Generalized Nesterov's accelerated proximal gradient algorithms with convergence rate of order o(1/k 2)

van Ngai Huynh, Anh Son • Ta

Noname manuscript No.
(will be inserted by the editor)

Generalized Nesterov's accelerated proximal gradient algorithms with convergence rate of order $o(1/k^2)$

Huynh Van Ngai · Ta Anh Son

Received: date / Accepted: date

Abstract The accelerated gradient method initiated by Nesterov is now recognized to be one of the most powerful tools for solving smooth convex optimization problems. This method improves significantly the convergence rate of function values from $O(1/k)$ of the standard gradient method down to $O(1/k^2)$. In this paper, we present two generalized variants of Nesterov's accelerated proximal gradient method for solving composition convex optimization problems in which the objective function is represented by the sum of a smooth convex function and a nonsmooth convex part. We show that with suitable ways to pick the sequences of parameters, the convergence rate for the function values of this proposed method is actually of order $o(1/k^2)$. Especially, when the objective function is $p-$uniformly convex for $p > 2$, the convergence rate is of order $O\left(\ln k/k^{2p/(p-2)}\right)$, and the convergence is linear if the objective function is strongly convex. By-product, we derive a forward-backward algorithm generalizing the one by Attouch-Peypouquet [1], which produces a convergence sequence with a convergence rate of the function values of order $o(1/k^2)$.

Key words: convex optimization, forward-backward method, Nesterov accelerated gradient method, proximal mapping, subdifferential

Mathematics Subject Classification: 49J52, 90C26, 90C30, 49M37, 65K05, 90C25

Research under Vingroup Innovation Foundation (VINIF) annual research support program in project code VINIF.2019.DA09

First author
Department of Mathematics and Statistics, 170 An Duong Vuong, Quy Nhon, Viet Nam
E-mail: vanngaihuynh@gmail.com

Second author
School of Applied Mathematics and Informatics Ha Noi University of Science and Technology
E-mail: son.taanh@hust.edu.vn

1 Introduction

Consider the *composition convex* optimization problem of the form

$$\min\{f(x)+\Phi(x):\ \ x\in\mathbb{R}^n\},\tag{1}$$

where $\Phi:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ is a proper lower-semicontinuous convex function and $f:\mathbb{R}^n\to\mathbb{R}$ is a continuously differentiable convex function with $L-$Lipschitz continuous gradient on $\operatorname{dom}\Phi$, for $L>0$, that is,

$$\|\nabla f(x)-\nabla f(y)\|\le L\|x-y\|,\ \ \forall x,y\in\operatorname{dom}\Phi.\tag{2}$$

This class of convex optimization problems arises in many applications, especially, in image processing and in machine learning ([4,5]). Recalling, the algorithms of forward-backward type (or also called gradient proximal algorithms), generalizing the gradient projection method ([10,12]), which exploit the additive separability of the smooth part and the nonsmooth one of the objective function, play an important role for solving (1) (see [13,7] and the references given therein).

The celebrated acceleration scheme initiated by Nesterov in 1983 ([14], [15]) for solving smooth unconstrained convex optimization problem improves the theoretical convergence rate (for the function values) from $O(1/k)$ (of the standard gradient method) down $O(1/k^2)$. Nowadays this accelerated gradient method is recognized to be one of the most powerful first-order methods for solving smooth convex optimization problems. Later, this acceleration scheme was developed for solving composition convex optimization of the form (1) in which the objective function is represented by the sum of a smooth convex function and a nonsmooth one (see [11,15,17,18] and the references given therein). In [4], a combination of the forward-backward method with Nesterov's acceleration scheme for solving (1) was proposed, called the *fast iterative shrinkage-thresholding* algorithm (FISTA), and it was successfully applied to image processing. In [1] (see also [3]), it was shown that the convergence rate of the *accelerated forward-backward* method (with respect to a special sequence of parameters) is actually $o(1/k^2)$, rather than $O(1/k^2)$, with a proof relying on an appropriate finite-difference discretization of a differential inclusion (see [2] and the references given therein for further about this approach).

In this paper, we will develop two accelerated schemes which generalize the one by Nesterov [17]. We show that by updating sequences of parameters in a suitable way, the convergence rate for the function values is actually of the order $o(1/k^2)$ for the convex case, and is $O\left(\ln k/k^{2p/(p-2)}\right)$ for the $p-$uniformly convex case with $p>2$. Moreover, when the objective function is strongly convex, the convergence is linear. By-product, as a particular case, the established convergence results permit us to derive a forward-backward algorithm generalizing the one considered by Attouch- Peypouquet which produces convergence sequences with rate of order $o(1/k^2)$.

Let us recall some basis notations and properties. In the sequel, the space $\mathbb{R}^n$ is equipped with the canonical inner product $\langle\cdot\rangle$, and the subdifferential of a convex function $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ at $x \in \operatorname{dom} \varphi$ is denoted by $\partial\varphi(x)$, that is,

$$\partial\varphi(x) = \{x^* \in \mathbb{R}^n : \quad \langle x^*, y - x\rangle \leq \varphi(y) - \varphi(x) \quad \forall y \in \mathbb{R}^n\}.$$

We set $\partial\varphi(x) = \emptyset$ if $x \notin \operatorname{dom} \varphi$. The notation $\operatorname{prox}_\varphi$ denotes the *proximal mapping* of the function φ (see [7]). That is,

$$\operatorname{prox}_\varphi(x) = \operatorname{argmin}\{\varphi(y) + \frac{1}{2}\|y - x\|^2 : \quad y \in \mathbb{R}^n\}.$$

A function $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is called $p-$unformly convex with parameter μ, for some $\mu \geq 0$, $p \geq 2$, or called $(\mu, p)-$uniformly convex if for all $x, y \in \mathbb{R}^n$, $\lambda \in [0, 1]$ one has

$$\varphi(\lambda x + (1 - \lambda)y) \leq \lambda\varphi(x) + (1 - \lambda)\varphi(y) - \frac{\mu}{p}\lambda(1 - \lambda)\|x - y\|^p.$$

When $p = 2$, the function φ is called strongly convex (with parameter μ.) Note that if φ is $(\mu, p)-$uniformly convex, then for all $x, y \in \mathbb{R}^n$, all $x^* \in \partial\varphi(x)$, one has

$$\langle x^*, y - x\rangle \leq \varphi(y) - \varphi(x) - \frac{\mu}{p}\|y - x\|^p. \tag{3}$$

For a function f which is differentiable on a convex set $\Omega \subseteq \mathbb{R}^n$ such that the gradient of this function ∇f is $L - Lipschitz$ on Ω, the well-known inequality (see e.g., [16]) is useful in the sequel.

$$f(y) \leq f(x) + \langle\nabla f(x), y - x\rangle + \frac{L}{2}\|y - x\|^2, \ \ \forall x, y \in \Omega. \tag{4}$$

2 Generalized Nesterov's Algorithm and convergence rates

2.1 Algorithm

Firstly we introduce the following notion of support functions of a convex function at a point.

Definition 1 For a convex function $\Phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and a point $z \in \mathbb{R}^n$. A convex function $\Psi_z := \Psi_{z,\Phi} : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is called a lower support function to Φ at z if $\Psi_z \leq \Phi$ and $\Psi_z(z) = \Phi(z)$.

Obviously, the usual two lower support functions of a convex function Φ, at a point z : the first is itself Φ, and the second is the linear function

$$\Psi_z(x) := \Phi(z) + \langle z^*, x - z\rangle, \ x \in \mathbb{R}^n,$$

where $z^* \in \partial\Phi(z)$, when Φ is subdifferentiable at z.

In what follows we make use of the following assumptions:

(A1) The optimal solution set of problem (1) is nonempty.

(A2) The function $\Phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is proper lower semicontinuous convex; the function $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable convex function such that its gradient ∇f is $L-$Lipschitz (for some $L > 0$) on $\operatorname{dom} \Phi$.

Pick parameters $C, \kappa, \mu \geq 0$; a sequence of positive reals $\{\alpha_k\}$, and two sequences of nonnegative reals $\{\beta_k\}$ and $\{\gamma_k\}$. Assume that the sequences $\{\alpha_k\}$, $\{\beta_k\}$ verify the condition

$$A_k := \sum_{i=0}^{k} \alpha_k \geq B_k := \sum_{i=0}^{k} \beta_k, \quad \text{for all } \ k \in \mathbb{N}. \tag{5}$$

Pick a strongly convex function $h : \mathbb{R}^n \to \mathbb{R}$ with a strong convexity parameter $\rho > 0$, which has a minimizer at $y_0 \in \operatorname{dom} \Phi$. Without loss of generality, we can assume $h(y_0) = 0$. Then one has

$$h(x) \geq \frac{\rho}{2}\|x - y_0\|^2, \ \text{ for all } x \in \mathbb{R}^n. \tag{6}$$

The algorithm is stated in the following scheme.

Algorithm 1: Generalized Nesterov's accelerated proximal gradient algorithm (GAPGA)

`Initialization:` Initial data: y^0 as in (6). Set $k = 0$.

