\|^2 \leq F(w) + \frac{\gamma^{-1} - \rho}{2}\|w - \check{y}\|^2.$$

Also, we have

$$F(\check{x}) \leq F(w) + \frac{\gamma^{-1} - \rho}{2}\left[\|w - \check{y}\|^2 - \|\check{x} - w\|^2\right]. \tag{7}$$

From Definition 4, we obtain

$$\begin{aligned}\left[\|w-\check{y}\|^2-\|\check{x}-w\|^2\right] &= [\|w-\check{y}\|^2-\|P_\gamma(\check{y})-w\|^2] \\ &= \left[\|w-\check{y}\|^2-\|Prox_{\gamma g}(\check{y}-\gamma\nabla f(\check{y}))-w\|^2\right] \\ &= \left[\|w\|^2+\|\check{y}\|^2-2\langle w,\check{y}\rangle-\|Prox_{\gamma g}(\check{y}-\gamma\nabla f(\check{y}))\|^2-\|w\|^2\right. \\ &\left.+\,2\langle w, Prox_{\gamma g}(\check{y}-\gamma\nabla f(\check{y}))\rangle\right] \\ &= \left[\|\check{y}\|^2-\|Prox_{\gamma g}(\check{y}-\gamma\nabla f(\check{y}))\|^2\right. \\ &\left.-\,2\langle w,\check{y}-Prox_{\gamma g}(\check{y}-\gamma\nabla f(\check{y}))\rangle\right]. \end{aligned} \tag{8}$$

From Definition 5 and 4, $Prox_{\gamma g}(\check{y}-\gamma\nabla f(\check{y})) = P_\gamma(\check{y}) = \check{y}-\gamma G_\gamma(\check{y})$, Equation 8 becomes

$$\begin{aligned}\left[\|w-\check{y}\|^2-\|\check{x}-w\|^2\right] &= \left[\|\check{y}\|^2-\|\check{y}-\gamma G_\gamma(\check{y})\|^2-2\langle w,\gamma G_\gamma(\check{y})\rangle\right] \\ &= \left[-\gamma^2\|G_\gamma(\check{y})\|^2+2\langle\gamma G_\gamma(\check{y}),\check{y}-w\rangle\right]. \end{aligned} \tag{9}$$

Putting Equation 9 in Equation 7, we get

$$F(\check{x}) \le F(w)+(1-\rho\gamma)\langle G_\gamma(\check{y}),\check{y}-w\rangle-\frac{\gamma(1-\rho\gamma)}{2}\|G_\gamma(\check{y})\|^2.$$

□

Lemma 5. *Let assume (K), $\gamma \in \left(0,\frac{1}{\rho+L}\right)$ and $\zeta^* \in S$. If $\alpha > 3$, then $\sum_{t=1}^{\infty} t(F(\zeta_t)-F(\zeta^*)) \le \frac{(\alpha-1)^2\mathcal{W}(1)}{(\alpha-3)} < +\infty$.*

Proof. From Lemma 4, for any $a,\ b \in \mathbb{R}^n$ we get

$$F(a-\gamma G_\gamma(a)) \le F(b)+(1-\rho\gamma)\langle G_\gamma(a),a-b\rangle-\frac{\gamma(1-\rho\gamma)}{2}\|G_\gamma(a)\|^2. \tag{10}$$

Putting $a=v_t, b=\zeta_t$ and $a=v_t, b=\zeta^*$ in Equation 10, we obtain

$$F(v_t-\gamma G_\gamma(v_t)) \le F(\zeta_t)+(1-\rho\gamma)\langle G_\gamma(v_t),v_t-\zeta_t\rangle-\frac{\gamma(1-\rho\gamma)}{2}\|G_\gamma(v_t)\|^2, \tag{11}$$

and

$$F(v_t-\gamma G_\gamma(v_t)) \le F(\zeta^*)+(1-\rho\gamma)\langle G_\gamma(v_t),v_t-\zeta^*\rangle-\frac{\gamma(1-\rho\gamma)}{2}\|G_\gamma(v_t)\|^2, \tag{12}$$

respectively. We have $c_t = \frac{t-1}{\alpha-1}$ and $\alpha_t = 1-\frac{\alpha}{t} = \frac{c_t-1}{c_{t+1}}$. Multiplying $(c_{t+1}-1) \ge 0$ in Equation 11 and after that adding with Equation 12, since $\zeta_{t+1} = v_t-\gamma G_\gamma(v_t)$ we obtain

$$\begin{aligned} c_{t+1}F(\zeta_{t+1}) \le (c_{t+1}-1)F(\zeta_t)+F(\zeta^*)-\frac{\gamma c_{t+1}}{2}(1-\rho\gamma)\|G_\gamma(v_t)\|^2 \\ +\,(1-\rho\gamma)\langle G_\gamma(v_t),(c_{t+1}-1)(v_t-\zeta_t)+(v_t-\zeta^*)\rangle. \end{aligned} \tag{13}$$

$$\begin{aligned} v_t + (c_{t+1} - 1)(v_t - \zeta_t) &= \zeta_t + c_{t+1}(v_t - \zeta_t) \\ &= \zeta_t + c_{t+1}\alpha_t(\zeta_t - \zeta_{t-1}). \end{aligned}$$

As $c_{t+1}\alpha_t = c_t - 1$, we get

$$\begin{aligned} v_t + (c_{t+1} - 1)(v_t - \zeta_t) &= \zeta_t + (c_t - 1)(\zeta_t - \zeta_{t-1}) \\ &= \zeta_{t-1} + c_t(\zeta_t - \zeta_{t-1}). \end{aligned} \tag{14}$$

Let $b_t = \zeta_{t-1} + c_t(\zeta_t - \zeta_{t-1})$, then the Equation 13 becomes

$$\begin{aligned} c_{t+1}F(\zeta_{t+1}) \leq (c_{t+1} - 1)F(\zeta_t) + F(\zeta^*) - \frac{\gamma c_{t+1}}{2}(1 - \rho\gamma)\|G_\gamma(v_t)\|^2 \\ + (1 - \rho\gamma)\langle G_\gamma(v_t), b_t - \zeta^*\rangle. \end{aligned} \tag{15}$$

Subtracting $c_{t+1}F(\zeta^*)$ on each side of Equation 15, we get

$$\begin{aligned} c_{t+1}(F(\zeta_{t+1}) - F(\zeta^*)) \leq (c_{t+1} - 1)(F(\zeta_t) - F(\zeta^*)) - \frac{\gamma c_{t+1}}{2}(1 - \rho\gamma)\|G_\gamma(v_t)\|^2 \\ + (1 - \rho\gamma)\langle G_\gamma(v_t), b_t - \zeta^*\rangle. \end{aligned} \tag{16}$$

Now,

$$\begin{aligned} b_{t+1} - b_t &= \zeta_t + c_{t+1}(\zeta_{t+1} - \zeta_t) - \zeta_{t-1} - c_t(\zeta_t - \zeta_{t-1}) \\ &= c_{t+1}(\zeta_{t+1} - \zeta_t) - (c_t - 1)(\zeta_t - \zeta_{t-1}) \\ &= c_{t+1}(\zeta_{t+1} - \zeta_t - \alpha_t(\zeta_t - \zeta_{t-1})) \\ &= c_{t+1}(\zeta_{t+1} - v_t) = -\gamma c_{t+1}G_\gamma(v_t). \end{aligned} \tag{17}$$

Then $b_{t+1} - \zeta^* = b_t - \zeta^* - \gamma c_{t+1}G_\gamma(v_t)$ and we have

$$\|b_{t+1} - \zeta^*\|^2 = \|b_t - \zeta^*\|^2 - 2\gamma c_{t+1}\langle G_\gamma(v_t), b_t - \zeta^*\rangle + \gamma^2 c_{t+1}^2\|G_\gamma(v_t)\|^2. \tag{18}$$

Putting Equation 18 in Equation 16, we obtain

$$\begin{aligned} c_{t+1}(F(\zeta_{t+1}) - F(\zeta^*)) &\leq (c_{t+1} - 1)(F(\zeta_t) - F(\zeta^*)) \\ &\quad + \frac{(1 - \rho\gamma)}{2\gamma c_{t+1}}(\|b_t - \zeta^*\|^2 - \|b_{t+1} - \zeta^*\|^2). \end{aligned}$$

