Duan et al. *Journal of Inequalities and Applications* (2020) 2020:255
https://doi.org/10.1186/s13660-020-02522-6

Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH **Open Access**

New inertial proximal gradient methods for unconstrained convex optimization problems

Peichao Duan[1*], Yiqun Zhang[2] and Qinxiong Bu[2]

*Correspondence:
pcduancauc@126.com
[1]College of Science, Tianjin Key Lab for Advanced Signal Processing, Civil Aviation University of China, Tianjin 300300, China
Full list of author information is available at the end of the article

Abstract

The proximal gradient method is a highly powerful tool for solving the composite convex optimization problem. In this paper, firstly, we propose inexact inertial acceleration methods based on the viscosity approximation and proximal scaled gradient algorithm to accelerate the convergence of the algorithm. Under reasonable parameters, we prove that our algorithms strongly converge to some solution of the problem, which is the unique solution of a variational inequality problem. Secondly, we propose an inexact alternated inertial proximal point algorithm. Under suitable conditions, the weak convergence theorem is proved. Finally, numerical results illustrate the performances of our algorithms and present a comparison with related algorithms. Our results improve and extend the corresponding results reported by many authors recently.

MSC: 47H09; 47H10; 47J25; 65K10; 90C25

Keywords: Convex optimization; Viscosity approximation; Proximal operator; Inertial acceleration; Alternated inertial acceleration

1 Introduction

Let H be a real Hilbert space with the inner product $\langle\cdot,\cdot\rangle$ and the induced norm $\|\cdot\|$, and let C be a nonempty closed convex subset of H. Let $\Gamma_0(H)$ be a space of functions in H that are proper, convex, and lower semicontinuous. We will deal with the unconstrained convex optimization problem of the following type:

$$\min_{x\in H} f(x)+g(x), \tag{1.1}$$

where $f,g\in\Gamma_0(H)$. It is often the case where f is differentiable and g is subdifferentiable.

In 1978, problem (1.1) was first studied in [13] and provided a natural tool to study various generic optimization models under a common framework. In recent years, many researchers have already proposed some algorithms to solve problem (1.1) and have discussed a lot of weak and strong convergence results, such as [1, 6, 12, 23, 25], just to name a few. As we know, lots of important optimization problems can be cast in this form. See, for instance, [23], where the author introduced the properties and iterative methods for the lasso as a special case of (1.1); due to the involvement of the l_1 norm, which promotes sparsity, we can get a good result on solving the corresponding problem.

The following proposition is very useful for constructing the iterative algorithms.

Proposition 1.1 (see [23]) *Let $f, g \in \Gamma_0(H)$. Let $x^* \in H$ and $\lambda > 0$. Assume that f is finite-valued and differential on H. Then x^* is a solution to (1.1) if and only if x^* solves the fixed point equation*

$$x^* = \bigl(\operatorname{prox}_{\lambda g}(I - \lambda \nabla f)\bigr)x^*. \tag{1.2}$$

On the other hand, we know that the errors often are produced in the process of calculation. It is an important property of algorithms which guarantees the convergence of the iterate under summable errors. Many authors have studied algorithms with perturbations and their convergence. Some related results are found in [3–5]. In 2011, Boikanyo and Morosanu introduced [2] a proximal point algorithm with error sequence. Under the summability condition on errors and some additional conditions on the parameters, they obtained strong convergence theorem.

In 2016, Jin, Censor, and Jiang [11] presented the projected scaled gradient (PSG) method with bounded perturbations in a finite dimensional setting for solving the following minimization problem:

$$\min_{x \in C} f(x), \tag{1.3}$$

where f is a continuously differentiable, convex function. More precisely, the method generates a sequence according to

$$x_{n+1} = P_C\bigl(x_n - \lambda_n D(x_n)\nabla f(x_n) + e(x_n)\bigr), \quad n \geq 0, \tag{1.4}$$

and converges to a solution of problem (1.3) under suitable conditions, where $D(x_n)$ is a diagonal scaling matrix.

In 2017, Xu extended the method to infinite dimensional space and projected the superiorization techniques for the relaxed PSG [24]. The following iterative step was introduced:

$$x_{n+1} = (1 - \tau_n)x_n + \tau_n P_C\bigl(x_n - \gamma_n D(x_n)\nabla f(x_n) + e(x_n)\bigr), \quad n \geq 0, \tag{1.5}$$

where $\tau_n \in [0, 1]$. The weak convergence theorem was obtained in [24].

Quite recently, Guo and Cui [8] considered the modified proximal gradient method:

$$x_{n+1} = \alpha_n h(x_n) + (1 - \alpha_n)\operatorname{prox}_{\lambda_n g}(I - \lambda_n \nabla f)(x_n) + e(x_n), \quad n \geq 0, \tag{1.6}$$

where h is a contractive mapping. The algorithm converges strongly to a solution of problem (1.1).

To accelerate the convergence of iteration methods, Polyak [19] introduced the following algorithm that can speed up gradient descent:

$$\begin{cases} y_n = x_n + \delta_n(x_n - x_{n-1}), \\ x_{n+1} = y_n - \lambda_n \nabla F(x_n). \end{cases} \tag{1.7}$$

This modification was made immensely popular by Nesterov's accelerated gradient algorithm [18]. Generally, an inertial iteration for operator $\mathbf{P}$ writes

$$\begin{cases} y_n = x_n + \delta_n(x_n - x_{n-1}), \\ x_{n+1} = \mathbf{P}(y_n). \end{cases} \tag{1.8}$$

In 2009, Beck and Teboulle [1] proposed a fast iterative shrinkage-thresholding algorithm for linear inverse problems. By applying the inertial technique, $\{x_n\}$ is not employed on the previous point $\{x_{n-1}\}$, but rather at the point $\{y_n\}$ which uses a very specific linear combination of the previous two points $\{x_{n-1}, x_{n-2}\}$. Therefore, the convergence speed of the algorithm is greatly accelerated.

In 2015, for solving the maximal monotone inclusion problem, Mu and Peng [17] introduced alternated inertial proximal point iterates as follows:

$$x_{n+1} = J_{\lambda T}(y_n), \tag{1.9}$$

where y_n is defined as

$$y_n = \begin{cases} x_n + \delta_n(x_n - x_{n-1}), & n = odd, \\ x_n, & n = even. \end{cases} \tag{1.10}$$

In equation (1.9), T is a set-valued maximal monotone operator and $\lambda > 0$. This form is a lot less popular than general inertia. However, it has pretty good convergence properties and performance.

In 2017, Iutzeler and Hendrickx [10] proposed a generic acceleration for optimization algorithm via relaxation and inertia, they also used alternated inertial acceleration in their algorithm. They obtained the convergence of the iterative sequence under some suitable assumptions.

Very recently, Shehu and Gibali [21] studied a new alternated inertial procedure for solving split feasibilities. Under some mild assumptions, they showed that the sequence converges strongly.

In this paper, mainly inspired and motivated by the above works, we introduce several iterative algorithms. Firstly, we combine the contractive mapping and proximal operator to propose an inertial acceleration proximal gradient method with errors for solving problem (1.1). Under more general and flexible conditions, we prove that the sequence converges strongly. Further, we extend the algorithm to a more generalized viscosity inertial acceleration method. Secondly, we propose a kind of alternating inertial proximal point algorithm with errors to solve problem (1.1), then we prove that the sequence converges weakly under appropriate conditions. Finally, we present several numerical examples to illustrate the effectiveness of our iterative schemes.

2 Preliminaries

We start by recalling some lemmas, definitions, and propositions needed in the proof of the main results.

Recall that given a closed subset C of a real Hilbert space H, for any $x \in H$, there exists a unique nearest point in C denoted by $P_C x$ such that

$$\|x - P_C x\| \leq \|x - y\|, \quad \forall y \in C.$$

Such a $P_C x$ is called the metric projection of H onto C.

Lemma 2.1 (see [14]) *Let C be a nonempty closed convex subset of a real Hilbert space H. Given $x \in H$ and $z \in C$, then $y = P_C x$ if and only if we have the relation*

$$\langle x - y, y - z\rangle \geq 0, \quad \forall z \in C.$$

Lemma 2.2 *Let H be a real Hilbert space, the following statements hold*:

(i) $\|x + y\|^2 = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2$, $\forall x, y \in H$.
(ii) $\|x + y\|^2 \leq \|x\|^2 + 2\langle x + y, y\rangle$, $\forall x, y \in H$.
(iii) $\|\alpha x + (1 - \alpha)y\|^2 = \alpha\|x\|^2 + (1 - \alpha)\|y\|^2 - \alpha(1 - \alpha)\|x - y\|^2$ *for all* $\alpha \in \mathbb{R}$ *and* $x, y \in H$.