`Repeat:` For $k = 0, 1, ...,$

1. Find

$$\begin{aligned} x_k &= \operatorname{argmin}\left\{\Phi(y) + \langle \nabla f(y_k), y - y_k\rangle + \tfrac{1}{2\kappa}\|y - y_k\|^2 : \ \ y \in \mathbb{R}^n\right\} \\ &= \operatorname{prox}_{\kappa\Phi}\left(y_k - \kappa \nabla f(y_k)\right). \end{aligned} \tag{7}$$

2. Find

$$\begin{aligned} z_k = \operatorname{argmin}_{x\in\mathbb{R}^n}&\{Ch(x) + \textstyle\sum_{i=0}^{k-1} \alpha_i[f(y_i) + \langle \nabla f(y_i), x - y_i\rangle + \Psi_{z_i}(x) \\ &+ \tfrac{1}{2}\mu\gamma_i\|x - y_i\|^2] + \alpha_k[f(y_k) + \langle \nabla f(y_k), x - y_k\rangle + \Phi(x) + \tfrac{1}{2}\mu\gamma_k\|x - y_k\|^2]\} \end{aligned} \tag{8}$$

3. Set Ψ_{z_k} is a support function to Φ at z_k such that

$$\begin{aligned} &\min_{x\in\mathbb{R}^n}\{Ch(x) + \textstyle\sum_{i=0}^{k-1} \alpha_i[f(y_i) + \langle \nabla f(y_i), x - y_i\rangle + \Psi_{z_i}(x) + \tfrac{1}{2}\mu\gamma_i\|x - y_i\|^2] \\ &+\alpha_k[f(y_k) + \langle \nabla f(y_k), x - y_k\rangle + \Phi(x) + \tfrac{1}{2}\mu\gamma_k\|x - y_k\|^2]\} \\ &= \min_{x\in\mathbb{R}^n}\{Ch(x) + \textstyle\sum_{i=0}^{k} \alpha_i[f(y_i) + \langle \nabla f(y_i), x - y_i\rangle + \Psi_{z_i}(x) + \tfrac{1}{2}\mu\gamma_i\|x - y_i\|^2]\}. \end{aligned} \tag{9}$$

4. Set

$$\tau_k := \frac{\alpha_{k+1}}{A_{k+1} - B_k}, \ \ y_{k+1} = \tau_k z_k + (1 - \tau_k)x_k.$$

Remark 1.

(i). In Nesterov's original accelerated schemes ([16,17]), $\tau_k := \frac{\alpha_{k+1}}{A_{k+1}}$. which is a particular case of Algorithm 1 with $\beta_k := 0$, $k \in \mathbb{N}$.

(ii). In Step 3 of Algorithm 1, we can take $\Psi_{z_k} = \Phi$. If we set $\Psi_{z_k} = \Phi$, for all $k \in \mathbb{N}$, Algorithm 1 gives a generalized variant of Nesterov's accelerated dual averaging algorithm. An another way to choose Ψ_{z_k} is as follows. As in Step 2, z_k is a minimizer of the convex function in the right hand of (8), then there is $z_k^* \in \partial\Phi(z_k)$ such that

$$0 \in C\partial h(z_k)+\sum_{i=0}^{k-1}\alpha_i[\nabla f(y_i)+\partial\Psi_{z_i}(z_k)]+\alpha_k[\nabla f(y_k)+z_k^*]+\mu\sum_{i=0}^{k}\alpha_i\gamma_i(z_i-y_i). \tag{10}$$

Then the support function

$$\Psi_{z_k}(x) := \langle z_k^*, x - z_k\rangle + \Phi(z_k), \quad x \in \mathbb{R}^n, \tag{11}$$

verify condition (9) in step 3.

Especially, when $h(x) := \frac{1}{2}\|x-y_0\|^2$, and for all $k \in \mathbb{N}$, the support function Ψ_{z_k} is defined by (11) for all $k \in \mathbb{N}$, then in view of (10) for k and $(k+1)$, one has, for all $k \in \mathbb{N}$, for some $z_{k+1}^* \in \partial\Phi(z_{k+1})$,

$$\begin{aligned}0 \in\ & (C + \mu\alpha_{k+1}\gamma_{k+1})z_{k+1} - (C + \mu\alpha_k\gamma_k)z_k \\ & +\mu\alpha_k\gamma_k y_k - \alpha_{k+1}\gamma_{k+1}y_{k+1} + \alpha_{k+1}[z_{k+1}^* + \nabla f(y_{k+1})].\end{aligned}$$

Thus equivalently,

$$\begin{aligned}& z_{k+1} = \mathrm{prox}_{\frac{\alpha_{k+1}}{C+\mu\alpha_{k+1}\gamma_{k+1}}\Phi}\left[\tfrac{1}{C+\mu\alpha_{k+1}\gamma_{k+1}}W_{k+1}\right]; \\ & W_{k+1} := (C + \mu\alpha_k\gamma_k)z_k - \mu\alpha_k\gamma_k y_k + \alpha_{k+1}\gamma_{k+1}y_{k+1} - \alpha_{k+1}\nabla f(y_{k+1}).\end{aligned} \tag{12}$$

In particular, when $\mu = 0$, the sequence $\{z_k\}$ is defined recurrently by

$$z_{k+1} = \mathrm{prox}_{\frac{\alpha_{k+1}}{C}\Phi}\left[z_k - \frac{\alpha_{k+1}}{C}\nabla f(y_{k+1})\right]. \tag{13}$$

This is exactly the (accelerated) scheme of the proximal gradient methods.

2.2 Convergence

The following theorem gives an estimate for function values $f(x_k) + \Phi(x_k)$, and it is crucial to derive the subsequent convergence rates. Let us introduce the functions F_k, G_k by respectively,

$$\begin{aligned}F_k(x) = Ch(x) &+ \textstyle\sum_{i=0}^{k-1}\alpha_i[f(y_i) + \langle\nabla f(y_i), x - y_i\rangle + \Psi_{z_i}(x) + \frac{1}{2}\mu\gamma_i\|x - y_i\|^2] \\ &+\alpha_k[f(y_k) + \langle\nabla f(y_k), x - y_k\rangle + \Phi(x) + \tfrac{1}{2}\mu\gamma_k\|x - y_k\|^2], \quad x \in \mathbb{R}^n.\end{aligned} \tag{14}$$

$$G_k(x) = Ch(x) + \textstyle\sum_{i=0}^{k}\alpha_i[f(y_i) + \langle\nabla f(y_i), x - y_i\rangle + \Psi_{z_i}(x) + \frac{1}{2}\mu\gamma_i\|x - y_i\|^2]. \tag{15}$$

Theorem 1 *Let $\{x_k\}$ and $\{y_k\}$ be sequences generated by Algorithm 1. Suppose that $\kappa \le 1/L$ and the sequences $\{\alpha_k\}$, $\{\beta_k\}$ and $\{\gamma_k\}$ satisfy the condition*

$$\left(C\rho + \mu \sum_{i=0}^{k-1} \alpha_i\gamma_i\right)(A_k - B_{k-1}) \ge \alpha_k^2/\kappa, \ \textit{ for all } k \in \mathbb{N}. \tag{16}$$

Then one has for all $k \in \mathbb{N}$,

$$\begin{array}{l}\sum_{i=0}^{k} \beta_i[f(x_i) + \Phi(x_i)] + (A_k - B_k)[f(x_k) + \Phi(x_k)] \\ +\frac{1}{2}(1/\kappa - L)\sum_{i=0}^{k}(A_i - B_{i-1})\|x_i - y_i\|^2 \le \min_{x\in\mathbb{R}^n} F_k(x).\end{array} \tag{17}$$

where, we set $B_{-1} = 0$. Moreover, if f is $\mu-$strong convex, then (17) holds if $\gamma_k = 1$, $k \in \mathbb{N}$, and the sequences $\{\alpha_k\}$, $\{\beta_k\}$ verifying the condition

$$\left(C\rho + \mu \sum_{i=0}^{k-1} \alpha_i\right)(A_k - B_{k-1}) \ge \alpha_k^2(\kappa^{-1} - \mu), \ \textit{ for all } k \in \mathbb{N}. \tag{18}$$

Proof. We prove (17) by induction on $k \in \mathbb{N}$. For $k = 0$, one has

$$\begin{array}{l}\min_{x\in\mathbb{R}^n} F_0(x) \\ = \min\left\{Ch(x) + \alpha_0[f(y_0) + \langle\nabla f(y_0), x - y_0\rangle + \Phi(x) + \mu\gamma_0\|x - y_0\|^2] : \ x \in \mathbb{R}^n\right\} \\ \ge \alpha_0 \min\left\{\frac{1}{2}(C\rho + \alpha_0\mu\gamma_0)\alpha_0^{-1}\|x - y_0\|^2 + f(y_0) + \langle\nabla f(y_0), x - y_0\rangle + \Phi(x) : \ x \in \mathbb{R}^n\right\} \\ \overset{\text{by (16)}}{\ge} \ \alpha_0 \min\left\{\frac{1}{2\kappa}\|x - y_0\|^2 + f(y_0) + \langle\nabla f(y_0), x - y_0\rangle + \Phi(x) : \ x \in \mathbb{R}^n\right\} \\ = \alpha_0\left[\frac{1}{2\kappa}\|x_0 - y_0\|^2 + f(y_0) + \langle\nabla f(y_0), x_0 - y_0\rangle + \Phi(x_0)\right] \\ \overset{\text{by (4)}}{\ge} \ \frac{1}{2}(\kappa^{-1} - L)\alpha_0\|x_0 - y_0\|^2 + \alpha_0[f(x_0) + \Phi(x_0)].\end{array}$$

That is, (17) holds for $k = 0$. Suppose that (17) holds for some $k \in \mathbb{N}$. We shall show that (17) holds for $k + 1$. Since F_k attains minimum at z_k; $\min_{x\in\mathbb{R}^n} G_k(x) = F_k(z_k) = \min_{x\in\mathbb{R}^n} F_k(x)$ and G_k is strongly convex with parameter $s_k := C\rho + \mu\sum_{i=0}^{k}\alpha_i\gamma_i$, by using the induction assumption, one has for $x \in \mathbb{R}^n$,

$$\begin{array}{l}F_{k+1}(x) = G_k(x) + \alpha_{k+1}[f(y_{k+1}) + \langle\nabla f(y_{k+1}), x - y_{k+1}\rangle + \Phi(x) + \frac{1}{2}\mu\gamma_{k+1}\|x - y_{k+1}\|^2] \\ \ge \min_{x\in\mathbb{R}^n} G_k(x) + \frac{1}{2}s_k\|x - z_k\|^2 + \alpha_{k+1}[f(y_{k+1}) + \langle\nabla f(y_{k+1}), x - y_{k+1}\rangle \\ +\frac{1}{2}\mu\gamma_{k+1}\|x - y_{k+1}\|^2 + \Phi(x)] \\ = \min_{x\in\mathbb{R}^n} F_k(x) + \frac{1}{2}s_k\|x - z_k\|^2 + \alpha_{k+1}[f(y_{k+1}) + \langle\nabla f(y_{k+1}), x - y_{k+1}\rangle \\ +\frac{1}{2}\mu\gamma_{k+1}\|x - y_{k+1}\|^2 + \Phi(x)] \\ \ge \sum_{i=0}^{k}\beta_i[f(x_i) + \Phi(x_i)] + (A_k - B_k)[f(x_k) + \Phi(x_k)] \\ +\frac{1}{2}(1/\kappa - L)\sum_{i=0}^{k}(A_i - B_{i-1})\|x_i - y_i\|^2 \\ +\frac{1}{2}s_k\|x - z_k\|^2 + \alpha_{k+1}[f(y_{k+1}) + \langle\nabla f(y_{k+1}), x - y_{k+1}\rangle + \frac{1}{2}\mu\|x - y_{k+1}\|^2 + \Phi(x)].\end{array} \tag{19}$$