Also,

$$\begin{aligned} c_{t+1}^2(F(\zeta_{t+1}) - F(\zeta^*)) &\leq (c_{t+1}^2 - c_{t+1})(F(\zeta_t) - F(\zeta^*)) \\ &\quad + \frac{(1 - \rho\gamma)}{2\gamma}(\|b_t - \zeta^*\|^2 - \|b_{t+1} - \zeta^*\|^2). \end{aligned}$$

Equivalently,

$$\begin{aligned}
c_{t+1}^2(F(\zeta_{t+1}) - F(\zeta^*)) + \frac{(1-\rho\gamma)}{2\gamma}\|b_{t+1} - \zeta^*\|^2 \\
&\leq c_t^2(F(\zeta_t) - F(\zeta^*)) + \frac{(1-\rho\gamma)}{2\gamma}\|b_t - \zeta^*\|^2 \\
&+ (c_{t+1}^2 - c_{t+1} - c_t^2)(F(\zeta_t) - F(\zeta^*)). \qquad (19)
\end{aligned}$$

Define $\mathcal{W}(t) := c_t^2(F(\zeta_t) - F(\zeta^*)) + \frac{(1-\rho\gamma)}{2\gamma}\|b_t - \zeta^*\|^2$, then Equation 19 becomes

$$\mathcal{W}(t+1) - (c_{t+1}^2 - c_{t+1} - c_t^2)(F(\zeta_t) - F(\zeta^*)) \leq \mathcal{W}(t). \qquad (20)$$

Since $c_{t+1}^2 - c_{t+1} - c_t^2 = -\frac{t(\alpha-3)+1}{(\alpha-1)^2} \leq 0$ for $\alpha \geq 3$, from Equation 20, we obtain

$$\mathcal{W}(t+1) + \frac{t(\alpha-3)+1}{(\alpha-1)^2}(F(\zeta_t) - F(\zeta^*)) \leq \mathcal{W}(t). \qquad (21)$$

Since the sequence $(\mathcal{W}(t))$ is positive and non-increasing, by summing Equation 21 for $t = 1, 2, \cdots, N$, we obtain

$$\mathcal{W}(N+1) + \sum_{t=1}^{N} \frac{t(\alpha-3)+1}{(\alpha-1)^2}(F(\zeta_t) - F(\zeta^*)) \leq \mathcal{W}(1).$$

Also, we get

$$\sum_{t=1}^{N} \frac{t(\alpha-3)+1}{(\alpha-1)^2}(F(\zeta_t) - F(\zeta^*)) \leq \mathcal{W}(1).$$

Then as $N \to \infty$,

$$\sum_{t=1}^{\infty} t(F(\zeta_t) - F(\zeta^*)) \leq \frac{(\alpha-1)^2\mathcal{W}(1)}{(\alpha-3)}.$$

□

Theorem 6. *Let us suppose that (K) holds,* $\gamma \in \left(0, \frac{1}{\rho+L}\right)$ *and* $\zeta^* \in S$*. Let* $(v_t)_{t\in\mathbb{N}}$ *and* $(\zeta_t)_{t\in\mathbb{N}}$ *be the sequences induced by NAPG*$_\alpha$*. When* $\alpha \geq 3$*, then*

(i) $F(\zeta_t) - F(\zeta^*) \leq \frac{(\alpha-1)^2}{t^2}\mathcal{W}(1) = \mathcal{O}(\frac{1}{t^2})$;

(ii) $\|\zeta_t - \zeta_{t-1}\| = \mathcal{O}\left(\frac{1}{t}\right)$.

Proof. From Equation 21, we obtain

$$\mathcal{W}(t) \leq \mathcal{W}(1).$$

The above equation implies that

$$c_t^2(F(\zeta_t) - F(\zeta^*)) + \frac{(1-\rho\gamma)}{2\gamma}\|b_t - \zeta^*\|^2 \leq \mathcal{W}(1), \tag{22}$$

since $c_t = \frac{t-1}{\alpha-1}$ then

$$F(\zeta_t) - F(\zeta^*) \leq \frac{(\alpha-1)^2}{(t-1)^2}\mathcal{W}(1) \leq \frac{(\alpha-1)^2}{t^2}\mathcal{W}(1) = \mathcal{O}\Big(\frac{1}{t^2}\Big).$$

Again, since $b_t = \zeta_{t-1} + c_t(\zeta_t - \zeta_{t-1})$, we obtain

$$\|\zeta_t - \zeta_{t-1}\| = \mathcal{O}\Big(\frac{1}{t}\Big).$$

□

Lemma 7. *Let assume the assumption (K), $\gamma \in \big(0, \frac{1}{\rho+L}\big)$ and $\zeta^* \in S$. If $\alpha > 3$, then*

$$\sum_{t=1}^{\infty} t\|\zeta_{t+1} - \zeta_t\|^2 \leq \frac{\gamma}{1-\rho\gamma}\Big[(\alpha-1)^2\mathcal{W}(1) + \frac{3(\alpha-1)^2\mathcal{W}(1)}{\alpha-3}\Big] < +\infty.$$

Proof. Using $\zeta_{t+1} = v_t - \gamma G_\gamma(v_t)$ and $v_t - \zeta_t = \alpha_t(\zeta_t - \zeta_{t-1})$ in Equation 11, we get

$$\begin{aligned}
F(\zeta_{t+1}) &\leq F(\zeta_t) + (1-\rho\gamma)\langle G_\gamma(v_t), v_t - \zeta_t\rangle - \frac{\gamma(1-\rho\gamma)}{2}\|G_\gamma(v_t)\|^2 \\
&= F(\zeta_t) + \frac{(1-\rho\gamma)}{\gamma}\langle v_t - \zeta_{t+1}, v_t - \zeta_t\rangle - \frac{(1-\rho\gamma)}{2\gamma}\|\zeta_{t+1} - v_t\|^2 \\
&= F(\zeta_t) + \frac{(1-\rho\gamma)}{\gamma}\langle \zeta_t - \zeta_{t+1} + \alpha_t(\zeta_t - \zeta_{t-1}), \alpha_t(\zeta_t - \zeta_{t-1})\rangle \\
&- \frac{(1-\rho\gamma)}{2\gamma}\|\zeta_{t+1} - \zeta_t - \alpha_t(\zeta_t - \zeta_{t-1})\|^2, \\
&\leq F(\zeta_t) + \frac{\alpha_t(1-\rho\gamma)}{\gamma}\langle \zeta_t - \zeta_{t+1}, \zeta_t - \zeta_{t-1}\rangle + \frac{\alpha_t^2(1-\rho\gamma)}{2}\|\zeta_t - \zeta_{t-1}\|^2 \\
&- \frac{(1-\rho\gamma)}{2}\|\zeta_{t+1} - \zeta_t\|^2 - \frac{\alpha_t^2(1-\rho\gamma)}{2\gamma}\|\zeta_t - x_{\gamma-1}\|^2 \\
&+ \frac{\alpha_t(1-\rho\gamma)}{\gamma}\langle \zeta_{t+1} - \zeta_t, \zeta_t - \zeta_{t-1}\rangle.
\end{aligned}$$

It implies that

$$F(\zeta_{t+1}) + \frac{(1-\rho\gamma)}{2}\|\zeta_{t+1} - \zeta_t\|^2 \leq F(\zeta_t) + \frac{\alpha_t^2(1-\rho\gamma)}{2\gamma}\|\zeta_t - \zeta_{t-1}\|^2. \tag{23}$$

Since $\alpha_t \leq \alpha_{t+1}$, then the above equation becomes

$$F(\zeta_{t+1}) + \frac{(1-\rho\gamma)}{2}\|\zeta_{t+1} - \zeta_t\|^2 \leq F(\zeta_t) + \frac{\alpha_{t+1}^2(1-\rho\gamma)}{2s}\|\zeta_t - \zeta_{t-1}\|^2. \tag{24}$$