Definition 2.3 A mapping $F : H \to H$ is said to be

(i) Lipschitzian if there exists a positive constant L such that

$$\|Fx - Fy\| \leq L\|x - y\|, \quad \forall x, y \in H.$$

In particular, if $L = 1$, F is called nonexpansive. If $L \in [0, 1)$, F is called contractive.

(ii) α-*averaged* mapping(α-*av* for short) if

$$F = (1 - \alpha)I + \alpha T,$$

where $\alpha \in (0, 1)$ and $T : H \to H$ is nonexpansive.

Proposition 2.4 ([22])

(i) *If $T_1, T_2, \ldots, T_n$ are averaged mappings, then we can get that $T_n T_{n-1} \cdots T_1$ is averaged. In particular, if T_i is α_i-av for each $i = 1, 2$, where $\alpha_i \in (0, 1)$, then $T_2 T_1$ is $(\alpha_2 + \alpha_1 - \alpha_2\alpha_1)$-av.*
(ii) *If the mappings $\{T_i\}_{i=1}^N$ are averaged and have a common fixed point, then we have*

$$\bigcap_{i=1}^{N} \mathrm{Fix}(T_i) = \mathrm{Fix}(T_1 \cdots T_N).$$

Here, the notation $\mathrm{Fix}(T)$ *denotes the set of fixed points of the mapping T; that is,* $\mathrm{Fix}(T) := \{x \in H : Tx = x\}$.

(iii) *If T is ν-ism, then, for any $\tau > 0$, τT is $\frac{\nu}{\tau}$-ism.*
(iv) *T is averaged if and only if $I - T$ is ν-ism for some $\nu > \frac{1}{2}$. Indeed, for any* $0 < \alpha < 1$, *T is α-averaged if and only if $I - T$ is $\frac{1}{2\alpha}$-ism.*

Definition 2.5 (see [16]) The proximal operator of $\varphi \in \Gamma_0(H)$ is defined by

$$\mathrm{prox}_{\varphi}(x) = \arg\min_{\nu \in H}\left\{\varphi(\nu) + \frac{1}{2}\|\nu - x\|^2\right\}, \quad x \in H.$$

The proximal operator of φ of order $\lambda > 0$ is defined as the proximal operator of $\lambda\varphi$, that is,

$$\operatorname{prox}_{\lambda\varphi}(x) = \arg\min_{\nu\in H}\left\{\varphi(\nu) + \frac{1}{2\lambda}\|\nu - x\|^2\right\}, \quad x \in H.$$

Lemma 2.6 *The proximal identity*

$$\operatorname{prox}_{\lambda\varphi} x = \operatorname{prox}_{\mu\varphi}\left(\frac{\mu}{\lambda}x + \left(1 - \frac{\mu}{\lambda}\right)\operatorname{prox}_{\lambda\varphi} x\right) \tag{2.1}$$

holds for $\varphi \in \Gamma_0(H)$, $\lambda > 0$ *and* $\mu > 0$.

Lemma 2.7 (Demiclosedness principle, see [7]) *Let* H *be a real Hilbert space, and let* $T : H \to H$ *be a nonexpansive mapping with* $\operatorname{Fix}(T) \neq \emptyset$. *If* $\{x_n\}$ *is a sequence in* H *weakly converging to* x *and if* $\{(I - T)x_n\}$ *converges strongly to* y, *then* $(I - T)x = y$; *in particular, if* $y = 0$, *then* $x \in \operatorname{Fix}(T)$.

Lemma 2.8 (see [9]) *Assume that* $\{s_n\}$ *is a sequence of nonnegative real numbers such that*

$$s_{n+1} \le (1 - \gamma_n)s_n + \gamma_n\mu_n, \quad n \ge 0,$$
$$s_{n+1} \le s_n - \eta_n + \varphi_n, \quad n \ge 0,$$

where $\{\gamma_n\}$ *is a sequence in* $(0, 1)$, $\{\eta_n\}$ *is a sequence of nonnegative real numbers and* $\{\mu_n\}$ *and* $\{\varphi_n\}$ *are two sequences in* $\mathbb{R}$ *such that*

(i) $\sum_{n=0}^{\infty} \gamma_n = \infty$,

(ii) $\lim_{n\to\infty} \varphi_n = 0$,

(iii) $\lim_{k\to\infty} \eta_{n_k} = 0$ *implies* $\limsup_{k\to\infty} \mu_{n_k} \le 0$ *for any subsequence* $\{n_k\} \subset \{n\}$.

Then $\lim_{n\to\infty} s_n = 0$.

Lemma 2.9 (see [7]) *Let* C *be a nonempty closed convex subset of a real Hilbert space* H. *Let* $\{x_n\}$ *be a sequence in* H *satisfying the properties*:

(i) $\lim_{n\to\infty} \|x_n - z\|$ *exists for each* $z \in C$,

(ii) $\omega_w(x_n) \subset C$, *where* $\omega_w(x_n) := \{x : \exists x_{n_j} \rightharpoonup x\}$ ($\{x_{n_j}\}$ *is a subsequence of* $\{x_n\}$) *denotes the weak* ω*-limit set of* $\{x_n\}$.

Then $\{x_n\}$ *converges weakly to a point in* C.

Lemma 2.10 (see [20]) *Let* $\{s_n\}$ *be a sequence of nonnegative numbers satisfying the generalized nonincreasing property*

$$s_{n+1} \le s_n + \sigma_n, \quad n \ge 0,$$

where $\{\sigma_n\}$ *is a sequence of nonnegative numbers such that* $\sum_{n=0}^{\infty} \sigma_n < \infty$. *Then* $\{s_n\}$ *is bounded and* $\lim_{n\to\infty} s_n$ *exists.*

3 Main results

3.1 Inertial proximal gradient algorithm

In this section, we combine a viscosity iterative method for approximating the unique fixed point of the following variational inequality problem (VIP for short):

$$\langle (I-h)x^*, \tilde{x}-x^* \rangle \geq 0, \quad \forall \tilde{x} \in \mathrm{Fix}(V_\lambda), \tag{3.1}$$

where $h : H \to H$ is ρ-*contractive* and V_λ is nonexpansive.

We propose an inertial acceleration algorithm.

Algorithm 1

1. Choose $x_0, x_1 \in H$ and set $n := 1$.
2. Given x_n, x_{n-1}, compute

$$y_n = x_n + \delta_n(x_n - x_{n-1}). \tag{3.2}$$

3. Calculate the next iterate via

$$x_{n+1} = \alpha_n h(y_n) + (1-\alpha_n)\big(\mathrm{prox}_{\lambda_n g}\big(y_n - \lambda_n D(y_n)\nabla f(y_n) + e(y_n)\big)\big). \tag{3.3}$$

4. If $\|x_n - x_{n+1}\| < \epsilon$, then stop. Otherwise, set $n = n+1$ and go to *2*.

Rewrite iteration (3.3) as follows:

$$\begin{aligned} x_{n+1} &= \alpha_n h(y_n) + (1-\alpha_n)\,\mathrm{prox}_{\lambda_n g}\big(y_n - \lambda_n \nabla f(y_n) + \hat{e}_n\big) \\ &= \alpha_n h(y_n) + (1-\alpha_n)\big(\mathrm{prox}_{\lambda_n g}\big(y_n - \lambda_n \nabla f(y_n)\big) + \tilde{e}_n\big), \end{aligned} \tag{3.4}$$

where $\hat{e}_n = \lambda_n \theta(y_n) + e(y_n)$, $\theta(y_n) = \nabla f(y_n) - D(y_n)\nabla f(y_n)$, and

$$\tilde{e}_n = \mathrm{prox}_{\lambda_n g}\big(y_n - \lambda_n \nabla f(y_n) + \hat{e}_n\big) - \mathrm{prox}_{\lambda_n g}\big(y_n - \lambda_n \nabla f(y_n)\big).$$

Note that $\|\tilde{e}_n\| \leq \|\hat{e}_n\| \leq \|e(y_n)\| + \lambda_n\|\theta(y_n)\|$, it is easy to get $\sum_{n=0}^{\infty} \|\tilde{e}_n\| < \infty$ from conditions (iii)–(iv) of Theorem 3.1. We use S to denote the solution set of problem (1.1).