By the convexity of f and Φ,

$$f(x_k) \ge f(y_{k+1}) + \langle\nabla f(y_{k+1}), x_k - y_{k+1}\rangle, \tag{20}$$

and

$$(A_k - B_k)\Phi(x_k) + \alpha_{k+1}\Phi(x) \ge (A_{k+1} - B_k)\Phi(\tau_k x + (1-\tau_k)x_k). \tag{21}$$

Hence, for all $x \in \mathbb{R}^n$,

$$\begin{aligned}&(A_k - B_k)[f(x_k) + \Phi(x_k)] + \tfrac{1}{2}s_k\|x - z_k\|^2 + \alpha_{k+1}[f(y_{k+1}) + \langle \nabla f(y_{k+1}), x - y_{k+1}\rangle + \Phi(x)]\\ &\ge (A_{k+1} - B_k)[f(y_{k+1}) + \tfrac{1}{2}s_k(A_{k+1} - B_k)^{-1}\|x - z_k\|^2\\ &+\tau_k\langle \nabla f(y_{k+1}), x - z_k\rangle + \Phi(\tau_k x + (1-\tau_k)x_k)].\end{aligned} \tag{22}$$

By setting $y := \tau_k x + (1-\tau_k)x_k$, and in view of (16), $s_k(A_{k+1}-B_k)^{-1} \ge \tau_k^2\kappa^{-1}$, the preceding relation implies

$$\begin{aligned}&(A_k - B_k)[f(x_k) + \Phi(x_k)] + \tfrac{1}{2}s_k\|x - z_k\|^2 + \alpha_{k+1}[f(y_{k+1}) + \langle \nabla f(y_{k+1}), x - y_{k+1}\rangle + \Phi(x)]\\ &\ge (A_{k+1} - B_k)[f(y_{k+1}) + \tfrac{1}{2}\kappa^{-1}\tau_k^2\|x - z_k\|^2 + \tau_k\langle \nabla f(y_{k+1}), x - z_k\rangle + \Phi(y)]\\ &= (A_{k+1} - B_k)[f(y_{k+1}) + \tfrac{1}{2}\kappa^{-1}\|y - y_{k+1}\|^2 + \langle \nabla f(y_{k+1}), y - y_{k+1}\rangle + \Phi(y)]\\ &\overset{\text{by (7)}}{\ge} (A_{k+1} - B_k)[f(y_{k+1}) + \tfrac{1}{2}\kappa^{-1}\|x_{k+1} - y_{k+1}\|^2 + \langle \nabla f(y_{k+1}), x_{k+1} - y_{k+1}\rangle + \Phi(x_{k+1})]\\ &\overset{\text{by (4)}}{\ge} (A_{k+1} - B_k)[\tfrac{1}{2}(\kappa^{-1} - L)\|x_{k+1} - y_{k+1}\|^2 + f(x_{k+1}) + \Phi(x_{k+1})].\end{aligned} \tag{23}$$

This estimate together with (19) yield

$$\begin{aligned}\min_{x\in\mathbb{R}^n} F_{k+1}(x) &\ge \textstyle\sum_{i=0}^{k}\beta_i[f(x_i) + \Phi(x_i)] + \tfrac{1}{2}(\kappa^{-1} - L)\sum_{i=0}^{k}(A_i - B_{i-1})\|x_i - y_i\|^2\\ &+(A_{k+1} - B_k)[\tfrac{1}{2}(\kappa^{-1} - L)\|x_{k+1} - y_{k+1}\|^2 + f(x_{k+1}) + \Phi(x_{k+1})]\\ &= \textstyle\sum_{i=0}^{k+1}\beta_i[f(x_i) + \Phi(x_i)] + (A_{k+1} - B_{k+1})[f(x_{k+1}) + \Phi(x_{k+1})]\\ &+\tfrac{1}{2}(\kappa^{-1} - L)\textstyle\sum_{i=0}^{k+1}(A_i - B_{i-1})\|x_i - y_i\|^2.\end{aligned}$$

That is, (17) holds for $k+1$, and it completes the proof of the first part.

Suppose now f is $\mu-$strong convex and $\gamma_k = 1$, $k \in \mathbb{N}$. The proof is the same as above, just a different point is as follows. Instead of (20), by the strongly convexity of f with parameter μ,

$$f(x_k) \ge f(y_{k+1}) + \langle \nabla f(y_{k+1}), x_k - y_{k+1}\rangle + \frac{1}{2}\mu\|x_k - y_{k+1}\|^2. \tag{24}$$

By using this and the inequality

$$\begin{aligned}&(A_k - B_k)\|x_k - y_{k+1}\|^2 + \alpha_{k+1}\|x - y_{k+1}\|^2 \ge (A_{k+1} - B_k)\|x + (1-\tau_k)x_k - y_{k+1}\|^2\\ &= (A_{k+1} - B_k)\tau_k^2\|x - z_k\|^2,\end{aligned}$$

estimate (22) is now changed to

$$\begin{aligned}&(A_k - B_k)[f(x_k) + \Phi(x_k)] + \tfrac{1}{2}s_k\|x - z_k\|^2 + \alpha_{k+1}[f(y_{k+1}) + \langle \nabla f(y_{k+1}), x - y_{k+1}\rangle\\ &+\tfrac{1}{2}\mu\|x - y_{k+1}\|^2 + \Phi(x)] \ge (A_{k+1} - B_k)[f(y_{k+1} + \tfrac{1}{2}s_k(A_{k+1} - B_k)^{-1}\|x - z_k\|^2\\ &+\tau_k\langle \nabla f(y_{k+1}), x - z_k\rangle + \Phi(\tau_k x + (1-\tau_k)x_k)]\\ &+\mu[(A_k - B_k)\|x_k - y_{k+1}\|^2 + \alpha_{k+1}\|x - y_{k+1}\|^2]\\ &\ge (A_{k+1} - B_k)[f(y_{k+1}) + \tfrac{1}{2}[s_k(A_{k+1} - B_k)^{-1} + \mu\tau_k^2]\|x - z_k\|^2\\ &+\tau_k\langle \nabla f(y_{k+1}), x - z_k\rangle + \Phi(\tau_k x + (1-\tau_k)x_k)]\\ &\overset{\text{by (18)}}{\ge} (A_{k+1} - B_k)[f(y_{k+1}) + \tfrac{1}{2\kappa}\tau_k^2\|x - z_k\|^2\\ &+\tau_k\langle \nabla f(y_{k+1}), x - z_k\rangle + \Phi(\tau_k x + (1-\tau_k)x_k)],\end{aligned} \tag{25}$$

where $s_k := C\rho + \mu \sum_{i=0}^{k} \alpha_i = C\rho + \mu A_k$. The remain estimates are the same as before, by using (25) instead of (22), and condition (18). □

Corollary 1 *In Algorithm 1, pick $\alpha_k = k$; $\beta_k = k/2$; $\mu = 0$, and $C, \kappa > 0$ such that $C\rho \geq \kappa^{-1} \geq L$. Then condition (16) is satisfied, and therefore for a minimizer x^* of problem (1), one has*

$$\begin{array}{l} \frac{1}{2}\sum_{i=0}^{k} i[f(x_i) + \Phi(x_i) - f(x^*) - \Phi(x^*)] \\ +\frac{1}{4}k(k+1)[f(x_k) + \Phi(x_k) - f(x^*) - \Phi(x^*)] \\ +\frac{1}{8}(1/\kappa - L)\sum_{i=0}^{k} i(3i-1)\|x_i - y_i\|^2 \leq Ch(x^*). \end{array} \tag{26}$$

As a result,

$$\lim_{k\to\infty} \min_{i=[k/2],\ldots,k} k^2[f(x_i) + \Phi(x_i) - f(x^*) - \Phi(x^*)] = 0, \tag{27}$$

where $[k/2]$ stands for the integer part of $k/2$. Therefore if $\{f(x_k) + \Phi(x_k)\}$ is a decreasing sequence, then

$$\lim_{k\to\infty} k^2[f(x_k) + \Phi(x_k) - f(x^*) - \Phi(x^*)] = 0. \tag{28}$$

Proof. By checking directly, we see that (16) is satisfied for $\alpha_k = k$, $\beta_k = k/2$, $\mu = 0$, and $C\rho \geq \kappa^{-1} \geq L$. Hence, by set $x = x^*$ in (17), then using the convexity of f, we obtain (26). This relation implies

$$\sum_{i=0}^{\infty} i[f(x_i) + \Phi(x_i) - f(x^*) - \Phi(x^*)] < +\infty.$$

Therefore

$$\lim_{k\to\infty} \sum_{i=[k/2]}^{k} i[f(x_i) + \Phi(x_i) - f(x^*) - \Phi(x^*)] = 0.$$

One has

$$\begin{array}{l} \sum_{i=[k/2]}^{k} i[f(x_i) + \Phi(x_i) - f(x^*) - \Phi(x^*)] \\ \geq \min_{i=[k/2],\ldots,k} k^2[f(x_i) + \Phi(x_i) - f(x^*) - \Phi(x^*)] \sum_{i=[k/2]}^{k} \frac{i}{k^2} \\ \geq \min_{i=[k/2],\ldots,k} k^2[f(x_i) + \Phi(x_i) - f(x^*) - \Phi(x^*)] \frac{k(3k+2)}{8k^2} \\ \geq \frac{3}{8} \min_{i=[k/2],\ldots,k} k^2[f(x_i) + \Phi(x_i) - f(x^*) - \Phi(x^*)], \end{array}$$

which shows (27). □

We consider the case where f is $p-$uniformly convex, $p > 2$, with parameter μ, or called $(\mu, p)-$unformly convex.