Subtracting $F(\zeta^*)$ on each side of Equation 24, we get

$$\begin{aligned} F(\zeta_{t+1}) - F(\zeta^*) + \frac{(1-\rho\gamma)}{2s}\|\zeta_{t+1} - \zeta_t\|^2 & \\ \leq F(\zeta_t) - F(\zeta^*) + \frac{\alpha_{t+1}^2(1-\rho\gamma)}{2s}\|\zeta_t - \zeta_{t-1}\|^2. \end{aligned} \tag{25}$$

Let $\xi_t := F(\zeta_t) - F(\zeta^*)$, $\mu_t := \frac{(1-\rho\gamma)}{2s}\|\zeta_{t+1} - \zeta_t\|^2$, then Equation 25 becomes

$$\begin{aligned} \xi_{t+1} + \mu_t &\leq \xi_t + \alpha_{t+1}^2 \mu_{t-1} \\ &= \xi_t + \left(\frac{t+1-\alpha}{t+1}\right)^2 \mu_{t-1}. \end{aligned} \tag{26}$$

As $\alpha > 3$, then $t+1-\alpha \leq t+1-2 = t-1$. Therefore, Equation 26 becomes

$$\xi_{t+1} + \mu_t \leq \xi_t + \left(\frac{t-1}{t+1}\right)^2 \mu_{t-1}, \tag{27}$$

and then

$$(t+1)^2\mu_t - (t-1)^2\mu_{t-1} \leq (t+1)^2(\xi_t - \xi_{t+1}). \tag{28}$$

Now, for $t \geq 1$,

$$\begin{aligned} (t+1)^2(\xi_t - \xi_{t+1}) &= t^2\xi_t - (t+1)^2\xi_{t+1} + (2t+1)\xi_t \\ &\leq t^2\xi_t - (t+1)^2\xi_{t+1} + 3t\xi_t. \end{aligned} \tag{29}$$

Now,

$$(t+1)^2\mu_t - (t-1)^2\mu_{t-1} = t^2\mu_t - (t-1)^2\mu_{t-1} + (2t+1)\mu_t.$$

From Equation 28, Equation 29 and using $\mu_t > 0$, we get

$$\begin{aligned} 2t\mu_t + t^2\mu_t - (t-1)^2\mu_{t-1} &\leq (t+1)^2\mu_t - (t-1)^2\mu_{t-1} \\ &\leq (t+1)^2(\xi_t - \xi_{t+1}) \\ &\leq t^2\xi_t - (t+1)^2\xi_{t+1} + 3t\xi_t. \end{aligned}$$

Summing the above equation for $t = 1, 2, \cdots, N$, we acquire

$$2\sum_{t=1}^{N} t\mu_t + N^2\mu_N \leq \xi_1 - (N+1)^2\xi_{N+1} + 3\sum_{t=1}^{N} t\xi_t.$$

If $\alpha > 3$, using Theorem 6 and Lemma 5 we obtain

$$2\sum_{k=1}^{N} t\mu_t \leq (\alpha-1)^2\mathcal{W}(1) + \frac{3(\alpha-1)^2\mathcal{W}(1)}{\alpha-3}.$$

Now taking $N \to \infty$, we have our result. □

Lemma 8. *Let assume* (K), $\gamma \in \left(0, \frac{1}{\rho+L}\right)$ *and* $\zeta^* \in S$. *If* $\alpha > 3$, *then*

$$\lim_{t\to\infty}\left[t^2\|\zeta_{t+1}-\zeta_t\|^2 + (t+1)^2(F(\zeta_{t+1}) - F(\zeta^*))\right] \quad \textit{exists.}$$

Proof. From Equation 26, we obtain

$$\xi_{t+1} + \mu_t \leq \xi_t + \left(\frac{t+1-\alpha}{t+1}\right)^2\mu_{t-1}.$$

Since $\alpha > 3$,

$$\xi_{t+1} + \mu_t \leq \xi_t + \left(\frac{t-1}{t}\right)^2\mu_{t-1},$$

equivalently,

$$t^2\mu_t - (t-1)^2\mu_{t-1} \leq t^2(\xi_t - \xi_{t+1}). \tag{30}$$

We have

$$(t+1)^2\xi_{t+1} - t^2\xi_t \leq t^2(\xi_{t+1} - \xi_t) + (2t+1)\xi_{t+1}. \tag{31}$$

Adding Equation 30 and 31, we obtain

$$\left[t^2\mu_t + (t+1)^2\xi_{t+1}\right] - \left[(t-1)^2\mu_{t-1} + t^2\xi_t\right] \leq 2(t+1)\xi_{t+1}. \tag{32}$$

From Lemma 5 the right-hand side of Equation 32 is summable. From Lemma 2, we claim that $\lim_{t\to\infty}(t-1)^2\mu_{t-1} + t^2\xi_t$ exists. □

The following theorem establishes an improved rate of convergence for the objective value gap, demonstrating faster decay of the function values toward the optimum.

Theorem 9. *Suppose that Assumption* (K) *holds,* $\gamma \in \left(0, \frac{1}{\rho+L}\right)$ *and* $\zeta^* \in S$. *Let* $(v_t)_{t\in\mathbb{N}}$ *and* $(\zeta_t)_{t\in\mathbb{N}}$ *be the sequences produced by NAPG*$_\alpha$. *When* $\alpha > 3$, *then*

(i) $F(\zeta_t) - F(\zeta^*) = o\left(\frac{1}{t^2}\right)$,

(ii) $\|\zeta_{t+1} - \zeta_t\| = o\left(\frac{1}{t}\right)$.

Proof. From Lemma 5 and 7, we deduce that

$$\sum_{t=1}^{\infty} \frac{1}{t}\Big[t^2\|\zeta_{t+1}-\zeta_t\|^2 + (t+1)^2(F(\zeta_{t+1}) - F(\zeta^*))\Big] < +\infty.$$

Now, analyzing the above result with Lemma 8, we obtain

$$\lim_{t\to\infty}\big[t^2\|\zeta_{t+1}-\zeta_t\|^2 + (t+1)^2(F(\zeta_{t+1}) - F(\zeta^*))\big] = 0.$$

Since all the terms are non-negative, we conclude that $t\|\zeta_{t+1}-\zeta_t\| \to 0$ and $(t+1)^2(F(\zeta_{t+1}) - F(\zeta^*)) \to 0$, as we want to claim. □

3 Convergence of the iterates

Here, we will illustrate the convergence of the iterates obtained by the NAPG$_\alpha$ algorithm.

Theorem 10. *Suppose Assumption (K) holds, $\alpha > 3$, $\gamma \in \big(0, \frac{1}{\rho+L}\big)$ and $\zeta^* \in S$. Let $(v_t)_{t\in\mathbb{N}}$ and $(\zeta_t)_{t\in\mathbb{N}}$ be the sequences generated by NAPG$_\alpha$. Then the sequence (ζ_t) converges to a point in S.*

Proof. We define previously as $b_t = \zeta_{t-1} + c_t(\zeta_t - \zeta_{t-1})$, we write

$$\begin{aligned}
\|b_t - \zeta^*\|^2 &= \|\zeta_{t-1} + c_t(\zeta_t - \zeta_{t-1}) - \zeta^*\|^2 \\
&= \Big(\frac{t-1}{\alpha-1}\Big)^2\|\zeta_t - \zeta_{t-1}\|^2 + 2\frac{t-1}{\alpha-1}\langle \zeta_{t-1} - \zeta^*, \zeta_t - \zeta_{t-1}\rangle + \|\zeta_{t-1} - \zeta^*\|^2. \qquad (33)
\end{aligned}$$

$$\begin{aligned}
\|\zeta_t - \zeta^*\|^2 &= \|\zeta_t - \zeta_{t-1} + \zeta_{t-1} - \zeta^*\|^2 \\
&= \|\zeta_t - \zeta_{t-1}\|^2 + \|\zeta_{t-1} - \zeta^*\|^2 + 2\langle \zeta_{t-1} - \zeta^*, \zeta_t - \zeta_{t-1}\rangle.
\end{aligned}$$