Theorem 3.1 *Let $f, g \in \Gamma_0(H)$ and assume that (1.1) is consistent (i.e., $S \neq \emptyset$). Let h be ρ-contractive self-map of H with $0 \leq \rho < 1$ and ∇f is L-Lipschitzian. Assume that D is a diagonal scaling matrix. Given $x_0, x_1 \in H$, let $\{x_n\}$ be a sequence generated by Algorithm 1, where $\lambda_n \in (0, \frac{2}{L})$, $\alpha_n \in (0, \frac{2+\lambda_n L}{4})$. Suppose that*

(i) $\lim_{n\to\infty} \alpha_n = 0$, $\sum_{n=0}^{\infty} \alpha_n = \infty$;
(ii) $0 < \liminf_{n\to\infty} \lambda_n \leq \limsup_{n\to\infty} \lambda_n < \frac{2}{L}$;
(iii) $\sum_{n=0}^{\infty} \|e(y_n)\| < \infty$;
(iv) $\sum_{n=0}^{\infty} \|\theta(y_n)\| < \infty$;
(v) $\sum_{n=0}^{\infty} \delta_n \|x_n - x_{n-1}\| < \infty$.

Then $\{x_n\}$ converges strongly to x^*, where x^* is a solution of (1.1), which is also the unique solution of variational inequality problem (3.1).

Proof We divide the proof into several steps.

Step 1. Show that $\{x_n\}$ is bounded. For any $z \in S$,

$$\begin{aligned}\|y_n - z\| &= \big\|x_n + \delta_n(x_n - x_{n-1}) - z\big\| \\ &\le \|x_n - z\| + \delta_n\|x_n - x_{n-1}\|.\end{aligned} \tag{3.5}$$

Put $V_{\lambda_n} := \mathrm{prox}_{\lambda_n g}(I - \lambda_n \nabla f)$, from (3.4) and (3.5), we have

$$\begin{aligned}&\|x_{n+1} - z\| \\ &\quad = \big\|\alpha_n h(y_n) + (1-\alpha_n)(V_{\lambda_n} y_n + \tilde{e}_n) - z\big\| \\ &\quad = \big\|\alpha_n\big(h(y_n) - z\big) + (1-\alpha_n)(V_{\lambda_n} y_n - z) + (1-\alpha_n)\tilde{e}_n\big\| \\ &\quad \le \alpha_n\big\|h(y_n) - h(z)\big\| + \alpha_n\big\|h(z) - z\big\| + (1-\alpha_n)\|V_{\lambda_n} y_n - z\| + \|\tilde{e}_n\| \\ &\quad \le \alpha_n\rho\|y_n - z\| + \alpha_n\big\|h(z) - z\big\| + (1-\alpha_n)\|y_n - z\| + \|\tilde{e}_n\| \\ &\quad = \big(1 - \alpha_n(1-\rho)\big)\|y_n - z\| + \alpha_n\big\|h(z) - z\big\| + \|\tilde{e}_n\| \\ &\quad \le \big(1 - \alpha_n(1-\rho)\big)\|x_n - z\| + \delta_n\|x_n - x_{n-1}\| + \alpha_n\big\|h(z) - z\big\| + \|\tilde{e}_n\| \\ &\quad = \big(1 - \alpha_n(1-\rho)\big)\|x_n - z\| + \alpha_n(1-\rho)\frac{\|h(z) - z\| + (\delta_n\|x_n - x_{n-1}\| + \|\tilde{e}_n\|)/\alpha_n}{1-\rho}.\end{aligned} \tag{3.6}$$

From conditions (iii)–(v) and $\alpha_n > 0$, we get $\{(\delta_n\|x_n - x_{n-1}\| + \|\tilde{e}_n\|)/\alpha_n\}$ is bounded. Thus there exists some $M_1 > 0$ such that

$$M_1 \ge \sup\big\{\big\|h(z) - z\big\| + \big(\delta_n\|x_n - x_{n-1}\| + \|\tilde{e}_n\|\big)/\alpha_n\big\}$$

for all $n \ge 0$. Then the mathematical induction implies that

$$\|x_n - z\| \le \max\left\{\|x_0 - z\|, \frac{M_1}{1-\rho}\right\}.$$

Therefore, the sequence $\{x_n\}$ is bounded and so are $\{y_n\}$, $\{h(y_n)\}$, and $\{V_{\lambda_n} y_n\}$.

Step 2. Show that $\lim_{k\to\infty} \eta_{n_k} = 0$ implies

$$\lim_{k\to\infty} \|x_{n_k} - V_{\lambda_{n_k}} x_{n_k}\| = 0$$

for any sequence $\{n_k\} \subset \{n\}$. Firstly, fix $z \in S$, we have

$$\begin{aligned}\|y_n - z\|^2 &= \big\|x_n + \delta_n(x_n - x_{n-1}) - z\big\|^2 \\ &\le \|x_n - z\|^2 + 2\big\langle x_n - z + \delta_n(x_n - x_{n-1}), \delta_n(x_n - x_{n-1})\big\rangle \\ &\le \|x_n - z\|^2 + 2\delta_n\|x_n - x_{n-1}\|\big(\|x_n - z\| + \delta_n\|x_n - x_{n-1}\|\big).\end{aligned} \tag{3.7}$$

Then from (3.4) we get

$$\begin{aligned}&\|x_{n+1} - z\|^2 \\ &\quad = \big\|\alpha_n h(y_n) + (1-\alpha_n)(V_{\lambda_n} y_n + \tilde{e}_n) - z\big\|^2\end{aligned}$$

$$
\begin{aligned}
&\le \left\|\alpha_n h(y_n) + (1-\alpha_n)V_{\lambda_n}y_n - z\right\|^2 + 2(1-\alpha_n)\left\langle \alpha_n h(y_n) + (1-\alpha_n)V_{\lambda_n}y_n - z, \tilde{e}_n\right\rangle \\
&\quad + \|\tilde{e}_n\|^2 \\
&\le \alpha_n^2\left\|h(y_n) - z\right\|^2 + (1-\alpha_n)^2\|V_{\lambda_n}y_n - z\|^2 + 2\alpha_n(1-\alpha_n)\left\langle h(y_n) - z, V_{\lambda_n}y_n - z\right\rangle \\
&\quad + \left(2\alpha_n\left\|h(y_n) - z\right\| + 2(1-\alpha_n)\|y_n - z\| + \|\tilde{e}_n\|\right)\|\tilde{e}_n\| \\
&\le 2\alpha_n^2\left(\left\|h(y_n) - h(z)\right\|^2 + \left\|h(z) - z\right\|^2\right) + (1-\alpha_n)^2\|y_n - z\|^2 \\
&\quad + 2\alpha_n(1-\alpha_n)\left\langle h(y_n) - z, V_{\lambda_n}y_n - z\right\rangle + M_2\|\tilde{e}_n\| \\
&\le 2\alpha_n^2\left(\left\|h(y_n) - h(z)\right\|^2 + \left\|h(z) - z\right\|^2\right) + (1-\alpha_n)^2\|y_n - z\|^2 \\
&\quad + 2\alpha_n(1-\alpha_n)\left(\left\|h(y_n) - h(z)\right\|\|y_n - z\| + \left\langle h(z) - z, V_{\lambda_n}y_n - z\right\rangle\right) + M_2\|\tilde{e}_n\| \\
&\le \left(1 - \alpha_n\left(2 - \alpha_n\left(1 + 2\rho^2\right) - 2(1-\alpha_n)\rho\right)\right)\|y_n - z\|^2 \\
&\quad + 2\alpha_n(1-\alpha_n)\left\langle h(z) - z, V_{\lambda_n}y_n - z\right\rangle + 2\alpha_n^2\left\|h(z) - z\right\|^2 + M_2\|\tilde{e}_n\|,
\end{aligned}
\tag{3.8}
$$

where M_2 is some constant such that

$$
M_2 \ge \sup\left\{2\alpha_n\left\|h(y_n) - z\right\| + 2(1-\alpha_n)\|y_n - z\| + \|\tilde{e}_n\|\right\}.
$$