Corollary 2 *Let f is $(\mu,p)-$uniformly convex with $p>2$, $\mu>0$. Let $0<\kappa\leq L^{-1}$, and $C,\rho,m>0$ such that*

$$m\mu\kappa \geq \begin{cases} 2^{\frac{4}{p-2}}\frac{8p}{(p-2)^2} & \text{if } 2<p<6,\\ \frac{8p}{(p-2)^2} & \text{if } p\geq 6; \end{cases} \tag{29}$$

$$C\rho \geq \begin{cases} \kappa^{-1} & \text{if } 2<p<6,\\ \frac{p-2}{4}m\mu & \text{if } p\geq 6. \end{cases} \tag{30}$$

In Algorithm 1, set $\alpha_k = k^{\frac{p+2}{p-2}}$, $\beta_k=0$, and $\gamma_0=0$, $\gamma_k = mk^{-2}$ for $k\geq 1$. Then (16) is satisfied and for x^* being a minimizer of $f+\Phi$, and therefore one has for all $k\in\mathbb{N}$,

$$f(x_k)+\Phi(x_k)-f(x^*)-\Phi(x^*) \leq \frac{2p}{p-2}(Ch(x^*)+\frac{1}{2}(p/2)^{\frac{2}{p-2}}m^{\frac{p}{p-2}}(\ln k+1)k^{-\frac{2p}{p-2}}. \tag{31}$$

Proof. By using the inequalities

$$\sum_{i=1}^{k} i^\alpha \geq \sum_{i=0}^{k-1}\int_i^{i+1} x^\alpha dx = \frac{1}{\alpha+1}k^{\alpha+1}, \tag{32}$$

if $\alpha>0$ and if $-1<\alpha\leq 0$,

$$\sum_{i=1}^{k} i^\alpha \geq \sum_{i=1}^{k}\int_i^{i+1} x^\alpha dx = \frac{1}{\alpha+1}[(k+1)^{\alpha+1}-1], \tag{33}$$

one has for $k\geq 1$,

$$(C\rho+\mu\sum_{i=0}^{k-1}\alpha_i\gamma_i)A_k \geq \frac{p-2}{2p}k^{\frac{2p}{p-2}}[C\rho+\frac{p-2}{4}m\mu(k-1)^{\frac{4}{p-2}}],$$

if $2<p<6$, and if $p\geq 6$,

$$(C\rho+\mu\sum_{i=0}^{k-1}\alpha_i\gamma_i)A_k \geq \frac{p-2}{2p}k^{\frac{2p}{p-2}}[C\rho+\frac{p-2}{4}m\mu(k^{\frac{4}{p-2}}-1)].$$

By virtue of these two inequalities, it is easy to check directly the valid of (16). For $k\in\mathbb{N}^*$, let define

$$J_k := \left\{ i\in\{1,...,k\}:\ \ \|y_k-x^*\| \leq (mp/2)^{\frac{1}{p-2}}k^{-\frac{2}{p-2}} \right\}.$$

Then

$$\begin{aligned}\textstyle\sum_{i\in J_k}\alpha_i\gamma_i\|y_i-x^*\|^2 &\leq (p/2)^{\frac{2}{p-2}}m^{\frac{p}{p-2}}\textstyle\sum_{i=1}^k i^{\frac{6-p}{p-2}}i^{-\frac{4}{p-2}}\\ &= (p/2)^{\frac{2}{p-2}}m^{\frac{p}{p-2}}\textstyle\sum_{i=1}^k i^{-1} \leq (p/2)^{\frac{2}{p-2}}m^{\frac{p}{p-2}}(\ln k+1),\end{aligned} \tag{34}$$

where the last inequality follows from the one

$$\sum_{i=1}^{k} i^{-1} \le 1 + \sum_{i=2}^{k} \int_{i-1}^{i} x^{-1} dx = 1 + \ln k.$$

For $i \in \{1, ..., k\} \setminus J_k$, then $\|y_k - x^*\| > (mp/2)^{\frac{1}{p-2}} k^{-\frac{2}{p-2}}$, therefore

$$\begin{aligned} \tfrac{1}{p}\alpha_i \|y_i - x^*\|^p &= \tfrac{1}{p}\alpha_i \|y_i - x^*\|^{p-2} \|y_i - x^*\|^2 \\ &\ge \tfrac{1}{p}\alpha_i (mp/2) k^{-2} \|y_i - x^*\|^2 = \tfrac{1}{2}\alpha_i \gamma_i \|y_i - x^*\|^2. \end{aligned} \tag{35}$$

From the latter two relations, in view of (17), setting $x = x^*$, we derive that the following estimate

$$\begin{aligned}
&A_k[f(x_k) + \Phi(x_k)] \le F_k(x^*) \\
&\le Ch(x^*) + \textstyle\sum_{i=0}^{k-1} \alpha_i [f(y_i) + \langle \nabla f(y_i), x^* - y_i \rangle + \Phi(x^*) + \frac{1}{2}\mu\gamma_i \|x^* - y_i\|^2] \\
&\le Ch(x^*) + \textstyle\sum_{j \in J_k} \frac{1}{2}\mu\alpha_i\gamma_i \|x^* - y_i\|^2 + \sum_{i \in J_k} \alpha_i [f(y_i) + \langle \nabla f(y_i), x^* - y_i \rangle + \Phi(x^*)] \\
&+ \textstyle\sum_{i=0,\, i \notin J_k}^{k} \alpha_i [f(y_i) + \langle \nabla f(y_i), x^* - y_i \rangle + \Phi(x^*) + \frac{1}{2}\mu\gamma_i \|x^* - y_i\|^2] \\
&\le Ch(x^*) + \tfrac{1}{2}(p/2)^{\frac{2}{p-2}} m^{\frac{p}{p-2}} (\ln k + 1) + \textstyle\sum_{i \in J_k} \alpha_i [f(x^*) + \Phi(x^*)] \\
&+ \textstyle\sum_{i=0,\, i \notin J_k}^{k} \alpha_i [f(y_i) + \langle \nabla f(y_i), x^* - y_i \rangle + \Phi(x^*) + \frac{1}{p}\mu \|x^* - y_i\|^p] \\
&\overset{\text{by (3)}}{\le} Ch(x^*) + \tfrac{1}{2}(p/2)^{\frac{2}{p-2}} m^{\frac{p}{p-2}} (\ln k + 1) \\
&+ \textstyle\sum_{i \in J_k} \alpha_i [f(x^*) + \Phi(x^*)] + \sum_{i=0,\, i \notin J_k}^{k} [f(x^*) + \Phi(x^*)] \\
&\le Ch(x^*) + \tfrac{1}{2}(p/2)^{\frac{2}{p-2}} m^{\frac{p}{p-2}} (\ln k + 1) + A_k[f(x^*) + \Phi(x^*)].
\end{aligned}$$

By noting from (32) that

$$A_k = \sum_{i=1}^{k} k^{\frac{p+2}{p-2}} \ge k^{\frac{2p}{p-2}},$$

one has (31) and the proof is completed. □

Next we consider the case where f is $\mu-$strongly convex for $\mu > 0$. For sequences $\alpha_k := q^k$ (for some $q > 1$) and $\beta_k = 0$, $k \in \mathbb{N}$, relation (18) becomes

$$\left(C\rho + \mu \frac{q^k - 1}{q - 1} \right) \frac{q^{k+1} - 1}{q - 1} \ge q^{2k}(\kappa^{-1} - \mu), \quad \forall k \in \mathbb{N}.$$

Equivalently,

$$\mu \frac{q^{2k+1}}{(q-1)^2} + C\rho \frac{q^{k+1} - 1}{q - 1} - \mu \frac{q^{k+1} - 1}{(q-1)^2} \ge q^{2k}(\kappa^{-1} - \mu), \quad \forall k \in \mathbb{N}.$$

If we take $C > 0$ such that $C\rho(q-1) - \mu \ge 0$ as well as

$$C\rho \frac{q^{k+1} - 1}{q - 1} - \mu \frac{q^{k+1} - 1}{(q-1)^2} \ge (C\rho(q-1) - \mu) \frac{q^{k+1} - 1}{(q-1)^2} \ge 0 \ \ \forall k \in N.$$

Then, relation (18) holds if

$$\mu\frac{q^{2k+1}}{(q-1)^2} \geq q^{2k}(\kappa^{-1}-\mu), \quad \forall k \in \mathbb{N},$$

equivalently

$$\frac{\mu q}{(q-1)^2} \geq \kappa^{-1}-\mu.$$

Hence in summary, (18) holds for $\alpha_k = q^k$, $\beta_k = 0$ with

$$q = \frac{2\kappa^{-1}-\mu+\sqrt{4\kappa^{-1}\mu-3\mu^2}}{2(\kappa^{-1}-\mu)} \quad \text{and} \quad C\rho \geq \frac{\mu}{q-1}. \tag{36}$$

So one obtains the following corollary for the linear convergence of Algorithm 1 in the case of strong convexity.