Equivalently,

$$2\langle \zeta_{t-1} - \zeta^*, \zeta_t - \zeta_{t-1}\rangle = \|\zeta_t - \zeta^*\|^2 - \|\zeta_t - \zeta_{t-1}\|^2 - \|\zeta_{t-1} - \zeta^*\|^2. \qquad (34)$$

Putting Equation 34 in Equation 33, we obtain

$$\begin{aligned}
\|b_t - \zeta^*\|^2 &= \Big[\Big(\frac{t-1}{\alpha-1}\Big)^2 - \Big(\frac{t-1}{\alpha-1}\Big)\Big]\|\zeta_t - \zeta_{t-1}\|^2 \\
&\quad + \frac{t-1}{\alpha-1}\big[\|\zeta_t - \zeta^*\|^2 - \|\zeta_{t-1} - \zeta^*\|^2\big] + \|\zeta_{t-1} - \zeta^*\|^2.
\end{aligned}$$

We require to prove that $\lim_{t\to\infty}\|b_t - \zeta^*\|$ exists. By Lemma 8 and 7, it is enough to show that

$$\Delta_t = (t-1)\big[\|\zeta_t - \zeta^*\|^2 - \|\zeta_{t-1} - \zeta^*\|^2\big] + (\alpha-1)\|\zeta_{t-1} - \zeta^*\|^2$$

possesses a limit as $t \to \infty$. Using Equation 22, Lemma 8 and 7, we have (Δ_t) is bounded. Let $\sigma_t = \|\zeta_t - \zeta^*\|^2$, then

$$\Delta_t = (t-1)(\sigma_t - \sigma_{t-1}) + (\alpha - 1)\sigma_{t-1},$$

and

$$\begin{aligned}\Delta_{t+1} - \Delta_t &= t(\sigma_{t+1} - \sigma_t) + (\alpha-1)\sigma_t - (t-1)(\sigma_t - \sigma_{t-1}) - (\alpha-1)\sigma_{t-1} \\ &= t(\sigma_{t+1} - \sigma_t) + (\alpha-1)(\sigma_t - \sigma_{t-1}) - (t-1)(\sigma_t - \sigma_{t-1}) \\ &= t(\sigma_{t+1} - \sigma_t) - (t-\alpha)(\sigma_t - \sigma_{t-1}). \end{aligned} \tag{35}$$

From Equation 12, we have

$$F(v_t - \gamma G_\gamma(v_t)) \leq F(\zeta^*) + (1-\rho\gamma)\langle G_\gamma(v_t), v_t - \zeta^*\rangle - \frac{\gamma(1-\rho\gamma)}{2}\|G_\gamma(v_t)\|^2.$$

Since $\zeta_{t+1} = v_t - \gamma G_\gamma(v_t)$, we obtain

$$0 \leq F(\zeta_{t+1}) - F(\zeta^*) \leq \frac{\gamma(\rho\gamma - 1)}{2}\|G_\gamma(v_t)\|^2 + (1-\rho\gamma)\langle G_\gamma(v_t), v_t - \zeta^*\rangle.$$

Then

$$\begin{aligned} 0 &\leq -\frac{\gamma}{2}\|G_\gamma(v_t)\|^2 + \langle G_\gamma(v_t), v_t - \zeta^*\rangle \\ &= -\|\zeta_{t+1} - v_t\|^2 + 2\langle v_t - \zeta_{t+1}, v_t - \zeta^*\rangle. \end{aligned} \tag{36}$$

Now,

$$\begin{aligned} \|\zeta_{t+1} - \zeta^*\|^2 &= \|(\zeta_{t+1} - v_t) + (v_t - \zeta^*)\|^2 \\ &= \|\zeta_{t+1} - v_t\|^2 + \|v_t - \zeta^*\|^2 + 2\langle \zeta_{t+1} - v_t, v_t - \zeta^*\rangle. \end{aligned}$$

Also,

$$2\langle v_t - \zeta_{t+1}, v_t - \zeta^*\rangle = \|\zeta_{t+1} - v_t\|^2 + \|v_t - \zeta^*\|^2 - \|\zeta_{t+1} - \zeta^*\|^2. \tag{37}$$

Putting Equation 37 in Equation 36, we get

$$0 \leq -\|\zeta_{t+1} - v_t\|^2 + \|\zeta_{t+1} - v_t\|^2 + \|v_t - \zeta^*\|^2 - \|\zeta_{t+1} - \zeta^*\|^2.$$

Hence,

$$\|\zeta_{t+1} - \zeta^*\|^2 \leq \|v_t - \zeta^*\|^2.$$

Using $v_t = \zeta_t + \alpha_t(\zeta_t - \zeta_{t-1})$, we obtain

$$\begin{aligned}\|\zeta_{t+1} - \zeta^*\|^2 &\leq \|\zeta_t - \zeta^* + \alpha_t(\zeta_t - \zeta_{t-1})\|^2 \\ &= \|\zeta_t - \zeta^*\|^2 + \alpha_t^2\|\zeta_t - \zeta_{t-1}\|^2 + 2\alpha_t\langle \zeta_t - \zeta^*, \zeta_t - \zeta_{t-1}\rangle. \end{aligned} \tag{38}$$

Now,

$$\begin{aligned}\|\zeta_{t-1} - \zeta^*\|^2 &= \|(\zeta_t - \zeta^*) - (\zeta_t - \zeta_{t-1})\|^2 \\ &= \|\zeta_t - \zeta^*\|^2 + \|\zeta_t - \zeta_{t-1}\|^2 - 2\langle \zeta_t - \zeta^*, \zeta_t - \zeta_{t-1}\rangle.\end{aligned}$$

Also,

$$2\langle \zeta_t - \zeta^*, \zeta_t - \zeta_{t-1}\rangle = \|\zeta_t - \zeta^*\|^2 + \|\zeta_t - \zeta_{t-1}\|^2 - \|\zeta_{t-1} - \zeta^*\|^2. \tag{39}$$

Putting Equation 39 in Equation 38, we obtain

$$\begin{aligned}\|\zeta_{t+1} - \zeta^*\|^2 \leq \|\zeta_t - \zeta^*\|^2 &+ (\alpha_t^2 + \alpha_t)\|\zeta_t - \zeta_{t-1}\|^2 \\ &+ \alpha_t[\|\zeta_t - \zeta^*\|^2 - \|\zeta_{t-1} - \zeta^*\|^2].\end{aligned}$$

Since $\alpha_t = 1 - \frac{\alpha}{t} \leq 1$,

$$\begin{aligned}\|\zeta_{t+1} - \zeta^*\|^2 \leq \|\zeta_t - \zeta^*\|^2 &+ 2\|\zeta_t - \zeta_{t-1}\|^2 \\ &+ \left(\frac{t-\alpha}{t}\right)[\|\zeta_t - \zeta^*\|^2 - \|\zeta_{t-1} - \zeta^*\|^2].\end{aligned}$$

Then, we obtain

$$t[\|\zeta_{t+1} - \zeta^*\|^2 - \|\zeta_t - \zeta^*\|^2] - (t-\alpha)[\|\zeta_t - \zeta^*\|^2 - \|\zeta_{t-1} - \zeta^*\|^2] \leq 2t\|\zeta_t - \zeta_{t-1}\|^2.$$

From Equation 35, we get

$$\Delta_{t+1} - \Delta_t \leq 2t\|\zeta_t - \zeta_{t-1}\|^2.$$

From Lemma 7, the series $\sum_{t=1}^{\infty} t\|\zeta_t - \zeta_{t-1}\|^2$ is summable. From Lemma 2, $\lim_{t\to\infty} \Delta_t$ exists. Hence, $\lim_{t\to\infty} \|b_t - \zeta^*\|$ exits. From Lemma 8 and the definition of b_t, $\lim_{t\to\infty} \|\zeta_{t-1} - \zeta^*\|$ exists. Given that each sequential cluster point of (ζ_t) in S. By Lemma 1, we deduce that the sequence (ζ_t) converges to a some point in S. □

4 Convergence of gradient

The following lemma establishes the Lipschitz type continuity of the gradient equivalent operator, which is essential for developing the Lyapunov study and demonstrating the rapid convergence of the gradient operator for a composite function.