Put $\gamma_n := \alpha_n(2 - \alpha_n(1 + 2\rho^2) - 2(1-\alpha_n)\rho)$, using (3.4) and (3.7), we deduce that

$$
\begin{aligned}
&\|x_{n+1} - z\|^2 \\
&\quad \le (1-\gamma_n)\|x_n - z\|^2 + 2\delta_n(1-\gamma_n)\|x_n - x_{n-1}\|\left(\|x_n - z\| + \delta_n\|x_n - x_{n-1}\|\right) \\
&\qquad + 2\alpha_n(1-\alpha_n)\left\langle h(z) - z, V_{\lambda_n}y_n - z\right\rangle + 2\alpha_n^2\left\|h(z) - z\right\|^2 + M_2\|\tilde{e}_n\|.
\end{aligned}
\tag{3.9}
$$

Secondly, since V_{λ_n} *is* $\frac{2+\lambda_n L}{4}$-*av* by Proposition 2.4, we can rewrite

$$
V_{\lambda_n} = \operatorname{prox}_{\lambda_n g}(I - \lambda_n \nabla f) = (1-w_n)I + w_n T_n,
\tag{3.10}
$$

where $w_n = \frac{2+\lambda_n L}{4}$, T_n is nonexpansive and, by condition (ii), we get $\frac{1}{2} < \liminf_{n\to\infty} w_n \le \limsup_{n\to\infty} w_n < 1$. Combining (3.4), (3.8), and (3.10), we obtain

$$
\begin{aligned}
&\|x_{n+1} - z\|^2 \\
&\quad = \left\|\alpha_n h(y_n) + (1-\alpha_n)(V_{\lambda_n}y_n + \tilde{e}_n) - z\right\|^2 \\
&\quad \le \left\|\alpha_n h(y_n) + (1-\alpha_n)V_{\lambda_n}y_n - z\right\|^2 + M_2\|\tilde{e}_n\| \\
&\quad = \left\|V_{\lambda_n}y_n - z + \alpha_n\left(h(y_n) - V_{\lambda_n}y_n\right)\right\|^2 + M_2\|\tilde{e}_n\| \\
&\quad = \|V_{\lambda_n}y_n - z\|^2 + {\alpha_n}^2\left\|h(y_n) - V_{\lambda_n}y_n\right\|^2 + 2\alpha_n\left\langle V_{\lambda_n}y_n - z, h(y_n) - V_{\lambda_n}y_n\right\rangle + M_2\|\tilde{e}_n\| \\
&\quad = \left\|(1-w_n)y_n + w_n T_n y_n - z\right\|^2 + {\alpha_n}^2\left\|h(y_n) - V_{\lambda_n}y_n\right\|^2 \\
&\qquad + 2\alpha_n\left\langle V_{\lambda_n}y_n - z, h(y_n) - V_{\lambda_n}y_n\right\rangle + M_2\|\tilde{e}_n\| \\
&\quad = (1-w_n)\|y_n - z\|^2 + w_n\|T_n y_n - T_n z\|^2 - w_n(1-w_n)\|T_n y_n - y_n\|^2 \\
&\qquad + {\alpha_n}^2\left\|h(y_n) - V_{\lambda_n}y_n\right\|^2 + 2\alpha_n\left\langle V_{\lambda_n}y_n - z, h(y_n) - V_{\lambda_n}y_n\right\rangle + M_2\|\tilde{e}_n\|
\end{aligned}
$$

$$
\begin{aligned}
&\le \|y_n - z\|^2 - w_n(1-w_n)\|T_n y_n - y_n\|^2 + \alpha_n^2 \|h(y_n) - V_{\lambda_n} y_n\|^2 \\
&\quad + 2\alpha_n \langle V_{\lambda_n} y_n - z, h(y_n) - V_{\lambda_n} y_n\rangle + M_2\|\tilde{e}_n\| \\
&\le \|x_n - z\|^2 - w_n(1-w_n)\|T_n y_n - y_n\|^2 + \alpha_n^2 \|h(y_n) - V_{\lambda_n} y_n\|^2 \\
&\quad + 2\alpha_n \langle V_{\lambda_n} y_n - z, h(y_n) - V_{\lambda_n} y_n\rangle \\
&\quad + 2\delta_n \|x_n - x_{n-1}\| \big(\|x_n - z\| + \delta_n \|x_n - x_{n-1}\|\big) + M_2\|\tilde{e}_n\|. \qquad (3.11)
\end{aligned}
$$

Set

$$
\begin{aligned}
&s_n = \|x_n - z\|^2, \qquad \eta_n = w_n(1-w_n)\|T_n y_n - y_n\|^2, \\
&\mu_n = \frac{1}{2 - \alpha_n(1+2\rho^2) - 2(1-\alpha_n)\rho} \Bigg(2\alpha_n \|h(z) - z\|^2 + M_2 \frac{\|\tilde{e}_n\|}{\alpha_n} \\
&\qquad + \frac{2\delta_n \|x_n - x_{n-1}\|(\|x_n - z\| + \delta_n\|x_n - x_{n-1}\|)}{\alpha_n} \\
&\qquad + 2(1-\alpha_n)\langle h(z) - z, V_{\lambda_n} y_n - z\rangle\Bigg), \\
&\varphi_n = \alpha_n^2 \|h(y_n) - V_{\lambda_n} y_n\|^2 + 2\alpha_n \langle V_{\lambda_n} y_n - z, h(y_n) - V_{\lambda_n} y_n\rangle + M_2\|\tilde{e}_n\|.
\end{aligned}
$$

Since $\sum_{n=0}^{\infty} \gamma_n = \infty$ and $\varphi_n \to 0$ hold obviously, in order to complete the proof by using Lemma 2.8, it suffices to verify that $\eta_{n_k} \to 0$ $(k \to \infty)$ implies

$$
\limsup_{k\to\infty} \mu_{n_k} \le 0
$$

for any subsequence $\{n_k\} \subset \{n\}$.

Indeed, as $k \to \infty$, $\eta_{n_k} \to 0$ implies $\|T_{n_k} y_{n_k} - y_{n_k}\| \to 0$, from (3.10), we have

$$
\|y_{n_k} - V_{\lambda_{n_k}} y_{n_k}\| = w_{n_k}\|y_{n_k} - T_{n_k} y_{n_k}\| \to 0. \tag{3.12}
$$

Due to condition (v), it follows that

$$
\|y_{n_k} - x_{n_k}\| = \delta_{n_k}\|x_{n_k} - x_{n_k-1}\| \to 0. \tag{3.13}
$$

Thus, we have

$$
\begin{aligned}
&\lim_{k\to\infty} \|x_{n_k} - V_{\lambda_{n_k}} x_{n_k}\| \\
&\quad = \lim_{k\to\infty} \|x_{n_k} - y_{n_k} + y_{n_k} - V_{\lambda_{n_k}} y_{n_k} + V_{\lambda_{n_k}} y_{n_k} - V_{\lambda_{n_k}} x_{n_k}\| \\
&\quad \le \lim_{k\to\infty} \|x_{n_k} - y_{n_k}\| + \lim_{k\to\infty} \|y_{n_k} - V_{\lambda_{n_k}} y_{n_k}\| + \lim_{k\to\infty} \|V_{\lambda_{n_k}} y_{n_k} - V_{\lambda_{n_k}} x_{n_k}\| \\
&\quad \le \lim_{k\to\infty} 2\|x_{n_k} - y_{n_k}\| + \lim_{k\to\infty} \|y_{n_k} - V_{\lambda_{n_k}} y_{n_k}\|. \qquad (3.14)
\end{aligned}
$$

It follows from (3.12) and (3.13) that

$$
\lim_{k\to\infty} \|x_{n_k} - V_{\lambda_{n_k}} x_{n_k}\| = 0. \tag{3.15}
$$

Step 3. Show that

$$\omega_w(x_{n_k}) \subset S. \tag{3.16}$$

Take $\tilde{x} \in \omega_w(x_{n_k})$ and assume that $\{x_{n_{k_j}}\}$ is a subsequence of $\{x_{n_k}\}$ weakly converging to $\tilde{x}$. Without loss of generality, we still use $\{x_{n_k}\}$ to denote $\{x_{n_{k_j}}\}$. Assume $\lambda_{n_k} \to \lambda$, then $0 < \lambda < \frac{2}{L}$. Set $V_\lambda = \operatorname{prox}_{\lambda g}(I - \lambda \nabla f)$, then V_λ is nonexpansive. Set

$$t_k = x_{n_k} - \lambda_{n_k} \nabla f(x_{n_k}), \qquad z_k = x_{n_k} - \lambda \nabla f(x_{n_k}).$$