Corollary 3 *Let f is $\mu-$strongly convex for some $\mu > 0$, and let q, C such as (36). Then for the sequence $\{x_k\}$ generated by Algorithm 1 with sequences $\alpha_k := q^k$, $\beta_k = 0$, and $\gamma_k = 1$, $k \in \mathbb{N}$, and a minimizer x^* of problem (1), one has*

$$f(x_k) + \Phi(x_k) - f(x^*) - \Phi(x^*) \leq \frac{(q-1)Ch(x^*)}{q^{k+1}-1}, \quad \textit{for all } k \in \mathbb{N}. \tag{37}$$

Proof. Relation (37) follows directly from (17) by noticing that as f is $\mu-$strongly convex, for all $i = 1, ..., k$,

$$f(y_i) + \langle \nabla f(y_i), x - y_i \rangle + \mu\|x - y_i\|^2 \leq f(x), \quad \forall x \in \mathbb{R}^n,$$

therefore, $F_k(x) \leq Ch(x) + A_k(f(x) + \Phi(x))$, for all $x \in \mathbb{R}^n$. □

Note that the linear convergence of the standard gradient method and Nesterov's accelerated schemes in the case of strongly convexity was well established in the literature (see e.g., [15]). Alternatively, in the papers [6,9], some geometric descent methods with linear convergence rates for minimizing smooth strongly convex functions have been proposed. Then in [8], this method has been generalized for convex composite minimization of the form (1). More recently, some results on the linear convergence of several first-order methods for smooth convex optimization problems in which the objective function is not necessarily strongly convex, have been derived in [19].

3 Generalized accelerated forward-backward algorithm

In [1], the authors have considered the following accelerated forward-backward scheme for solving (1):

$$\begin{cases} y_k = x_k + \frac{k-1}{k+\alpha-1}(x_k - x_{k-1}), \\ x_{k+1} = \text{prox}_{\kappa\Phi}(y_k - \kappa\nabla f(y_k)), \end{cases} \tag{38}$$

where $\alpha > 0, \kappa > 0$. The authors have established the rate of convergence of order $o(1/k^2)$ when $\alpha > 3$ and $\kappa \leq 1/L$, when ∇f is assumed to be $L-$Lipschitz on the whole space $\mathbb{R}^n$.
By considering the operator $G_\kappa : \mathbb{R}^n \to \mathbb{R}^n$, defined by

$$G_\kappa(y) = \frac{1}{\kappa}[y - \mathrm{prox}_{\kappa\Phi}(y - \kappa\nabla f(y))], \ \ y \in \mathbb{R}^n,$$

and setting

$$z_k = \frac{k+\alpha-1}{\alpha-1}y_k - \frac{k}{\alpha-1}x_k,$$

then we can rewrite the scheme (38) as follows.

$$\begin{cases} z_{k+1} = z_k - \frac{\kappa(k+\alpha-1)}{\alpha-1}G_\kappa(y_k), \\ y_k = \frac{\alpha-1}{k+\alpha-1}z_k + \frac{k}{k+\alpha-1}x_k, \\ x_{k+1} = \mathrm{prox}_{\kappa\Phi}(y_k - \kappa\nabla f(y_k)). \end{cases} \tag{39}$$

Obviously, the sequence $\{z_k\}$ in the scheme (39) can be represented equivalently

$$z_{k+1} = \mathrm{argmin}_{x\in\mathbb{R}^n}\left\{\frac{1}{2\kappa}\|x\|^2 + \sum_{i=0}^{k}\alpha_i\langle G_\kappa(y_i), x\rangle\right\},$$

where $\alpha_i = \frac{i+\alpha-1}{\alpha-1}$, for $i \in \mathbb{N}$. In view of this representation, we propose the generalized accelerated forward-backward algorithm: Given a $\rho-$strongly convex function $h : \mathbb{R}^n \to \mathbb{R}$ $(\rho > 0)$ as before; parameters $C, \mu > 0$, $0 < \kappa \leq 1/L$, and a sequence of positive reals $\{\alpha_k\}$; sequences of nonnegative reals $\{\beta_k\}$, and $\{\gamma_k\}$ as in Section 2. Set

$$A_k = \sum_{i=0}^{k}\alpha_k, \ B_k = \sum_{i=0}^{k}\beta_k,$$

and also assume that $A_k \geq B_k$ for all $k \in \mathbb{N}$, and denote $A_{-1} = B_{-1} = 0$.

In this section, in assumption (A2), instead of the $L-$Lipschitz continuity of ∇f on $\mathrm{dom}\,\Phi$, we assume that

(H) *The gradient ∇f is $L-$Lipschitz on the whole space $\mathbb{R}^n$.*

Algorithm 2: Generalized accelerated forward-backward algorithm (GAFBA)

`Initialization:` Initial data: $x_0 = z_0 = y_0 \in \mathbb{R}^n$, with y_0 as in (6). Set $k = 0$.

`Repeat:` For $k = 0, 1, ...,$

1. Set
$$\tau_k := \frac{\alpha_k}{A_k - B_{k-1}}, \ \ y_k = \tau_k z_k + (1-\tau_k)x_k.$$
2. Find
$$x_{k+1} = \mathrm{prox}_{\kappa\Phi}(y_k - \kappa\nabla f(y_k)). \tag{40}$$

4. Set

$$G_\kappa(y_k) = \frac{1}{\kappa}[y_k - \mathrm{prox}_{\kappa\Phi}(y_k - \kappa\nabla f(y_k))] = \frac{1}{\kappa}(y_k - x_{k+1}).$$

3. Find

$$z_{k+1} = \mathrm{argmin}\{Ch(x) + \sum_{i=0}^{k} \alpha_i[\langle G_\kappa(y_i), x - y_i\rangle + \frac{1}{2}\mu\gamma_i\|x - y_i\|^2] : \ x \in \mathbb{R}^n\} \tag{41}$$

By a straightforward computation, scheme (38) with $\alpha > 3$, is a particular case of Algorithm 2 with $h(x) = \frac{1}{2}\|x\|^2$, $C = \kappa^{-1}$, $\mu = 0$ and

$$\alpha_k = \frac{k + \alpha - 1}{\alpha - 1}, \ \ \beta_k = \frac{(\alpha - 3)(2k + 1)}{2(\alpha - 1)^2} + 2\alpha - 5, \ \ k \in \mathbb{N}. \tag{42}$$

Let us introduce the following functions E_k, $k \in \mathbb{N}$, which plays a role of an "estimating function" for Algorithm 2, as the one of the functions F_k for Algorithm 1.

$$E_k(x) = Ch(x) + \sum_{i=0}^{k} \alpha_i[f(x_{i+1}) + \Phi(x_{k+1}) + \langle G_\kappa(y_i), x - y_i\rangle + \frac{\kappa}{2}\|G_\kappa(y_i)\|^2 + \frac{1}{2}\mu\gamma_i\|x - y_i\|^2]. \tag{43}$$

The following property of the operator G_κ (see e.g., [1,4,21]) plays a key role in the proof of the convergence result,

$$(f + \Phi)(y - \kappa G_\kappa(y)) + \langle G_\kappa(y), x - y\rangle \le (f + \Phi)(x) - \frac{\kappa}{2}\|G_\kappa(y)\|^2, \ \ \forall x, y \in \mathbb{R}^n. \tag{44}$$

More generally, either f or Φ is $(\mu, p)-$uniformly convex for some $\mu \ge 0, p \ge 2$, one has

Lemma 1 *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is constinuously differentiable with $L-$Lipschitz continuous gradient on $\mathbb{R}^n$ and $0 < \kappa \le 1/L$. then for $\bar{y} = y - \kappa G_\kappa(y) = prox_{\kappa\Phi}(y - \kappa\nabla f(y))$, one has*

(i) If f is $(\mu, p)-$uniformly convex, then

$$(f+\Phi)(\bar{y}) + \langle G_\kappa(y), x - y\rangle + \frac{\mu}{p}\|x - y\|^p \le (f+\Phi)(x) - \frac{\kappa}{2}\|G_\kappa(y)\|^2, \ \ \forall x, y \in \mathbb{R}^n. \tag{45}$$

(ii) If Φ is $(\mu, p)-$uniformly convex, then

$$(f+\Phi)(\bar{y}) + \langle G_\kappa(y), x - y\rangle + \frac{\mu}{p}\|x - \bar{y}\|^p \le (f+\Phi)(x) - \frac{\kappa}{2}\|G_\kappa(y)\|^2, \ \ \forall x, y \in \mathbb{R}^n. \tag{46}$$

Proof. As

$$\bar{y} = \text{prox}_{\kappa\Phi}(y-\kappa\nabla f(y)) = \text{argmin}\{\Phi(x)+\langle\nabla f(y), x-y\rangle+\frac{1}{2\kappa}\|x-y\|^2 : \ \ x\in\mathbb{R}^n\},$$

one has

$$-\kappa^{-1}(\bar{y}-y)-\nabla f(y)\in\partial\Phi(\bar{y}). \tag{47}$$

Firstly for part (i), this relation implies

$$\langle -\kappa^{-1}(\bar{y}-y)-\nabla f(y), x-\bar{y}\rangle \le \Phi(x)-\Phi(\bar{y}), \ \ \forall x\in\mathbb{R}^n.$$

Equivalently, for $x\in\mathbb{R}^n$,

$$\begin{aligned}&(f+\Phi)(\bar{y})+\langle G_\kappa(y), x-y\rangle+\tfrac{\kappa}{2}\|G_\kappa(y)\|^2\\ &\le \Phi(x)+[f(y)+\langle\nabla f(y), x-y\rangle]+[f(\bar{y})-f(y)-\langle\nabla f(y),\bar{y}-y\rangle-\tfrac{1}{2\kappa}\|\bar{y}-y\|^2].\end{aligned}$$

Relation (45) follows directly from this relation, since f is $(\mu,p)-$uniformly convex,

$$f(y)+\langle\nabla f(y), x-y\rangle\le f(x)-\frac{\mu}{p}\|x-y\|^p,$$

and as ∇f is $L-$Lipschitz continuous,

$$f(\bar{y})-f(y)-\langle\nabla f(y),\bar{y}-y\rangle-\frac{1}{2\kappa}\|\bar{y}-y\|^2\le 0.$$

For (ii), Φ is $(\mu,p)-$uniformly convex, (47) implies

$$\langle -\kappa^{-1}(\bar{y}-y)-\nabla f(y), x-\bar{y}\rangle\le\Phi(x)-\Phi(\bar{y})-\frac{\mu}{p}\|x-\bar{y}\|^2, \ \ \forall x\in\mathbb{R}^n,$$

and as before, equivalently,

$$\begin{aligned}&(f+\Phi)(\bar{y})+\langle G_\kappa(y), x-y\rangle+\tfrac{\kappa}{2}\|G_\kappa(y)\|^2\\ &\le\Phi(x)-\tfrac{\mu}{p}\|x-\bar{y}\|^2+[f(y)+\langle\nabla f(y), x-y\rangle]\\ &+[f(\bar{y})-f(y)-\langle\nabla f(y),\bar{y}-y\rangle-\tfrac{\kappa}{2}\|\bar{y}-y\|^2],\end{aligned}$$

which implies (46) by the convexity of f, as well as the $L-$Lipschitz continuity of ∇f. □

We are now ready to state the convergence result of Algorithm 2.