Lemma 11. *Suppose Assumption (K) holds, and G_γ is the gradient mapping on $\mathbb{R}^n$, then for all $u, v \in \mathbb{R}^n$*

$$\|G_\gamma(u) - G_\gamma(v)\| \leq \frac{2 + L\gamma}{\gamma}\|u - v\|.$$

Proof. Since $\gamma G_\gamma(u) = u - Prox_{\gamma g}(u - \gamma\nabla f(u))$, for all $u,\ v \in \mathbb{R}^n$, we have

$$\begin{aligned}\gamma\|G_\gamma(u) - G_\gamma(v)\| &= \|(u - v) - (Prox_{\gamma g}(u - \gamma\nabla f(u)) - Prox_{\gamma g}(v - \gamma\nabla f(v)))\| \\ &\leq \|u - v\| + \|Prox_{\gamma g}(u - \gamma\nabla f(u)) - Prox_{\gamma g}(v - \gamma\nabla f(v))\|.\end{aligned}$$

Since proximal operator is firmly non-expansive, hence, the proximal operator is Lipschitz continuous with constant 1 [Proposition 12.28, [7]], then the above equation becomes

$$\begin{aligned}\gamma\|G_\gamma(u) - G_\gamma(v)\| &\leq \|u - v\| + \|u - v - \gamma(\nabla f(u) - \nabla f(v))\| \\ &\leq 2\|u - v\| + \gamma\|\nabla f(u) - \nabla f(v)\|.\end{aligned}$$

Since ∇f is L-Lipschitz continuous, we obtain

$$\gamma\|G_\gamma(u) - G_\gamma(v)\| \leq (2 + L\gamma)\|u - v\|. \tag{40}$$

□

Remark 2. *The gradient mapping G_γ satisfies the Lipschitz type continuity for the convex settings.*

Lemma 12. *Let assume (K) and $\gamma \in (0, \frac{1}{L+\rho})$, let $\check{y} \in \mathbb{R}^n$ and $\check{x} = P_\gamma(\check{y})$. Then for all $w \in \mathbb{R}^n$,*

$$\begin{aligned}F(\check{x}) \leq F(w) + (1 - \rho\gamma)\langle G_\gamma(\check{y}), \check{y} - w\rangle - \frac{\gamma(1 - \rho\gamma)}{2}\|G_\gamma(\check{y})\|^2 \\ - \frac{\gamma^2(\beta - \rho)}{2(2 + L\gamma)^2}\|G_\gamma(\check{y}) - G_\gamma(w)\|^2.\end{aligned}$$

Proof. From Equation 6, we have

$$\begin{aligned}g(\check{x}) + f(\check{x}) + \frac{\gamma^{-1} - \rho}{2}\|w - \check{x}\|^2 \leq g(w) + f(\check{y}) + \langle\nabla f(\check{y}), w - \check{y}\rangle \\ + \frac{\gamma^{-1} - \rho}{2}\|z - \check{y}\|^2 + \frac{\rho}{2}\|z - \check{y}\|^2.\end{aligned} \tag{41}$$

From Definition 1, $f(\check{y})) + \langle\nabla f(\check{y}), w - \check{y}\rangle \leq f(w) - \frac{\beta}{2}\|w - \check{y}\|^2$ holds, and as $F = f + g$. Equation 41 becomes

$$F(\check{x}) + \frac{\gamma^{-1} - \rho}{2}\|w - \check{x}\|^2 \leq F(w) + \frac{\gamma^{-1} - \rho}{2}\|w - \check{y}\|^2 - \frac{\beta}{2}\|w - \check{y}\|^2 + \frac{\rho}{2}\|w - \check{y}\|^2.$$

Equivalently,

$$F(\check{x}) + \frac{\gamma^{-1} - \rho}{2}\|w - \check{x}\|^2 \leq F(w) + \frac{\gamma^{-1} - \rho}{2}\|w - \check{y}\|^2 - \frac{\beta - \rho}{2}\|w - \check{y}\|^2. \tag{42}$$

Using Lemma 11, Equation 42 becomes

$$F(\check{x}) + \frac{\gamma^{-1} - \rho}{2}\|w - \check{x}\|^2 \leq F(w) + \frac{\gamma^{-1} - \rho}{2}\|w - \check{y}\|^2 - \frac{\gamma^2(\beta - \rho)}{2(2 + L\gamma)^2}\|G_\gamma(\check{y}) - G_\gamma(w)\|^2.$$

From Equation 9, we obtain

$$\begin{aligned} F(\check{x}) \leq F(w) + (1 - \rho\gamma)\langle G_\gamma(\check{y}), \check{y} - w\rangle - \frac{\gamma(1 - \rho\gamma)}{2}\|G_\gamma(\check{y})\|^2 \\ - \frac{\gamma^2(\beta - \rho)}{2(2 + L\gamma)^2}\|G_\gamma(\check{y}) - G_\gamma(w)\|^2. \end{aligned}$$

□

Theorem 13. *Let assume (K). Let $(\zeta_t)_{t\in\mathbb{N}}$ be the sequence generated by NAPG$_\alpha$, where $\alpha \geq 3$, and $\gamma \in \left(0, \frac{1}{\rho+L}\right)$, $\zeta^* \in S$. Then the following convergence rate is satisfied:*

$$\sum_{t=1}^{\infty} t(t + 1 - \alpha)\|G_\gamma(\zeta_t)\|^2 \leq \frac{2(2 + L\gamma)^2(\alpha - 1)^2}{\gamma^2(\beta - \rho)}\mathcal{W}(1) < +\infty.$$

Proof. From Lemma 12, we get

$$\begin{aligned} F(a - \gamma G_\gamma(a)) \leq F(b) + (1 - \rho\gamma)\langle G_\gamma(a), a - b\rangle - \frac{\gamma(1 - \rho\gamma)}{2}\|G_\gamma(a)\|^2 \\ - \frac{\gamma^2(\beta - \rho)}{2(2 + L\gamma)^2}\|G_\gamma(a) - G_\gamma(b)\|^2, \end{aligned}$$

for $a,\ b \in \mathbb{R}^n$. Put $a = v_t, b = \zeta_t$ and then $a = v_t, b = \zeta^*$ in the above equation, we obtain the respective equations as

$$\begin{aligned} F(v_t - \gamma G_\gamma(v_t)) \leq F(\zeta_t) + (1 - \rho\gamma)\langle G_\gamma(v_t), v_t - \zeta_t\rangle - \frac{\gamma(1 - \rho\gamma)}{2}\|G_\gamma(v_t)\|^2 \\ - \frac{\gamma^2(\beta - \rho)}{2(2 + L\gamma)^2}\|G_\gamma(v_t) - G_\gamma(\zeta_t)\|^2, \end{aligned} \tag{43}$$

and

$$\begin{aligned} F(v_t - \gamma G_\gamma(v_t)) \leq F(\zeta^*) + (1 - \rho\gamma)\langle G_\gamma(v_t), v_t - \zeta^*\rangle - \frac{\gamma(1 - \rho\gamma)}{2}\|G_\gamma(v_t)\|^2 \\ - \frac{\gamma^2(\beta - \rho)}{2(2 + L\gamma)^2}\|G_\gamma(v_t) - G_\gamma(\zeta^*)\|^2. \end{aligned}$$

From Remark 1, we have $G_\gamma(\zeta^*) = 0$ then the above equation becomes

$$F(v_t - \gamma G_\gamma(v_t)) \le F(\zeta^*) + (1-\rho\gamma)\langle G_\gamma(v_t), v_t - \zeta^*\rangle - \frac{\gamma(1-\rho\gamma)}{2}\|G_\gamma(v_t)\|^2 - \frac{\gamma^2(\beta-\rho)}{2(2+L\gamma)^2}\|G_\gamma(v_t)\|^2. \tag{44}$$