Using the proximal identity of Lemma 2.6, we deduce that

$$\begin{aligned}
&\|V_{\lambda_{n_k}} x_{n_k} - V_\lambda x_{n_k}\| \\
&\quad = \|\operatorname{prox}_{\lambda_{n_k} g} t_k - \operatorname{prox}_{\lambda g} z_k\| \\
&\quad = \left\|\operatorname{prox}_{\lambda g}\left(\frac{\lambda}{\lambda_{n_k}} t_k + \left(1 - \frac{\lambda}{\lambda_{n_k}}\right) \operatorname{prox}_{\lambda_{n_k} g} t_k\right) - \operatorname{prox}_{\lambda g} z_k\right\| \\
&\quad \le \left\|\frac{\lambda}{\lambda_{n_k}} t_k + \left(1 - \frac{\lambda}{\lambda_{n_k}}\right) \operatorname{prox}_{\lambda_{n_k} g} t_k - z_k\right\| \\
&\quad \le \frac{\lambda}{\lambda_{n_k}} \|t_k - z_k\| + \left(1 - \frac{\lambda}{\lambda_{n_k}}\right) \|\operatorname{prox}_{\lambda_{n_k} g} t_k - z_k\| \\
&\quad = \frac{\lambda}{\lambda_{n_k}} |\lambda_{n_k} - \lambda| \left\|\nabla f(x_{n_k})\right\| + \left(1 - \frac{\lambda}{\lambda_{n_k}}\right) \|\operatorname{prox}_{\lambda_{n_k} g} t_k - z_k\|.
\end{aligned} \tag{3.17}$$

Since $\{x_n\}$ is bounded, ∇f is Lipschitz continuous, and $\lambda_{n_k} \to \lambda$, we immediately derive from the last relation that $\|V_{\lambda_{n_k}} x_{n_k} - V_\lambda x_{n_k}\| \to 0$. As a result, we find

$$\|x_{n_k} - V_\lambda x_{n_k}\| \le \|x_{n_k} - V_{\lambda_{n_k}} x_{n_k}\| + \|V_{\lambda_{n_k}} x_{n_k} - V_\lambda x_{n_k}\| \to 0. \tag{3.18}$$

Using Lemma 2.7, we get $\omega_w(x_{n_k}) \subset S$. Meanwhile, we have

$$\begin{aligned}
&\limsup_{k\to\infty} \langle h(x^*) - x^*, V_{\lambda_{n_k}} y_{n_k} - x^* \rangle \\
&\quad = \limsup_{k\to\infty} \langle h(x^*) - x^*, V_{\lambda_{n_k}} x_{n_k} - x^* \rangle \\
&\quad = \limsup_{k\to\infty} \langle h(x^*) - x^*, x_{n_k} - x^* \rangle \\
&\quad = \langle h(x^*) - x^*, \tilde{x} - x^* \rangle, \quad \forall \tilde{x} \in S.
\end{aligned} \tag{3.19}$$

Also, since x^* is the unique solution of variational inequality problem (3.1), we get

$$\limsup_{k\to\infty} \langle h(x^*) - x^*, x_{n_k} - x^* \rangle \le 0,$$

and hence $\limsup_{k\to\infty} \mu_{n_k} \le 0$. □

Furthermore, we extend Algorithm 1 to a more generalized viscosity iterative algorithm. Suppose that the contractive mappings sequence $\{h_n(x)\}$ is uniformly convergent on any B, where B is any bounded subset of H. Assume that the solution set $S \neq \emptyset$, next we prove that the sequence $\{x_n\}$ generated by Algorithm 2 converges strongly to a point $x^* \in S$, which also solves variational inequality (3.1).

A more general inertial iterative algorithm is as follows.

Algorithm 2

1. Choose $x_0, x_1 \in H$ and set $n := 1$.
2. Given x_n, x_{n-1}, compute

$$y_n = x_n + \delta_n(x_n - x_{n-1}). \tag{3.20}$$

3. Calculate the next iterate via

$$x_{n+1} = \alpha_n h_n(y_n) + (1 - \alpha_n)(\mathrm{prox}_{\lambda_n g}\big(y_n - \lambda_n D(y_n)\nabla f(y_n) + e(y_n)\big). \tag{3.21}$$

4. If $\|x_n - x_{n+1}\| < \epsilon$, then stop. Otherwise, set $n = n + 1$ and go to *2*.

Theorem 3.2 *Let $f, g \in \Gamma_0(H)$ and assume that* (1.1) *is consistent. Let $\{h_n\}$ be a sequence of ρ_n-contractive self-mappings of H with $0 < \rho_l = \liminf_{n\to\infty} \rho_n \leq \limsup_{n\to\infty} \rho_n = \rho_u < 1$ and $\{h_n(x)\}$ is uniformly convergent on any B, where B is any bounded subset of H. Assume that ∇f is L-Lipschizian and D is a diagonal scaling matrix. Given $x_0, x_1 \in H$, define the sequence $\{x_n\}$ by Algorithm* 2, *where $\lambda_n \in (0, \frac{2}{L})$, $\alpha_n \in (0, \frac{2+\lambda_n L}{4})$. Suppose that*

(i) $\lim_{n\to\infty} \alpha_n = 0$, $\sum_{n=0}^{\infty} \alpha_n = \infty$;
(ii) $0 < \liminf_{n\to\infty} \lambda_n \leq \limsup_{n\to\infty} \lambda_n < \frac{2}{L}$;
(iii) $\sum_{n=0}^{\infty} \|e(y_n)\| < \infty$;
(iv) $\sum_{n=0}^{\infty} \|\theta(y_n)\| < \infty$;
(v) $\sum_{n=0}^{\infty} \delta_n \|x_n - x_{n-1}\| < \infty$.

Then $\{x_n\}$ converges strongly to x^*, where x^* is a solution of (1.1), *which is also the unique solution of variational inequality problem* (3.1).

Proof Using the uniform convergence of the sequence of contractive mapping $\{h_n\}$ and consulting [6], we have $\lim_{n\to\infty} h_n = h$. It is not hard to complete the proof by using some similar techniques as in Theorem 3.1. □

3.2 Alternated inertial proximal gradient algorithm

In the light of the ideas of [10, 17, 21] and more related references, combining the proximal gradient method, we consider the following algorithm.

Algorithm 3

1. Choose $x_0, x_1 \in H$ and set $n := 1$.
2. Given x_n, x_{n-1}, compute

$$y_n = \begin{cases} x_n + \delta_n(x_n - x_{n-1}), & n = odd, \\ x_n, & n = even. \end{cases} \tag{3.22}$$

3. Calculate the next iterate via

$$x_{n+1} = \mathrm{prox}_{\lambda_n g}\big(y_n - \lambda_n D(y_n)\nabla f(y_n) + e(y_n)\big). \tag{3.23}$$

4. If $\|x_n - x_{n+1}\| < \epsilon$, then stop. Otherwise, set $n = n + 1$ and go to *2*.

Similar to (3.3), we rewrite (3.23) as follows:

$$x_{n+1} = \mathrm{prox}_{\lambda_n g}\big(y_n - \lambda_n \nabla f(y_n)\big) + \tilde{e}_n, \tag{3.24}$$

where $\hat{e}_n = \lambda_n \theta(y_n) + e(y_n)$, $\theta(y_n) = \nabla f(y_n) - D(y_n)\nabla f(y_n)$, and

$$\tilde{e}_n = \mathrm{prox}_{\lambda_n g}\big(y_n - \lambda_n \nabla f(y_n) + \hat{e}_n\big) - \mathrm{prox}_{\lambda_n g}\big(y_n - \lambda_n \nabla f(y_n)\big).$$

Theorem 3.3 *Let $f, g \in \Gamma_0(H)$ and assume that (1.1) is consistent (i.e., $S \neq \emptyset$). Assume that ∇f is L-Lipschitzian and D is a diagonal scaling matrix. Given $x_0, x_1 \in H$, let $\{x_n\}$ be a sequence generated by Algorithm 3, where $\lambda_n \in (0, \frac{2}{L})$. Suppose that*

(i) $0 < \liminf_{n\to\infty} \lambda_n \le \limsup_{n\to\infty} \lambda_n < \frac{2}{L}$;
(ii) $\sum_{n=0}^{\infty} \|e(y_n)\| < \infty$;
(iii) $\sum_{n=0}^{\infty} \|\theta(y_n)\| < \infty$;
(iv) $\sum_{n=0}^{\infty} \delta_n \|x_n - x_{n-1}\| < \infty$.

Then $\{x_n\}$ converges weakly to a solution of the minimization problem of (1.1).