Theorem 2 *Let $\{x_k\}$ be the sequences defined by Algorithm 2. Suppose that $\kappa\le 1/L$ and the sequences $\{\alpha_k\}$, $\{\beta_k\}$ and $\{\gamma_k\}$ satisfy the condition (16) in Theorem 1. Then for all $k\in\mathbb{N}$,*

$$\sum_{i=0}^{k}\beta_i[f(x_{i+1})+\Phi(x_{i+1})]+(A_k-B_k)[f(x_{k+1})+\Phi(x_{k+1})]\le\min_{x\in\mathbb{R}^n}E_k(x). \tag{48}$$

where, we set $B_{-1}=0$. Moreover, if f is $\mu-$strong convex, then (48) holds if $\gamma_k=1$, $k\in\mathbb{N}$, and the sequences $\{\alpha_k\}$, $\{\beta_k\}$ verifying the condition (18) in Theorem 1.

As results, for a minimizer x^* of problem (1), one has *(i) For $\mu = 0$, and any two sequences of positive reals $\{\alpha_k\}$ and $\{\beta_k\}$ with $\alpha_k \geq \beta_k$ for $k \in \mathbb{N}$ and*

$$0 < \liminf_{k\to\infty} \frac{\beta_k}{k} \leq \limsup_{k\to\infty} \frac{\alpha_k}{k} < +\infty, \quad \limsup_{k\to\infty} \frac{\beta_k}{\alpha_k} < 1,$$

then we can find $C_0 > 0$ satisfying the condition

$$C_0\rho(A_k - B_{k-1}) \geq \alpha_k \kappa^{-1}, \quad \forall k \in \mathbb{N}, \tag{49}$$

and therefore for all $C \geq C_0$, for the sequence $\{x_k\}$ generated by Algorithm 2, one has

$$\lim_{k\to\infty} k^2 \min_{i=[k/2],\ldots,k} [f(x_i) + \Phi(x_i) - f(x^*) - \Phi(x^*)] = 0.$$

(ii) Suppose that f is $(\mu, p)-$uniformly convex with $\mu > 0$, $p > 2$. Then with the same conditions as in Corollary 2, one has

$$f(x_k) + \Phi(x_k) - f(x^*) - \Phi(x^*) = O\left(\frac{\ln k}{k^{2p/(p-2)}}\right).$$

(iii) If f is $\mu-$strongly convex, then with $q > 1$ $C > 0$ as in Corollary 3, and the sequences $\alpha_k = q^k$, $\beta_k = 0$ and $\gamma_k = 1$, for $k \in \mathbb{N}$, one has

$$f(x_k) + \Phi(x_k) - f(x^*) - \Phi(x^*) = O\left(q^{-k}\right).$$

Proof. Similarly to the proof of Theorem 1, we prove (48) by induction on $k \in \mathbb{N}$. For $k = 0$, since $C\rho + \alpha_0\mu\gamma_0 \geq \kappa^{-1}$, one has

$$\begin{aligned} E_0(x) &= Ch(x) + \alpha_0[f(x_1) + \Phi(x_1) + \langle G_\kappa(y_0), x - y_0\rangle + \tfrac{\kappa}{2}\|G_\kappa(y_0)\|^2 + \tfrac{1}{2}\mu\gamma_0\|x - y_0\|^2] \\ &\geq \tfrac{1}{2}(C\rho + \alpha_0\mu\gamma_0)\|x - y_0\|^2 + \alpha_0[f(x_1) + \Phi(x_1) + \langle G_\kappa(y_0), x - y_0\rangle + \tfrac{\kappa}{2}\|G_\kappa(y_0)\|^2] \\ &\geq \alpha_0(f(x_1) + \Phi(x_1) + \tfrac{\kappa}{2}[\kappa^{-1}(x - y_0) - G_\kappa(y_0)]^2 \geq \alpha_0(f(x_1) + \Phi(x_1)), \end{aligned}$$

for all $x \in \mathbb{R}^n$, showing (48) holds for $k = 0$. Assuming (48) holds for $k-1 \in \mathbb{N}$, we will show that it holds for k. As $z_k = \operatorname{argmin}_{x\in\mathbb{R}^n} E_{k-1}(x)$, since E_{k-1} is $(C\rho + \mu\sum_{i=0}^{k-1}\alpha_i\gamma_i)-$strongly convex, one has

$$E_{k-1}(x) \geq \min_{x\in\mathbb{R}^n} E_{k-1}(x) + \frac{s_{k-1}}{2}\|x - z_k\|^2,$$

for $x \in \mathbb{R}^n$, where $s_{k-1} = C\rho + \mu\sum_{i=0}^{k-1}\alpha_i\gamma_i$, which implies

$$\begin{aligned} E_k(x) &= E_{k-1}(x) + \alpha_k[f(x_{k+1}) + \Phi(x_{k+1}) + \langle G_\kappa(y_k), x - y_k\rangle + \tfrac{\kappa}{2}\|G_\kappa(y_k)\|^2 + \tfrac{1}{2}\mu\gamma_k\|x - y_k\|^2] \\ &\geq \textstyle\sum_{i=0}^{k-1}\beta_i[f(x_{i+1}) + \Phi(x_{i+1})] + (A_{k-1} - B_{k-1})[f(x_k) + \Phi(x_k)] + \tfrac{s_{k-1}}{2}\|x - z_k\|^2 \\ &\quad + \alpha_k[f(x_{k+1}) + \Phi(x_{k+1}) + \langle G_\kappa(y_k), x - y_k\rangle + \tfrac{\kappa}{2}\|G_\kappa(y_k)\|^2 + \tfrac{1}{2}\mu\gamma_k\|x - y_k\|^2] \end{aligned} \tag{50}$$

In view of inequality (44), noticing $x_{k+1} = y_k - \kappa G_\kappa(y_k)$,

$$f(x_k) + \Phi(x_k) \geq f(x_{k+1}) + \Phi(x_{k+1}) + \langle G_\kappa(y_k), x_k - y_k\rangle + \frac{\kappa}{2}\|G_\kappa(y_k)\|^2,$$

therefore (50) implies

$$\begin{aligned}
E_k(x) &\geq \textstyle\sum_{i=0}^k \beta_i[f(x_{i+1})+\varPhi(x_{i+1})]+(A_k-B_k)[f(x_{k+1})+\varPhi(x_{k+1})]\\
&\quad +(A_k-B_{k-1})[\tfrac{s_{k-1}}{2(A_k-B_{k-1})}\|x-z_k\|^2+\langle G_\kappa(y_k),\tau_k x+(1-\tau_k)x_k-y_k\rangle+\tfrac{\kappa}{2}\|G_\kappa(y_k)\|^2]\\
&\overset{\text{by (16)}}{\geq} \textstyle\sum_{i=0}^k \beta_i[f(x_{i+1})+\varPhi(x_{i+1})]+(A_k-B_k)[f(x_{k+1})+\varPhi(x_{k+1})]\\
&\quad +(A_k-B_{k-1})[\tfrac{\kappa^{-1}\tau_k^2}{2}\|x-z_k\|^2+\tau_k\langle G_\kappa(y_k),x-z_k\rangle+\tfrac{\kappa}{2}\|G_\kappa(y_k)\|^2]\\
&= \textstyle\sum_{i=0}^k \beta_i[f(x_{i+1})+\varPhi(x_{i+1})]+(A_k-B_k)[f(x_{k+1})+\varPhi(x_{k+1})]+\tfrac{\kappa}{2}\|\kappa^{-1}\tau_k(x-z_k)-G_\kappa(y_k)\|^2\\
&\geq \textstyle\sum_{i=0}^k \beta_i[f(x_{i+1})+\varPhi(x_{i+1})]+(A_k-B_k)[f(x_{k+1})+\varPhi(x_{k+1})],
\end{aligned}$$

showing (48) holds for k.

For (i), with the assumptions on $\{\alpha_k\}$, $\{\beta_k\}$, there are $0<\eta_1<\eta_2<\eta_3<\eta_4$ and $r_1,r_2,r_3,r_4\in\mathbb{R}$ such that for all k sufficiently large, one has

$$\eta_1 k+r_1\leq\beta_k\leq\eta_2 k+r_2,\;\; \eta_3 k+r_3\leq\alpha_k\leq\eta_4 k+r_4.$$

Hence, for k sufficiently large,

$$A_k-B_{k-1}\geq\eta_3\frac{k(k+1)}{2}+kr_3-\eta_2\frac{k(k-1)}{2}-(k-1)r_2=O(k^2),$$

and $\alpha_k^2\leq(\eta_4 k+r_4)^2=O(k^2)$, so we can find out $C_0>0$ such that

$$C_0\rho(A_k-B_{k-1})\geq\alpha_k^2\kappa^{-1}.$$

That is, condition (16) is satisfied for all $C\geq C_0$. Next by inequality (44), for x^* being a minimizer of problem (1), one has

$$\begin{aligned}
E_k(x^*) &= Ch(x^*)+\textstyle\sum_{i=0}^k \alpha_i[f(x_{i+1})+\varPhi(x_{k+1})+\langle G_\kappa(y_i),x^*-y_i\rangle+\tfrac{\kappa}{2}\|G_\kappa(y_i)\|^2]\\
&\leq Ch(x^*)+A_k(f(x^*)+\varPhi(x^*)).
\end{aligned}$$

Therefore, (48) implies

$$\sum_{i=0}^{\infty}\beta_i(f(x_{i+1})+\varPhi(x_{i+1}))<+\infty,$$

and since $\beta_i=O(i)$ as $i\to\infty$, with the same argument as in the proof Corollary 1, one derives the conclusion of part (i).