Multiplying $(c_{t+1}-1)$ in Equation 43 and then adding with Equation 44 and by considering $\zeta_{t+1} = v_t - \gamma G_\gamma(v_t)$, we obtain

$$\begin{aligned} c_{t+1}F(\zeta_{t+1}) \le\ & (c_{t+1}-1)F(\zeta_t) + F(\zeta^*) - \frac{\gamma c_{t+1}}{2}(1-\rho\gamma)\|G_\gamma(v_t)\|^2 \\ & + (1-\rho\gamma)\langle G_\gamma(v_t), (c_{t+1}-1)(v_t - \zeta_t) + (v_t - \zeta^*)\rangle \\ & - \frac{\gamma^2(\beta-\rho)}{2(2+L\gamma)^2}\Big[\|G_\gamma(v_t)\|^2 + (c_{t+1}-1)\|G_\gamma(v_t) - G_\gamma(\zeta_t)\|^2\Big]. \end{aligned} \tag{45}$$

Using Equation 14 and $b_t := \zeta_{t-1} + c_t(\zeta_t - \zeta_{t-1})$, Equation 45 becomes

$$\begin{aligned} c_{t+1}F(\zeta_{t+1}) \le\ & (c_{t+1}-1)F(\zeta_t) + F(\zeta^*) - \frac{\gamma c_{t+1}}{2}(1-\rho\gamma)\|G_\gamma(v_t)\|^2 \\ & + (1-\rho\gamma)\langle G_\gamma(v_t), b_t - \zeta^*\rangle \\ & - \frac{\gamma^2(\beta-\rho)}{2(2+L\gamma)^2}\Big[\|G_\gamma(v_t)\|^2 + (c_{t+1}-1)\|G_\gamma(v_t) - G_\gamma(\zeta_t)\|^2\Big]. \end{aligned} \tag{46}$$

Subtracting $c_{t+1}F(\zeta^*)$ on each sides of Equation 46, we obtain

$$\begin{aligned} c_{t+1}(F(\zeta_{t+1}) - F(\zeta^*)) \le\ & (c_{t+1}-1)(F(\zeta_t) - F(\zeta^*)) - \frac{\gamma c_{t+1}}{2}(1-\rho\gamma)\|G_\gamma(v_t)\|^2 \\ & + (1-\rho\gamma)\langle G_\gamma(v_t), b_t - \zeta^*\rangle \\ & - \frac{\gamma^2(\beta-\rho)}{2(2+L\gamma)^2}\Big[\|G_\gamma(v_t)\|^2 + (c_{t+1}-1)\|G_\gamma(v_t) - G_\gamma(\zeta_t)\|^2\Big]. \end{aligned} \tag{47}$$

Using Equation 18, Equation 47 becomes

$$\begin{aligned} c_{t+1}(F(\zeta_{t+1}) - F(\zeta^*)) \le\ & (c_{t+1}-1)(F(\zeta_t) - F(\zeta^*)) \\ & + \frac{(1-\rho\gamma)}{2\gamma c_{t+1}}\left[\|b_t - \zeta^*\|^2 - \|b_{t+1} - \zeta^*\|^2\right] \\ & - \frac{\gamma^2(\beta-\rho)}{2(2+L\gamma)^2}\Big[\|G_\gamma(v_t)\|^2 + (c_{t+1}-1)\|G_\gamma(v_t) - G_\gamma(\zeta_t)\|^2\Big]. \end{aligned} \tag{48}$$

Now multiply c_{t+1} on both sides and by simplifying, we get

$$\begin{aligned} c_{t+1}^2(F(\zeta_{t+1}) - F(\zeta^*)) + \frac{(1-\rho\gamma)}{2\gamma}\|b_{t+1} - \zeta^*\|^2 - (c_{t+1}^2 - c_{t+1} - c_t^2)(F(\zeta_t) - F(\zeta^*)) \\ \leq c_t^2(F(\zeta_t) - F(\zeta^*)) + \frac{(1-\rho\gamma)}{2\gamma}\|b_t - \zeta^*\|^2 \\ - \frac{\gamma^2(\beta-\rho)c_{t+1}}{2(2+L\gamma)^2}\Big[\|G_\gamma(v_t)\|^2 + (c_{t+1}-1)\|G_\gamma(v_t) - G_\gamma(\zeta_t)\|^2\Big]. \quad (49) \end{aligned}$$

Define $\mathcal{W}(t) := c_t^2(F(\zeta_t) - F(\zeta^*)) + \frac{(1-\rho\gamma)}{2\gamma}\|b_t - \zeta^*\|$, then the above equation becomes

$$\begin{aligned} \mathcal{W}(t+1) - &(c_{t+1}^2 - c_{t+1} - c_t^2)(F(\zeta_t) - F(\zeta^*)) \\ &\leq \mathcal{W}(t) - \frac{\gamma^2(\beta-\rho)c_{t+1}}{2(2+L\gamma)^2}\Big[\|G_\gamma(v_t)\|^2 + (c_{t+1}-1)\|G_\gamma(v_t) - G_\gamma(\zeta_t)\|^2\Big]. \quad (50) \end{aligned}$$

Now, we have

$$\begin{aligned} \|G_\gamma(v_t)\|^2 + (c_{t+1}-1)\|G_\gamma(v_t) - G_\gamma(\zeta_t)\|^2 \\ = \|G_(v_t)\|^2 + (c_{t+1}-1)\|G_\gamma(v_t)\|^2 \\ + (c_{t+1}-1)\|G_\gamma(\zeta_t)\|^2 + 2(1-c_{t+1})\langle G_\gamma(v_t), G_\gamma(\zeta_t)\rangle. \quad (51) \end{aligned}$$

By using Cauchy-Schwarz inequality, Equation 51 becomes

$$\begin{aligned} &\|G_\gamma(v_t)\|^2 + (c_{t+1}-1)\|G_\gamma(v_t) - G_\gamma(\zeta_t)\|^2 \\ &\geq c_{t+1}\|G_(v_t)\|^2 + (c_{t+1}-1)\|G_\gamma(\zeta_t)\|^2 - 2(1-c_{t+1})\|G_\gamma(v_t)\|\|G_\gamma(\zeta_t)\| \\ &= c_{t+1}\|G_(v_t)\|^2 + (c_{t+1}-1)\|G_\gamma(\zeta_t)\|^2 + 2(c_{t+1}-1)\|G_\gamma(v_t)\|\|G_\gamma(\zeta_t)\|. \quad (52) \end{aligned}$$

Using Equation 52, Equation 50 becomes

$$\begin{aligned} &\mathcal{W}(t+1) - (c_{t+1}^2 - c_{t+1} - c_t^2)(F(\zeta_t) - F(\zeta^*)) \\ &\leq \mathcal{W}(t) - \frac{\gamma^2(\beta-\rho)}{2(2+L\gamma)^2}\Big[c_{t+1}^2\|G_(v_t)\|^2 + c_{t+1}(c_{t+1}-1)\|G_\gamma(\zeta_t)\|^2 \\ &+ 2c_{t+1}(c_{t+1}-1)\|G_\gamma(v_t)\|\|G_\gamma(\zeta_t)\|\Big]. \quad (53) \end{aligned}$$

Also, we obtain

$$\begin{aligned} &\mathcal{W}(t+1) - (c_{t+1}^2 - c_{t+1} - c_t^2)(F(\zeta_t) - F(\zeta^*)) \\ &+ \frac{\gamma^2(\beta-\rho)}{2(2+L\gamma)^2}c_{t+1}(c_{t+1}-1)\|G_\gamma(\zeta_t)\|^2 \leq \mathcal{W}(t). \quad (54) \end{aligned}$$

As $c_{t+1}^2 - c_{t+1} - c_t^2 \le 0$, we get from Equation 54

$$\mathcal{W}(t+1) + \frac{\gamma^2(\beta-\rho)}{2(2+L\gamma)^2} c_{t+1}(c_{t+1}-1)\|G_\gamma(\zeta_t)\|^2 \le \mathcal{W}(t).$$

By summing the above equation for $t = 1, 2, \cdots, N$, we obtain

$$\mathcal{W}(N+1) + \sum_{t=1}^{N} \frac{\gamma^2(\beta-\rho)}{2(2+L\gamma)^2} c_{t+1}(c_{t+1}-1)\|G_\gamma(\zeta_t)\|^2 \le \mathcal{W}(1).$$