Proof *Step 1.* Show that $\{x_n\}$ is bounded. For any $z \in S$,

$$\begin{aligned} \|x_{2n+2} - z\| &= \|V_{\lambda_{2n+1}} y_{2n+1} + \tilde{e}_{2n+1} - z\| \\ &\le \|y_{2n+1} - z\| + \|\tilde{e}_{2n+1}\| \\ &= \big\|x_{2n+1} + \delta_{2n+1}(x_{2n+1} - x_{2n}) - z\big\| + \|\tilde{e}_{2n+1}\| \\ &\le \|x_{2n+1} - z\| + \delta_{2n+1}\|x_{2n+1} - x_{2n}\| + \|\tilde{e}_{2n+1}\|. \end{aligned} \tag{3.25}$$

Applying conditions (ii) and (iv), we deduce that $\{x_{2n}\}$ is bounded. Since

$$\begin{aligned} \|x_{2n+1} - z\| &= \|V_{\lambda_{2n}} y_{2n} + \tilde{e}_{2n} - z\| \\ &= \|V_{\lambda_{2n}} x_{2n} + \tilde{e}_{2n} - z\| \\ &\le \|x_{2n} - z\| + \|\tilde{e}_{2n}\|. \end{aligned} \tag{3.26}$$

It is easy to get that $\{x_n\}$ is bounded and so are $\{y_n\}$ and $\{V_{\lambda_n} y_n\}$. Also, it follows from (3.25) and (3.26) that $\{x_n\}$ is quasi-Fejer monotone with respect to S. By Lemma 2.10, $\lim_{n\to\infty} \|x_n - z\|$ exists.

Step 2. Show that $\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0$ and $\lim_{n\to\infty} \|x_n - V_{\lambda_n} x_n\| = 0$. Firstly, fix $z \in S$, by Lemma 2.2 and Schwartz's inequality, we have

$$\begin{aligned} \|y_{2n+1} - z\|^2 &= \big\|x_{2n+1} + \delta_{2n+1}(x_{2n+1} - x_{2n}) - z\big\|^2 \\ &\le \|x_{2n+1} - z\|^2 + 2\big\langle x_{2n+1} - z + \delta_{2n+1}(x_{2n+1} - x_{2n}), \delta_{2n+1}(x_{2n+1} - x_{2n})\big\rangle \\ &\le \|x_{2n+1} - z\|^2 \\ &\quad + 2\delta_{2n+1}\|x_{2n+1} - x_{2n}\|\big(\|x_{2n+1} - z\| + \delta_{2n+1}\|x_{2n+1} - x_{2n}\|\big). \end{aligned} \tag{3.27}$$

Since V_{λ_n} *is* $\frac{2+\lambda_n L}{4}$*-av*, we see that

$$V_{\lambda_n} = \mathrm{prox}_{\lambda_n g}(I - \lambda_n \nabla f) = (1 - w_n)I + w_n T_n, \tag{3.28}$$

where $w_n = \frac{2+\lambda_n L}{4}$, T_n is nonexpansive. From condition (ii), we get $\frac{1}{2} < \liminf_{n\to\infty} w_n \le \limsup_{n\to\infty} w_n < 1$. Combining (3.23) and (3.26), we obtain

$$\begin{aligned}
\|x_{2n+2}-z\|^2 &= \|V_{\lambda_{2n+1}}y_{2n+1}+\tilde{e}_{2n+1}-z\|^2 \\
&= \|V_{\lambda_{2n+1}}y_{2n+1}-z\|^2 + 2\langle V_{\lambda_{2n+1}}y_{2n+1}-z, \tilde{e}_{2n+1}\rangle + \|\tilde{e}_{2n+1}\|^2 \\
&\le \|y_{2n+1}-z\|^2 + \|\tilde{e}_{2n+1}\|\bigl(2\|y_{2n+1}-z\| + \|\tilde{e}_{2n+1}\|\bigr) \\
&\le \|x_{2n+1}-z\|^2 + 2\delta_{2n+1}\|x_{2n+1}-x_{2n}\|\bigl(\|x_{2n+1}-z\| + \delta_{2n+1}\|x_{2n+1}-x_{2n}\|\bigr) \\
&\quad + M_3\|\tilde{e}_{2n+1}\|,
\end{aligned} \tag{3.29}$$

where $M_3 = \sup\{2\|y_{2n+1}-z\| + \|\tilde{e}_{2n+1}\|\}$.

With the help of equality (3.28), we have

$$\begin{aligned}
&\|x_{2n+1}-z\|^2 \\
&\quad = \|V_{\lambda_{2n}}y_{2n}+\tilde{e}_{2n}-z\|^2 \\
&\quad = \bigl\|(1-w_{2n})x_{2n}+w_{2n}T_{2n}x_{2n}-z\bigr\|^2 + 2\langle V_{\lambda_{2n}}x_{2n}-z, \tilde{e}_{2n}\rangle + \|\tilde{e}_{2n}\|^2 \\
&\quad \le (1-w_{2n})\|x_{2n}-z\|^2 + w_{2n}\|T_{2n}x_{2n}-T_{2n}z\|^2 - w_{2n}(1-w_{2n})\|T_{2n}x_{2n}-x_{2n}\|^2 \\
&\qquad + \bigl(2\|x_{2n}-z\| + \|\tilde{e}_{2n}\|\bigr)\|\tilde{e}_{2n}\| \\
&\quad \le \|x_{2n}-z\|^2 - w_{2n}(1-w_{2n})\|T_{2n}x_{2n}-x_{2n}\|^2 + M_4\|\tilde{e}_{2n}\|,
\end{aligned} \tag{3.30}$$

where $M_4 = \sup\{2\|x_{2n}-z\| + \|\tilde{e}_{2n}\|\}$.

Substituting (3.30) into (3.29), we get

$$\begin{aligned}
&\|x_{2n+2}-z\|^2 \\
&\quad \le \|x_{2n}-z\|^2 + 2\delta_{2n+1}\|x_{2n+1}-x_{2n}\|\bigl(\|x_{2n+1}-z\| + \delta_{2n+1}\|x_{2n+1}-x_{2n}\|\bigr) \\
&\qquad - w_{2n}(1-w_{2n})\|T_{2n}x_{2n}-x_{2n}\|^2 + M_3\|\tilde{e}_{2n+1}\| + M_4\|\tilde{e}_{2n}\|.
\end{aligned} \tag{3.31}$$

Hence, we have the following result:

$$\begin{aligned}
&w_{2n}(1-w_{2n})\|T_{2n}x_{2n}-x_{2n}\|^2 \\
&\quad \le \|x_{2n}-z\|^2 - \|x_{2n+2}-z\|^2 + 2\delta_{2n+1}\|x_{2n+1}-x_{2n}\|\bigl(\|x_{2n+1}-z\| + \delta_{2n+1}\|x_{2n+1}-x_{2n}\|\bigr) \\
&\qquad + M_3\|\tilde{e}_{2n+1}\| + M_4\|\tilde{e}_{2n}\|.
\end{aligned} \tag{3.32}$$

Noting the fact that $\frac{1}{2} < \liminf_{n\to\infty} w_n \le \limsup_{n\to\infty} w_n < 1$, we deduce from (3.32) that

$$\sum_{n=0}^{\infty} \|T_{2n}x_{2n}-x_{2n}\|^2 < \infty. \tag{3.33}$$

In particular, $\lim_{n\to\infty} \|T_{2n}x_{2n}-x_{2n}\| = 0$. Now we have

$$\|x_{2n+1}-x_{2n}\| \le w_{2n}\|T_{2n}x_{2n}-x_{2n}\| + \|\tilde{e}_{2n}\| \to 0. \tag{3.34}$$

Similarly, we argue that

$$\sum_{n=0}^{\infty} \|T_{2n+1}y_{2n+1} - y_{2n+1}\|^2 < \infty. \tag{3.35}$$

Observe that

$$x_{2n+2} = (1 - w_{2n+1})y_{2n+1} + w_{2n+1}T_{2n+1}y_{2n+1} + \tilde{e}_{2n+1}. \tag{3.36}$$

From (3.35) and condition (ii), we get

$$\|x_{2n+2} - y_{2n+1}\| \le w_{2n+1}\|T_{2n+1}y_{2n+1} - y_{2n+1}\| + \|\tilde{e}_{2n+1}\| \to 0. \tag{3.37}$$