For (ii) and (iii), from (48), invoking the inequality (45), with the same arguments as in the proofs of Corollaries 2 and 3, respectively, one derives the desired conclusions □

Note that for the scheme (38), then $\rho=1$; the sequences α_k and β_k are defined as (42), one has $A_k-B_{k-1}=\alpha_k^2$, for all $k\in\mathbb{N}$, so we can take $C_0=\kappa^{-1}$. Generally, we are going to consider Algorithm 2 when $h(x):=\frac{1}{2}\|x-y_0\|^2$, $x\in\mathbb{R}^n$; $\mu=0$, and sequences $\{\alpha_k\}$ and $\{\beta_k\}$ satisfying the condition

$$A_k-B_{k-1}=\alpha_k^2,\;\; k\in\mathbb{N}.$$

In this case, $\tau_k = 1/\alpha_k$, moreover (16) is verified for $C := \kappa^{-1}$, and the formula of z_k can be represented equivalently,

$$z_{k+1} = z_k - \kappa\alpha_k G_\kappa(y_k), \quad k \in \mathbb{N}. \tag{51}$$

Hence, recalling $y_k - x_{k+1} = \kappa G_\kappa(y_k)$, and $z_k = (y_k - (1-\tau_k)x_k)\tau_k^{-1}$, y_{k+1} can be rewritten as

$$\begin{aligned} y_{k+1} &= \tau_{k+1}z_{k+1} + (1-\tau_{k+1})x_{k+1} \\ &= \tau_{k+1}[(y_k - (1-\tau_k)x_k)\tau_k^{-1} - \alpha_k(y_k - x_{k+1})] + (1-\tau_{k+1})x_{k+1} \\ &= x_{k+1} + \tfrac{\tau_{k+1}(1-\tau_k)}{\tau_k}(x_{k+1} - x_k) \\ &= x_{k+1} + \tfrac{\alpha_k - 1}{\alpha_{k+1}}(x_{k+1} - x_k) \end{aligned}$$

Thus Algorithm 2 can be rewritten simply in the following scheme generalizing (38):

$$\begin{cases} y_k = x_k + \frac{\alpha_{k-1}-1}{\alpha_k}(x_k - x_{k-1}), \\ x_{k+1} = \mathrm{prox}_{\kappa\Phi}(y_k - \kappa\nabla f(y_k)). \end{cases} \tag{52}$$

For this scheme, we establish the following convergence result which generalizes Theorems 1 and 3 in [1].

Theorem 3 *Let $\{\alpha_k\}$, $\{\beta_k\}$ be sequences of positive reals such that for some $0 < c_1, c_2 < 1$,*

$$c_1\alpha_k \le \beta_k \le c_2\alpha_k, \quad A_k - B_{k-1} = \alpha_k^2, \quad k \in \mathbb{N}. \tag{53}$$

Consider Algorithm 2 with $h(x) := \frac{1}{2}\|x - y_0\|^2$; $C = \kappa^{-1} \ge L$, and $\mu = 0$, or equivalently the scheme (52). Then one has

$$\lim_{k\to\infty} k^2[f(x_k)+\Phi(x_k)-f(x^*)-\Phi(x^*)] = 0 \qquad \text{and} \qquad \lim_{k\to\infty} k\|x_{k+1}-x_k\| = 0, \tag{54}$$

that is, $f(x_k) + \Phi(x_k) - f(x^*) - \Phi(x^*) = o(k^{-2})$ and $\|x_{k+1} - x_k\| = o(k^{-1})$, where x^* is a minimizer of problem (1). Moreover, the whole sequence $\{x_k\}$ converges to a minimizer of problem (1).

The following lemma is needed.

Lemma 2 *Suppose that sequences of positive reals $\{\alpha_k\}$, $\{\beta_k\}$ satisfy (53). Then $\{\alpha_k\}$ is an increasing sequence and there are $0 < a_1 < a_2$, and $b > 0$ such that*

$$a_1 k \le \alpha_k \le a_2 k + b, \quad \text{for all } k \in \mathbb{N}.$$

Proof. Since $\alpha_k > 0$ and $A_k - B_{k-1} = \alpha_k^2$,

$$\alpha_k = \frac{1}{2}(1 + \sqrt{1 + 4(A_{k-1} - B_{k-1})}), \quad k \in \mathbb{N}.$$

As $0 \le B_{k-1} \le c_2 A_{k-1}$, it implies that

$$\frac{1}{2}\sqrt{(1-c_2)A_{k-1}} \le \alpha_k \le \frac{1}{2}(1 + \sqrt{1 + 4A_{k-1}}), \quad k \in \mathbb{N}.$$

We prove the lemma by induction. Pick

$0 < a_1 < \min\{\alpha_1, \alpha_2/2, (1-c_2)/3\}$ and $b \geq \max\{\alpha_0, \alpha_1, \alpha_2, 16\}$, $a_2 > 2b$.

Obviously, (2) holds for $k = 0, 1, 2$. Assume that (2) holds for $k-1 \geq 2$. Then $a_1 \frac{k(k-1)}{2} \leq A_{k-1} \leq a_2 \frac{k(k-1)}{2} + bk$, therefore,

$$\sqrt{(1-c_2)a_1 k(k-1)/2} \leq \alpha_k \leq \frac{1}{2}(1 + \sqrt{1 + 4(a_2 k(k-1)/2 + bk)}).$$

As $a_1 \leq (1-c_2)/3$, $(1-c_2)a_1 k(k-1)/2 \geq a_1^2 k^2$, so one has $\alpha_k \geq a_1 k$. On the other hand, by making use of the inequality

$$\frac{1}{2}(1 + \sqrt{1+x}) \leq \sqrt{x}, \quad \text{for } x \geq 16,$$

one derives

$$\begin{aligned} \alpha_k &\leq \tfrac{1}{2}(1 + \sqrt{1 + 4(a_2 k(k-1)/2 + bk)}) \leq 2\sqrt{a_2 k(k-1)/2 + bk} \\ &\leq 2\sqrt{a_2 k^2} \leq a_2 k \leq a_2 k + b. \end{aligned}$$

That is (2) holds for k, so the lemma is proved. □

Proof of Theorem 3. Denoting by $\theta_k = f(x_k) + \Phi(x_k) - f(x^*) - \Phi(x^*) (\geq 0)$, $k \in \mathbb{N}$, relation (44) in Theorem 2 (with $\mu = 0$) and (44) imply

$$\begin{aligned} &\textstyle\sum_{i=0}^{k} \beta_i [f(x_{i+1}) + \Phi(x_{i+1}) - f(x^*) - \Phi(x^*)] \\ &+(A_k - B_k)[f(x_{k+1}) + \Phi(x_{k+1}) - f(x^*) - \Phi(x^*)] \\ &\leq \min_{x \in \mathbb{R}^n} E_k(x) \leq E_k(x^*) \leq \tfrac{\kappa^{-1}}{2}\|x^* - y_0\|^2 + A_k(f(x^*) + \Phi(x^*)), \end{aligned}$$

which implies immediately $\sum_{i=0}^{\infty} \beta_i \theta_{i+1} < +\infty$. Then in view of Lemma 2, $\sum_{i=0}^{\infty} k\theta_{i+1} < +\infty$, which follows $\sum_{i=0}^{\infty} \alpha_i \theta_{i+1} < +\infty$, and $\sum_{i=0}^{\infty} \alpha_i \theta_i < +\infty$, as well.

Note that $y_k - x_{k+1} = \kappa G_\kappa(y_k)$ and $y_k - x_k = \frac{\alpha_{k-1}-1}{\alpha_k}(x_k - x_{k-1})$, relation (44) gives

$$(f + \Phi)(x_{k+1}) + \kappa^{-1}\langle y_k - x_{k+1}, x_k - y_k \rangle + \frac{\kappa^{-1}}{2}\|y_k - x_{k+1}\|^2 \leq (f + \Phi)(x_k).$$

Equivalently,

$$\theta_{k+1} + \frac{\kappa^{-1}}{2}\|x_{k+1} - x_k\|^2 \leq \theta_k + \frac{\kappa^{-1}}{2}\frac{(\alpha_{k-1}-1)^2}{\alpha_k^2}\|x_k - x_{k-1}\|^2.$$

Therefore,

$$\alpha_k^2 \theta_{k+1} + \frac{\kappa^{-1}}{2}\alpha_k^2\|x_{k+1} - x_k\|^2 \leq \alpha_k^2 \theta_k + \frac{\kappa^{-1}}{2}\alpha_{k-1}^2\|x_k - x_{k-1}\|^2 - \frac{\kappa^{-1}}{2}(2\alpha_{k-1} - 1)\|x_k - x_{k-1}\|^2.$$

By $A_k - B_{k-1} = \alpha_k^2$, then $\alpha_k^2 - \alpha_{k-1}^2 = \alpha_k - \beta_{k-1}$, thus the preceding inequality implies

$$\begin{aligned} \alpha_k^2 \theta_{k+1} + \tfrac{\kappa^{-1}}{2}\alpha_k^2\|x_{k+1} - x_k\|^2 &\leq \alpha_{k-1}^2 \theta_k + \tfrac{\kappa^{-1}}{2}\alpha_{k-1}^2\|x_k - x_{k-1}\|^2 \\ &+(\alpha_k - \beta_{k-1})\theta_k - \tfrac{\kappa^{-1}}{2}(2\alpha_{k-1} - 1)\|x_k - x_{k-1}\|^2. \end{aligned}$$