Also, we have

$$\sum_{t=1}^{N} \frac{\gamma^2(\beta-\rho)}{2(2+L\gamma)^2} c_{t+1}(c_{t+1}-1)\|G_\gamma(\zeta_t)\|^2 \le \mathcal{W}(1).$$

As $N \to \infty$, we get

$$\sum_{t=1}^{\infty} t(t+1-\alpha)\|G_\gamma(\zeta_t)\|^2 \le \frac{2(2+L\gamma)^2(\alpha-1)^2}{\gamma^2(\beta-\rho)} \mathcal{W}(1).$$

□

Remark 3. *Since $\sum_{t=1}^{\infty} t^2\|G_\gamma(\zeta_t)\|^2 < +\infty$, we can conclude that*

$$\min_{1\le j\le t} \|G_\gamma(\zeta_j)\|^2 \sum_{j=1}^{t} j^2 \le \sum_{j=1}^{t} j^2\|G_\gamma(\zeta_j)\|^2 \le \sum_{j=1}^{\infty} j^2\|G_\gamma(\zeta_j)\|^2 < +\infty.$$

Hence, we obtain

$$\min_{1\le j\le t} \|G_\gamma(\zeta_j)\|^2 = \mathcal{O}\left(\frac{1}{t^3}\right).$$

Similar result holds for v_t.

5 Ravine method

Gelfand and Tsetlin [18] initially introduced the Ravine method in 1961. The Ravine method produces the sequence $(v_t)_{k\in\mathbb{N}}$ which satisfy $u_t = v_t - \gamma\nabla f_\gamma(v_t)$ and then $v_{t+1} = u_t + (1 - \frac{\alpha}{t+1})(u_t - u_{t-1})$ for smooth convex function f.

Here, we define the Ravine accelerated proximal gradient descent method for Problem (P),

$$(\mathrm{RAPG}_\alpha) \begin{cases} u_t = v_t - \gamma G_\gamma(v_t) \\ v_{t+1} = u_t + (1 - \dfrac{\alpha}{t+1})(u_t - u_{t-1}). \end{cases}$$

In the following theorem, we show that the equivalence relation between the NAPG_α scheme and the RAPG_α scheme.

Theorem 14. *(i) Suppose $(u_t)_{t\in\mathbb{N}}$ is a sequence induced by the algorithm NAPG$_\alpha$. Let $(v_t)_{t\in\mathbb{N}}$ be the corresponding sequence given by $v_t = u_t + (1-\frac{\alpha}{t})(u_t - u_{t-1})$. Then $(v_t)_{t\in\mathbb{N}}$ satisfies the algorithm RAPG$_\alpha$.*

(ii) Let the sequence $(v_t)_{t\in\mathbb{N}}$ is induced from RAPG$_\alpha$, then the sequence $(u_t)_{t\in\mathbb{N}}$ stated by $u_{t+1} = v_t - \gamma G_\gamma(v_t)$ satisfies the algorithm NAPG$_\alpha$.

Proof. (i) Let us consider the iterates $(u_t)_{t\in\mathbb{N}}$ produced by $(NAPG_\alpha)$. Since $v_t = u_t + (1-\frac{\alpha}{t})(u_t - u_{t-1})$, then

$$\begin{aligned}v_{t+1} &= u_{t+1} + (1-\frac{\alpha}{t+1})(u_{t+1} - u_t)\\ &= v_t - \gamma G_\gamma(v_t) + (1-\frac{\alpha}{t+1})((v_t - \gamma G_\gamma(v_t)) - (v_{t-1} - \gamma G_\gamma(v_{t-1}))).\end{aligned}$$

Set $u_t = v_t - \gamma G_\gamma(v_t)$, we get

$$v_{t+1} = u_t + (1-\frac{\alpha}{t+1})(u_t - u_{t-1}).$$

Hence, the sequence $(v_t)_{k\in\mathbb{N}}$ complies RAPG$_\alpha$.
(ii) From the definition of v_t and u_t in RAPG$_\alpha$, we have

$$v_{t+1} = v_t - \gamma G_\gamma(v_t) + (1-\frac{\alpha}{t+1})((v_t - \gamma G_\gamma(v_t)) - (v_{t-1} - \gamma G_\gamma(v_{t-1}))).$$

Setup $u_{t+1} = v_t - \gamma G_\gamma(v_t)$, we get

$$v_{t+1} = u_{t+1} + (1-\frac{\alpha}{t+1})(u_{t+1} - u_t).$$

Equivalently, $v_t = u_t + (1-\frac{\alpha}{t})(u_t - u_{t-1})$. Combining the preceding relationship with the u_{t+1} definition, we obtain $(u_t)_{t\in\mathbb{N}}$ satisfies NAPG$_\alpha$. □

Theorem 15. *Suppose that Assumption (K) holds, $\alpha \geq 3$ and $\gamma \in \big(0, \frac{1}{\rho+L}\big)$, $\zeta^* \in S$. Let $(v_t)_{t\in\mathbb{N}}$ and $(u_t)_{t\in\mathbb{N}}$ be the sequences generated by RAPG$_\alpha$. Then*

(i) $F(u_t) - F(\zeta^*) = \mathcal{O}\big(\frac{1}{t^2}\big)$;

(ii) $\|G_\gamma(v_t)\| = \mathcal{O}\big(\frac{1}{t}\big)$;

when $\alpha > 3$

(iii) $F(u_t) - F(\zeta^*) = o\big(\frac{1}{t^2}\big)$;

(iv) $\lim v_t = \lim u_t \in S$.

Proof. From Lemma 4, we have

$$F(b-\gamma G_\gamma(b)) \leq F(a) + (1-\rho\gamma)\langle G_\gamma(b), b-a\rangle - \frac{\gamma}{2}(1-\rho\gamma)\|G_\gamma(b)\|^2.$$

for any $a,\ b \in \mathbb{R}^n$. Putting $a = u_{t-1}$ and $b = v_t$, and as $(\rho\gamma - 1) < 0$ we deduce that

$$F(u_t) \leq F(u_{t-1}) + (1-\rho\gamma)\langle G_\gamma(v_t), v_t - u_{t-1}\rangle.$$

By Cauchy-Schwarz inequality, we get

$$F(u_t) \leq F(u_{t-1}) + (1-\rho\gamma)\|G_\gamma(v_t)\|\|v_t - u_{t-1}\|. \tag{55}$$

From the algorithm RAPG$_\alpha$, we have

$$\begin{aligned}\|G_\gamma(v_t)\| &= \frac{1}{\gamma}\|u_t - v_t\| \\ &\leq \frac{1}{\gamma}\left(\|u_t - u_{t-1}\| + \|v_t - u_{t-1}\|\right) \\ &\leq \frac{1}{\gamma}\left(\|u_t - u_{t-1}\| + \|u_{t-1} - u_{t-2}\|\right).\end{aligned} \tag{56}$$

From Theorem 14, we get an equivalence relation between the two sequences (u_t) and (v_t). Hence, $\zeta_{t+1} = u_t$. Equation 56 becomes

$$\|G_\gamma(v_t)\| \leq \frac{1}{\gamma}\left(\|\zeta_{t+1} - \zeta_t\| + \|\zeta_t - \zeta_{t-1}\|\right). \tag{57}$$

From Theorem 6, Equation 57 becomes

$$\|G_\gamma(v_t)\| = \mathcal{O}\left(\frac{1}{t}\right).$$

In a similar way,

$$\|v_t - u_{t-1}\| \leq \|u_{t-1} - u_{t-2}\| = \|\zeta_t - \zeta_{t-1}\| = \mathcal{O}\left(\frac{1}{t}\right). \tag{58}$$

Entering the preceding estimates into Equation 55, as $\zeta_t = u_{t-1}$, we obtain

$$F(u_t) - F(\zeta^*) \leq F(\zeta_t) - F(\zeta^*) + \frac{C}{t^2} = \mathcal{O}\left(\frac{1}{t^2}\right). \tag{59}$$