It follows from (3.36) and condition (iv) that

$$\begin{aligned} \|x_{2n+2} - x_{2n+1}\| &\le \|x_{2n+2} - y_{2n+1}\| + \|y_{2n+1} - x_{2n+1}\| \\ &= \|x_{2n+2} - y_{2n+1}\| + \delta_{2n+1}\|x_{2n+1} - x_{2n}\| \to 0. \end{aligned} \tag{3.38}$$

Combining (3.34) and (3.38), we obtain $\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0$. This yields

$$\begin{aligned} \|x_n - V_{\lambda_n}x_n\| &\le \|x_n - x_{n+1}\| + \|x_{n+1} - V_{\lambda_n}y_n\| + \|V_{\lambda_n}y_n - V_{\lambda_n}x_n\| \\ &\le \|x_n - x_{n+1}\| + \|\tilde{e}_n\| + \|y_n - x_n\| \to 0. \end{aligned} \tag{3.39}$$

Step 3. Show that

$$\omega_w(x_n) \subset S. \tag{3.40}$$

Since λ_n is bounded, we may assume that the subsequence λ_{n_k} converges to some λ. It can be proved by a method similar to step 3 in Theorem 3.1. We conclude that (3.40) holds. By Lemma 2.9, we get $\{x_n\}$ converges weakly. □

4 Numerical illustrations

In this section, we consider the following two examples to demonstrate the effectiveness of the algorithms and convergence of Theorem 3.1 and Theorem 3.3.

Example 4.1 Let $H = \mathbb{R}^N$. Define $h(x) = \frac{1}{10}x$. Take $f(x) = \frac{1}{2}\|Ax - b\|^2$, then we obtain that $\nabla f(x) = A^T(Ax-b)$ with Lipschitz constant $L = \|A^TA\|$, where A^T represents the transpose of A. Take $g = \|x\|_1$, then

$$\operatorname{prox}_{\lambda g} x = \arg\min_{v\in H}\left\{\frac{1}{2\lambda}\|v - x\|^2 + \|v\|_1\right\}.$$

In [15], we know that

$$\operatorname{prox}_{\lambda_n\|\cdot\|_1} x = \left[\operatorname{prox}_{\lambda_n|\cdot|} x(1), \operatorname{prox}_{\lambda_n|\cdot|} x(2), \dots, \operatorname{prox}_{\lambda_n|\cdot|} x(N)\right]^T,$$

where $\operatorname{prox}_{\lambda_n|\cdot|} x(i) = \max\{|x(i)| - \lambda_n, 0\} \operatorname{sign}(x(i))$, and $x(i)$ denotes the ith element of x, $i = 1, 2, \ldots, N$. Let D be a diagonal matrix with the element $y_n(i)$. That is, $D_{ii} = y_n(i)$, $i = 1, 2, \ldots, N$. Given $\alpha_n = \frac{1}{100n}$, $\lambda_n = \frac{1}{30L}\frac{n+1}{n+2}$, and

$$\delta_n = \begin{cases} \frac{1}{n^2\|x_n - x_{n-1}\|}, & \|x_n - x_{n-1}\| \neq 0, \\ 0, & \|x_n - x_{n-1}\| = 0 \end{cases}$$

for every $n \geq 0$. Generate an $M * N$ random matrix A whose entries are sampled independently from uniformly distribution. Generate randomly a vector b from a Gaussian distribution of zero mean and unit variance.

According to the iterative process of Theorem 3.1, the sequence $\{x_n\}$ is generated by

$$\begin{cases} y_n = x_n + \delta_n(x_n - x_{n-1}), \\ x_{n+1} = \alpha_n h(y_n) + (1 - \alpha_n)(\operatorname{prox}_{\lambda_n g}(y_n - \lambda_n D(y_n)A^T(Ay_n - b) + e(y_n)). \end{cases} \tag{4.1}$$

Next, we use MATLAB software for numerical implementation. Set $M = 100$, $N = 1000$. Under the same parameters, contrast with iterative algorithm (4.2) in reference [6]. Take different error limit ϵ, we obtain the numerical experiment results in Table 1, where n and t denote the iterative number and running time(tic/toc), respectively. We use $\|x_{n+1} - x_n\| < \epsilon$ as the stopping criteria.

$$x_{n+1} = \alpha_n h(x_n) + (1 - \alpha_n)(\operatorname{prox}_{\lambda_n g}\big(x_n - \lambda_n DA^T(Ax_n - b) + e(x_n)\big). \tag{4.2}$$

In addition, we compare the values of $\|x_{n+1} - x_n\|$ at the same number of iterations of (4.1) and (4.2). The results can be seen in Fig. 1. We also present different running time and the number of iterations at different stopping criteria ϵ. See Fig. 2.

It can be easily seen from Table 1, Fig. 1, and Fig. 2 that Algorithm 1 is faster than iterative formula (4.2) without inertial step. At the same stopping criteria, the values of $\|x_{n+1} - x_n\|$ and $\|Ax_n - b\|$ of Algorithm 1 are smaller.

In what follows, we give an example in an infinite dimensional space.

Example 4.2 Suppose that $H = L^2([0, 1])$ with the norm $\|x\| = (\int_0^1 (x(t))^2\,dt)^{\frac{1}{2}}$ and the inner product $\langle x, y\rangle = \int_0^1 x(t)y(t)\,dt$, $\forall x, y \in H$. Define $h(x) = \frac{1}{2}x$ and $Ax(t) = tx(t)$. Let $f(x) = \frac{1}{2}\|Ax(t) - u(t)\|^2$ and $g(x)$ be the indicator function of C, respectively, where $u(t) \in H$ is a fixed function and $C = \{x \in H \mid \|x\| \leq 1\}$.

By the definition of f and g, we obtain

$$\nabla f(x) = A^*(Ax - u)$$

Table 1 Comparison of Algorithm 1 (IA) with the algorithm without inertia step (UA) for Example 4.1. $x_0 = \operatorname{randn}(N, 1)$

ϵ	IA				UA			
	n	t	$\|x_{n+1} - x_n\|$	$\|Ax_n - b\|$	n	t	$\|x_{n+1} - x_n\|$	$\|Ax_n - b\|$
10^{-3}	143	0.0546	7.8635×10^{-7}	7.9601	685	0.4510	9.8753×10^{-4}	8.7650
10^{-5}	636	0.8414	6.6975×10^{-8}	2.8313	5768	3.8915	9.2786×10^{-5}	3.6433
10^{-7}	1391	1.0010	8.3482×10^{-9}	0.8189	14,768	6.0098	9.9947×10^{-8}	1.3663
10^{-9}	2023	1.0035	1.7182×10^{-12}	0.1473	56,077	7.0788	9.9989×10^{-10}	0.5936

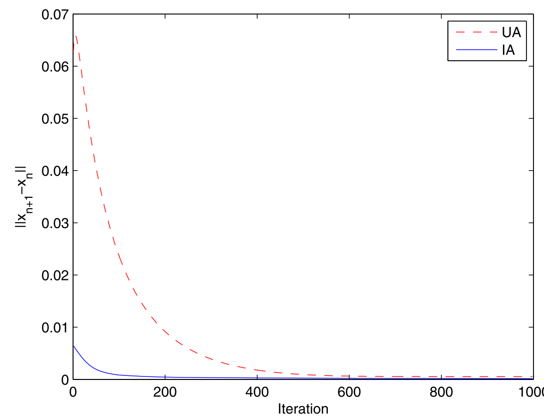

Figure 1 The comparison of $\|x_{n+1} - x_n\|$ of inertial acceleration (IA) and without inertial acceleration (UA) for $(M, N) = (100, 1000)$ of Example 4.1

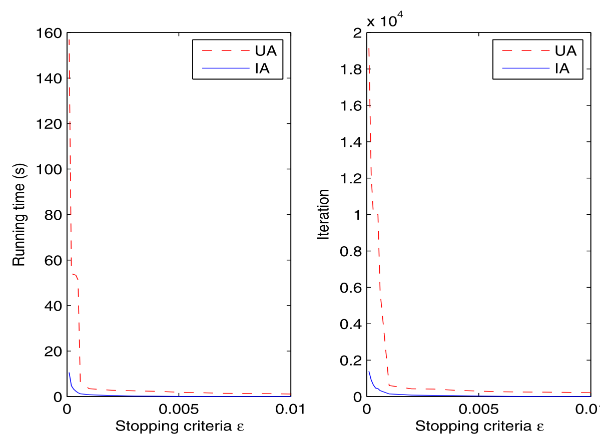

Figure 2 The comparison of running time and iteration steps of inertial acceleration (IA) and without inertial acceleration (UA) with the same stopping criteria for $(M, N) = (100, 1000)$ of Example 4.1

and

$$\operatorname{prox}_{\lambda g} x = \arg\min_{v \in H} \left\{ \frac{1}{2\lambda} \|v - x\|^2 + \iota_C(v) \right\} = P_C(x),$$

where ι_C denotes the indicator function and

$$\iota_C(x) = \begin{cases} 0, & \text{if } x \in C, \\ \infty, & \text{if } x \notin C. \end{cases}$$