Since $\sum_{k=0}^{\infty}(\alpha_k - \beta_{k-1})\theta_k \le \sum_{k=0}^{\infty} \alpha_k\theta_k < +\infty$, the inequality above yields

$$\lim_{k\to\infty}[\alpha_k^2\theta_{k+1} + \frac{\kappa^{-1}}{2}\alpha_k^2\|x_{k+1} - x_k\|^2] \text{ exists}$$

as well as

$$\sum_{k=0}^{\infty}\alpha_k\|x_{k+1} - x_k\|^2 < +\infty,$$

and consequently,

$$\sum_{k=0}^{\infty}\alpha_k(\theta_k + \frac{\kappa^{-1}}{2}\|x_{k+1} - x_k\|^2) < +\infty.$$

To complete the proof, we will show that this relation implies

$$\lim_{k\to\infty}[\alpha_k^2\theta_{k+1} + \frac{\kappa^{-1}}{2}\alpha_k^2\|x_{k+1} - x_k\|^2] = 0.$$

Indeed, if this is not the case, then

$$\lim_{k\to\infty}[\alpha_k^2\theta_{k+1} + \frac{\kappa^{-1}}{2}\alpha_k^2\|x_{k+1} - x_k\|^2] = \delta > 0,$$

which follows that

$$\begin{aligned}&\lim_{k\to\infty}\sum_{i=[k/2]}^{k}\alpha_k(\theta_k + \tfrac{\kappa^{-1}}{2}\|x_{k+1} - x_k\|^2) \ge \lim_{k\to\infty}\sum_{i=[k/2]}^{k}\tfrac{\delta}{\alpha_k}\\ &\overset{\text{by Lemma 2}}{\ge} \lim_{k\to\infty}\sum_{i=[k/2]}^{k}\tfrac{\delta}{a_2k+b} \ge \lim_{k\to\infty}\tfrac{\delta(k-[k/2])}{a_2k+b} = \tfrac{\delta}{2a_2} > 0,\end{aligned}$$

which contradicts the summable property of $\sum_{k=0}^{\infty}\alpha_k(\theta_k + \frac{\kappa^{-1}}{2}\|x_{k+1} - x_k\|^2)$. Hence $\lim_{k\to\infty}\alpha_k^2\theta_k = 0$ as well as $\lim_{k\to\infty}\alpha_k\|x_{k+1} - x_k\| = 0$. In view of Lemma 2, one obtains (54).

The proof of the convergence of the sequence $\{x_k\}$ follows the idea in [1,2], that, by virtue of Opial's Lemma [20], it suffices to show that for any minimizer x^* of (1), $\lim_{k\to\infty}\|x_k - x^*\|^2$ exists finitely. Indeed, considering the sequence

$$z_k = x_k + (\alpha_{k-1} - 1)(x_k - x_{k-1}),\ k \in \mathbb{N},$$

one has

$$\begin{aligned}&\|z_k - x^*\|^2 = \|x_k + (\alpha_{k-1} - 1)(x_k - x_{k-1}) - x^*\|^2\\ &= \|x_k - x^*\|^2 + (\alpha_{k-1} - 1)^2\|x_k - x_{k-1}\|^2 + 2(\alpha_{k-1} - 1)\langle x_k - x_{k-1}, x_k - x^*\rangle\\ &= [(\alpha_{k-1} - 1)^2 + \alpha_{k-1} - 1]\|x_k - x_{k-1}\|^2 + b_k,\end{aligned}$$

where,

$$b_k = \alpha_{k-1}\|x_k - x^*\|^2 - (\alpha_{k-1} - 1)\|x_{k-1} - x^*\|^2.$$

By Lemma 2, $\alpha_k = O(k)$, implying $(\alpha_{k-1} - 1)^2 + \alpha_{k-1} - 1 = O(k^2)$. Thus, since $\lim_{k\to\infty} k\|x_k - x_{k-1}\| = 0$, one has

$$\lim_{k\to\infty}[(\alpha_{k-1} - 1)^2 + \alpha_{k-1} - 1]\|x_k - x_{k-1}\|^2 = 0,$$

which follows that the convergence of $\{\|z_k - x^*\|^2\}$ is equivalent to the one of $\{b_k\}$. Thanks to Lemma 1 (for $\mu = 0$),

$$\langle G_\kappa(y_k), y_k - x^*\rangle - \frac{\kappa}{2}\|G_\kappa(y_k)\|^2 \geq (f + \Phi)(x_{k+1}) - (f + \Phi)(x^*) \geq 0.$$

Using this inequality, and $G_\kappa(y_k) = (y_k - x_{k+1})/\kappa$, it is easy to derive that

$$\begin{aligned}\|x_{k+1} - x^*\|^2 \leq \|y_k - x^*\|^2 &= \|x_k + (\alpha_{k-1} - 1)\alpha_k^{-1}(x_k - x_{k-1}) - x^*\|^2 \\ &= (1 + (\alpha_{k-1} - 1)\alpha_k^{-1})\|x_k - x^*\|^2 - (\alpha_{k-1} - 1)\alpha_k^{-1}\|x_{k-1} - x^*\|^2 \\ &+[(\alpha_{k-1} - 1)^2\alpha_k^{-2} + (\alpha_{k-1} - 1)\alpha_k^{-1}]\|x_k - x_{k-1}\|^2.\end{aligned}$$

By virtue of Lemma 2, there is a constant $c > 0$, such that

$$(\alpha_{k-1} - 1)^2\alpha_k^{-2} + (\alpha_{k-1} - 1)\alpha_k^{-1} \leq c, \quad \forall k \in \mathbb{N}_*.$$

Therefore, the preceding inequality yields immediately

$$b_{k+1} - b_k \leq c\alpha_k\|x_k - x_{k-1}\|^2,$$

and by $\sum_{k=1}^{\infty} \alpha_k\|x_k - x_{k-1}\|^2 < +\infty$, it follows the existence finitely of $\lim_{k\to\infty} b_k$, so is $\lim_{k\to\infty} \|z_k - x^*\|^2$. As $\lim_{k\to\infty}(\alpha_{k-1} - 1)\|x_k - x_{k-1}\| = 0$, since $\alpha_k = O(k)$ and $\lim_{k\to\infty} k\|x_k - x_{k-1}\| = 0$, the convergence of $\{\|x_k - x^*\|^2\}$ follows, and the proof is completed. □

References

1. Attouch H., Peypouquet J., The rate of convergence of Nesterov's accelerated forward-backward method is actually faster $O(1/k^2)$, *SIAM J. Optim.,* **26**(3), 1824-1834, (2016).
2. Attouch H., Chbani Z., Peypouquet J., and Redont P, Fast convergence of inertial dynamics and algorithms with asymptotic vanishing damping, *Math. Program.,* **168**(1-2), 123-175, (2018).
3. Attouch, H.; Peypouquet, J. Convergence rate of proximal inertial algorithms associated with Moreau envelopes of convex functions, *In: Bauschke H., Burachik R., Luke D. (eds) Splitting Algorithms, Modern Operator Theory, and Applications,* Springer, (2019).
4. Beck A., Teboulle M., A fast iterative shinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imag. Sci.,* **2**(1), 183-202, (2009).
5. Bubeck S., Convex Optimization: Algorithms and Complexity, *Foundations and Trends in Machine Learning,* **8**(3-4), 231-357,(2015).
6. Bubbeck S., Lee J.T., and Singh M., A geometric alternative to Nesterov's accelerated gradient descent, *Preprint, arXiv:1506.08187,* (2015).
7. Bauschke H., Combettes P., *Convex analysis and monotone operator theory in Hilbert spaces,* CMS Books in Mathematics, Springer, New York, (2011).
8. Chen S., Ma S., and Liu W., Geometric Descent Method for Convex Composite Minimization, *Preprint, arXiv: 1612.09034,* (2016).
9. Drusvyatskiy D., Fazel M., and Roy S., An optimal first order method based on optimal quadratic averaging, *SIAM J. Optim.,* **28**(1), 251-271, (2018).
10. Goldstein A. A., Convex programming in Hilbert spaces, *Bull. Am. Math.Soc.,* **70**, 709-710, (1964).
11. Kim D., Fessler J. A., Optimized first-order methods for smooth convex minimization, *Math. Program., Series A,* **159**, 81-107, (2016).

12. Levittin E.S., Polyak B. T., Constrained minimization problems, *USSR Comput. Math. Math. Phys.,* **6**, 1-50, (1966).
13. Lions P.L., Mercier B., Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.,* **16**, 964-979, (1979).
14. Nesterov Y., A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$, *Doklady AN SSSR* (translated as Soviet Math.Docl.) **269**, 543-547, (1983).
15. Nesterov Y., *Introductory Lectures on Convex Optimization: Basis course,* Kluwer, Boston (2003).
16. Nesterov Y., Smooth minimization of non-smooth functions, *Math. Program. Ser A.,* **103**, 127-152 (2005).
17. Nesterov Y., Gradient methods for minimization composite objective functions, *Math. Prog. Ser. A,* **140**, 125-161, (2013).
18. Nesterov Y., Universal Gradient methods for convex optimization problems, *Math. Prog. Ser. A,* **152**, 381-404, (2015).
19. Necoara I., Nesterov Y., and Glineur F., Linear convergence of first order methods for non-strongly convex optimization, *Math. Program. Series A,* **175**, 69-107, (2019).
20. Opial Z., Weak convergence of the sequence of successive approximations of nonexpansive mappings, *Bull. Am. Math. Soc.,* **73**, 591-597, (1967).
21. Parikh N., Boyd S., Proximal algorithms, *Found. Trends Optim.,* **1**, 123-231, (2013).