From Theorem 14, Equation 59 becomes

$$F(u_t) - F(\zeta^*) = \mathcal{O}\left(\frac{1}{t^2}\right).$$

For $\alpha > 3$, Theorem 9 implies that $F(\zeta_t) - F(\zeta^*) = o\left(\frac{1}{t^2}\right)$ and $\|\zeta_t - \zeta_{t-1}\| = o\left(\frac{1}{t}\right)$. Then, it implies that $F(u_t) - F(\zeta^*) = o\left(\frac{1}{t^2}\right)$. From Theorem 9, we have $\|\zeta_t - \zeta_{t-1}\| \to 0$, and from the NAPG$_\alpha$ algorithm, we have $v_t - \zeta_t = \alpha_t(\zeta_t - \zeta_{t-1})$. It implies that $\|v_t - \zeta_t\| \to 0$, i.e., $v_t - \zeta_t$ converges to 0. Since $(\zeta_t)_{t\in\mathbb{N}}$ converges when $\alpha > 3$ and $\zeta_{t+1} = u_t$, therefore the sequence $(v_t)_{t\in\mathbb{N}}$ converges to exactly the same limit as $(u_t)_{t\in\mathbb{N}}$. □

Theorem 16. *Let assume (K). Let $(v_t)_{t\in\mathbb{N}}$ be the sequence generated by RAPG$_\alpha$, where $\alpha \geq 3$ and $\gamma < \frac{1}{\rho+L}$. Then the following convergence rate is satisfied:*

$$\sum_{t=1}^{\infty} t^2 \|G_\gamma(v_t)\|^2 < +\infty.$$

Proof. From Equation 53, we have

$$\begin{aligned}
&\mathcal{W}(t+1) - (c_{t+1}^2 - c_{t+1} - c_t^2)(F(\zeta_t) - F(\zeta^*)) \\
&\leq \mathcal{W}(t) - \frac{\gamma^2(\beta-\rho)}{2(2+L\gamma)^2}\Big[c_{t+1}^2\|G_\gamma(v_t)\|^2 + c_{t+1}(c_{t+1}-1)\|G_\gamma(\zeta_t)\|^2 \\
&+ 2c_{t+1}(c_{t+1}-1)\|G_\gamma(v_t)\|\|G_\gamma(\zeta_t)\|\Big].
\end{aligned} \tag{60}$$

Also, we have

$$\begin{aligned}
&\mathcal{W}(t+1) - (c_{t+1}^2 - c_{t+1} - c_t^2)(F(\zeta_t) - F(\zeta^*)) \\
&+ \frac{\gamma^2(\beta-\rho)}{2(2+L\gamma)^2} c_{t+1}^2\|G_\gamma(v_t)\|^2 \leq \mathcal{W}(t).
\end{aligned} \tag{61}$$

Since, $(c_{t+1}^2 - c_{t+1} - c_t^2) \leq 0$, we obtain from Equation 61

$$\mathcal{W}(t+1) + \frac{\gamma^2(\beta-\rho)}{2(2+L\gamma)^2} c_{t+1}^2\|G_\gamma(v_t)\|^2 \leq \mathcal{W}(t).$$

Now by summing the above equation and as $c_{t+1} = \frac{t}{\alpha-1}$, we obtain

$$\sum_{t\in\mathbb{N}} t^2\|G_\gamma(v_t)\|^2 \leq \frac{2(2+L\gamma)^2(\alpha-1)^2}{\gamma^2(\beta-\rho)}\mathcal{W}(1) < +\infty.$$

□

6 Conclusion

Here, we construct NAPG$_\alpha$ method for the composite function, where the smooth component is strongly convex and the non-smooth component is weakly convex and establish that the convergence rate of the objective function satisfies $F(\zeta_t) - F(\zeta^*) = o\left(\frac{1}{t^2}\right)$ for $\alpha > 3$ while the gradient mapping exhibits a rapid convergence rate of $\|G_\gamma(\zeta_t)\|$ is $o\left(\frac{1}{t}\right)$ for $\alpha \geq 3$. These results are derived through a Lyapunov-based analysis of the additive structure comprising smooth and non-smooth components within the strongly-weakly convex class of functions. Furthermore, analogous convergence results are obtained for the RAPG$_\alpha$ scheme under the same class of functions.

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 and Alexandre Cabot. Convergence rates of inertial forward-backward algorithms. *SIAM Journal on Optimization*, 28:849–874, 2018.

[3] 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.

[4] Hedy Attouch, Zaki Chbani, and Hassan Riahi. Rate of convergence of the nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$. *ESAIM: Control, Optimisation and Calculus of Variations*, 25:2, 2019.

[5] Hedy Attouch and Jalal Fadili. From the ravine method to the nesterov method and vice versa: a dynamical system perspective. *SIAM Journal on Optimization*, 32:2074–2101, 2022.

[6] Hedy Attouch and Juan Peypouquet. The rate of convergence of nesterov's accelerated forward-backward method is actually faster than $1/k^2$. *SIAM Journal on Optimization*, 26:1824–1834, 2016.

[7] HH Bauschke and PL Combettes. Convex analysis and monotone operator theory in hilbert spaces, 2011 springer. *New York*, 2017.

[8] Amir Beck. *First-order methods in optimization*. SIAM, 2017.

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

[10] Axel Böhm and Stephen J Wright. Variable smoothing for weakly convex composite functions. *Journal of optimization theory and applications*, 188(3):628–649, 2021.

[11] Luca Calatroni and Antonin Chambolle. Backtracking strategies for accelerated descent methods with smooth composite objectives. *SIAM journal on optimization*, 29(3):1772–1798, 2019.

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

[13] Antonin Chambolle and Thomas Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016.

[14] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(11):1413–1457, 2004.

[15] Alexandre d'Aspremont, Damien Scieur, Adrien Taylor, et al. Acceleration methods. *Foundations and Trends® in Optimization*, 5(1-2):1–245, 2021.

[16] Mihai I Florea and Sergiy A Vorobyov. An accelerated composite gradient method for large-scale composite objective problems. *IEEE Transactions on Signal Processing*, 67(2):444–459, 2018.

[17] Mihai I Florea and Sergiy A Vorobyov. A generalized accelerated composite gradient method: Uniting nesterov's fast gradient method and fista. *IEEE Transactions on Signal Processing*, 68:3033–3048, 2020.

[18] IM Gelfand and M Zejtlin. Printszip nelokalnogo poiska v sistemah avtomatich, optimizatsii, dokl. *AN SSSR*, 137:295–298, 1961.

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

[20] Tim Hoheisel, Maxime Laborde, and Adam Oberman. A regularization interpretation of the proximal point method for weakly convex functions. *J. Dyn. Games*, 2020.

[21] Pham Duy Khanh, Boris S Mordukhovich, Vo Thanh Phat, and Dat Ba Tran. Inexact proximal methods for weakly convex functions. *Journal of Global Optimization*, 91:611–646, 2025.

[22] Feng-Yi Liao and Yang Zheng. A proximal descent method for minimizing weakly convex optimization. *arXiv preprint arXiv:2509.02804*, 2025.

[23] Pierre-Louis Lions and Bertrand Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[24] Dirk A Lorenz and Thomas Pock. An inertial forward-backward algorithm for monotone inclusions. *Journal of Mathematical Imaging and Vision*, 51(2):311–325, 2015.

[25] Bernard Martinet. Brève communication. régularisation d'inéquations variationnelles par approximations successives. *Revue française d'informatique et de recherche opérationnelle. Série rouge*, 4:154–158, 1970.

[26] Yurii Nesterov. A method for unconstrained convex minimization problem with the rate of convergence $o(1/k^2)$. In *Dokl. Akad. Nauk. SSSR*, volume 269, page 543, 1983.

[27] Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

[28] Boris T Polyak. Some methods of speeding up the convergence of iteration methods. *Ussr computational mathematics and mathematical physics*, 4(5):1–17, 1964.

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

[30] Kansei Ushiyama. A $\sqrt{2}$-accelerated fista for composite strongly convex problems. *arXiv preprint arXiv:2509.09295*, 2025.