We also deduce the adjoint operator of A is still A, i.e., $A^* = A$. Take $D(x_n) = I$, set the parameters $\alpha_n = \frac{1}{1000n}$ and $\lambda_n = \frac{n}{L*(n+1)}$, according to the iterative algorithm of Theorem 3.1, we get the following sequence $\{x_n\}$:

$$\begin{cases} y_n = x_n + \delta_n(x_n - x_{n-1}), \\ x_{n+1} = \frac{1}{1000n} \frac{1}{2} y_n + (1 - \frac{1}{1000n}) P_C(y_n - \frac{n}{L(n+1)} A(Ay_n - u)). \end{cases}$$

Table 2 Comparison of Algorithm 3 (AIA) with Algorithm 1 (IA) for Example 4.2. $u = e^t$, $x_0 = t$, $x_1 = t^2$

ϵ	AIA				IA			
	n	t	$\|x_{n+1} - x_n\|$	$\|x_n\|$	n	t	$\|x_{n+1} - x_n\|$	$\|x_n\|$
10^{-3}	17	0.0007	9.0023×10^{-4}	1.0001	21	0.0011	9.0302×10^{-4}	1.0001
10^{-7}	101	0.0025	9.9998×10^{-8}	1.0000	132	0.0034	9.9989×10^{-8}	1.0000
10^{-8}	316	0.0063	8.3482×10^{-9}	1.0000	715	0.0072	9.9947×10^{-8}	1.0000
10^{-9}	996	0.0206	1.7182×10^{-12}	1.0000	1087	0.0317	9.9989×10^{-10}	1.0000

Table 3 Comparison of Algorithm 3 (AIA) with Algorithm 1 (IA) for Example 4.2. $u = \sin t$, $x_0 = t$, $x_1 = 2t$

ϵ	AIA				IA			
	n	t	$\|x_{n+1} - x_n\|$	$\|x_n\|$	n	t	$\|x_{n+1} - x_n\|$	$\|x_n\|$
10^{-3}	127	0.0051	9.3714×10^{-4}	1.0001	130	0.0053	9.7883×10^{-4}	1.0001
10^{-5}	525	0.0145	8.2615×10^{-6}	1.0000	534	0.0182	9.1975×10^{-6}	1.0000
10^{-7}	1052	0.0299	9.4418×10^{-8}	1.0000	1077	0.0343	9.8397×10^{-8}	1.0000
10^{-8}	2011	0.0611	9.9554×10^{-9}	1.0000	2071	0.0697	9.9885×10^{-9}	1.0000

The numerical integration method used in this example is the trapezoidal formula. We test these two algorithms with different stopping criteria. The numerical results are shown in Table 2.

In what follows, we present a comparison of inertial proximal gradient algorithm (IA) and alternated inertial proximal gradient algorithm (AIA). Set $e(y_n) = \frac{1}{n^2}$ as the outer perturbation, the numerical results are reported in Table 3.

It is observed that the norm of x_n is close to 1 with the increase of iteration steps. From this example, the alternated inertia algorithm needs fewer iterations and less running time than inertia algorithm, but there is not much difference between the two algorithms.

Acknowledgements
The authors would like to thank the referee for valuable suggestions to improve the manuscript.

Funding
The authors thank the Foundation of Tianjin Key Lab for Advanced Signal Processing (2019ASP-TJ04) and the Scientific Research Project of Tianjin Municipal Education Commission (2019KJ133) for support.

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

Author details
[1]College of Science, Tianjin Key Lab for Advanced Signal Processing, Civil Aviation University of China, Tianjin 300300, China. [2]College of Science, Civil Aviation University of China, Tianjin 300300, China.

Publisher's Note

Received: 18 August 2020 Accepted: 26 November 2020 Published online: 07 December 2020

References

1. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**, 183–202 (2009)
2. Boikanyo, O.A., Morosanu, G.: Strong convergence of a proximal point algorithm with bounded error sequence. Optim. Lett. **72**, 415–420 (2013)

3. Censor, Y., Davidi, R., Herman, G.T.: Perturbation resilience and superiorization of iterative algorithms. Inverse Probl. **26**, 065008 (2010)
4. Davidi, R., Herman, G.T., Censor, Y.: Perturbation-resilient block-iterative projection methods with application to image reconstruction from projections. Int. Trans. Oper. Res. **16**, 505–524 (2009)
5. Dong, Q.L., Zhao, J., He, S.N.: Bounded perturbation resilience of the viscosity algorithm. J. Inequal. Appl. **2016**, 299 (2016)
6. Duan, P.C., Song, M.M.: General viscosity iterative approximation for solving unconstrained convex optimization problems. J. Inequal. Appl. **2015**, 334 (2015)
7. Geobel, K., Kirk, W.A.: Topics in Metric Fixed Point Theory. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge (1990)
8. Guo, Y.N., Cui, W.: Strong convergence and bounded perturbation resilience of a modified proximal gradient algorithm. J. Inequal. Appl. **2018**, 103 (2018)
9. He, S.N., Yang, C.P.: Solving the variational inequality problem defined on intersection of finite level sets. Abstr. Appl. Anal. **2013**, Article ID 942315 (2013)
10. Iutzeler, F., Hendricks, M.: A generic online acceleration scheme for optimization algorithms via relaxation and inertia. Optim. Methods Softw. **34**, 383–405 (2019)
11. Jin, W., Censor, Y., Jiang, M.: Bounded perturbation resilience of projected scaled gradient methods. Comput. Optim. Appl. **63**, 365–392 (2016)
12. Mahammand, A.A., Naseer, S., Xu, H.K.: Properties and iterative methods for *Q*-lasso. Abstr. Appl. Anal. **8**, Article ID 250943 (2013)
13. Mangasarian, O.L., Meyer, R.R., Robinson, S.M., Auslender, A.: Minimisation de fonctions localement Lipschitiziwnnes: appligramming a la programmation mi-convexe, mi-differential. In: Nonlinear Programming 3. Academic Press, New York (1978)
14. Marino, G., Xu, H.K.: Weak and strong convergence theorems for strict pseudo-contractions in Hilbert spaces. J. Math. Anal. Appl. **329**, 336–346 (2007)
15. Micchelli, C.A., Shen, L.X., Xu, Y.S.: Proximity algorithms for image models: denoising. Inverse Probl. **27**, 045009 (2011)
16. Moreau, J.J.: Proprietes des applications 'prox'. C. R. Acad. Sci. Paris Sér. A Math. **256**, 1069–1071 (1963)
17. Mu, Z.G., Peng, Y.: A note on the inertial proximal point method. Stat. Optim. Inf. Comput. **3**, 241–248 (2015)
18. Nesterov, Y.: A method for solving the convex programming problem with convergence rate $O(1/k^2)$. Dokl. Akad. Nauk SSSR **269**, 543–547 (1983)
19. Polyak, B.T.: Some methods of speeding up the convergence of iteration methods. USSR Comput. Math. Math. Phys. **4**, 1–17 (1964)
20. Polyak, B.T.: Introduction to Optimization. Optimization Software Inc. Publications Division, New York (1987)
21. Shehu, Y., Gibali, A.: New inertial relaxed method for solving split feasibilities. Optim. Lett. (2020). https://doi.org/10.1007/s11590-020-01603-1
22. Xu, H.K.: Averaged mappings and the gradient-projection algorithm. J. Optim. Theory Appl. **150**, 360–378 (2011)
23. Xu, H.K.: Properties and iterative methods for the lasso and its variants. Chin. Ann. Math. **35B**(3), 1–18 (2014)
24. Xu, H.K.: Bounded perturbation resilience and superiorization techniques for the projected scaled gradient method. Inverse Probl. **33**, 044008 (2017)
25. Yao, Z.S., Cho, S.Y., Kang, S.M., Zhu, L.J.: A regularized algorithm for the proximal split feasibility problem. Abstr. Appl. Anal. **2014**, Article ID 894272 (2014)