



Superiorization methodology and perturbation resilience of inertial proximal gradient algorithm with application to signal recovery

Nuttapol Pakkaranang^{1,2} · Poom Kumam^{1,2,3}  · Vasile Berinde⁴ · Yusuf I. Suleiman⁵

Published online: 27 February 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

In this paper, we construct a novel algorithm for solving non-smooth composite optimization problems. By using inertial technique, we propose a modified proximal gradient algorithm with outer perturbations, and under standard mild conditions, we obtain strong convergence results for finding a solution of composite optimization problem. Based on bounded perturbation resilience, we present our proposed algorithm with the superiorization method and apply it to image recovery problem. Finally, we provide the numerical experiments to show efficiency of the proposed algorithm and comparison with previously known algorithms in signal recovery.

Keywords Composite optimization problems · Proximal gradient algorithm · Inertial algorithm · Superiorization methodology · Bounded perturbation resilience

Mathematics Subject Classification 47H05 · 47H09 · 49M37 · 65K10

1 Introduction

Let $\Gamma_0(H)$ be a class of convex, lower semicontinuous, and proper functions from a Hilbert space H to $(\infty, +\infty]$. The *non-smooth composite optimization problem* (shortly, NSCOP) is defined by

$$\min_{x \in H} (\psi(x) + \varphi(x)), \quad (1)$$

where $\psi, \varphi \in \Gamma_0(H)$, ψ is differentiable, φ is not necessarily differentiable, and $\nabla\psi$ is Lipschitz continuous on H . The NSCOP can be traced back to classical work of Bauschke and Combettes [1] and it also has a typical research fields in linear inverse

✉ Poom Kumam
poom.kum@kmutt.ac.th

Extended author information available on the last page of the article

problems [2]. This class of optimization problem covers a large number of applications in applied mathematics, robotics, medical imaging, financial engineering, machine learning, signal processing and resource allocation, see e.g. [3–9]. Due to recent boom in automations and machine learning, smart iterative algorithms are indispensable in the fields of artificial intelligence. This, however, has led to an increase in demand for feasible and faster algorithms in which all the relevant components can be evaluated sufficiently quickly and at least time. One of the goals in the present paper is to seek to explore a novel and faster algorithm for solving NSCOP.

The important and well-known power tool for solving the problem (1) is *proximal gradient algorithm*. It is as an analogous tool for non-smooth, constrained, large-scale, or distributed versions of Newton’s method for solving unconstrained smooth optimization problems of modest size.

Assumption 1 [1]: (The existence of solution of NSCOP)

Let Ω be the set of all solutions of (1). If $\psi(x) + \varphi(x) := \Phi$ is coercive, that is,

$$\lim_{\|x\| \rightarrow \infty} \Phi(x) = +\infty.$$

Then $\Omega = \text{Arg min } \Phi \neq \emptyset$. This means that Φ has a minimizer over H .

Assumption 2 [1]: (Uniqueness of Solution of Proximal Mappings)

Let $\varphi \in \Gamma_0(H)$, then for each x , the function

$$\left(\frac{\|y - x\|^2}{2} + \varphi(y) \right) \text{ has exactly one minimizer over } H.$$

In view of this, the proximal operator of φ of order $\lambda > 0$ is defined by

$$\mathbf{prox}_{\lambda\varphi}(x) := \operatorname{argmin}_{y \in H} \left\{ \frac{\|y - x\|^2}{2\lambda} + \varphi(y) \right\}, \quad x \in H. \tag{2}$$

The *proximal gradient algorithms* (shortly, PGA) are used to split the contribution of the functions ψ and φ , in particular, the gradient descent step determined by ψ [10, 11] and the proximal step induced by φ [12–14]. PGA is easy to implement and applicable in problems that involve large or high-dimensional datasets. Xu [15] submitted that x^* is a solution of the NSCOP (1) if and only if x^* solves the fixed point equation

$$x^* = \mathbf{prox}_{\lambda\varphi}(x^* - \lambda \nabla \psi(x^*)). \tag{3}$$

He then established the weak convergence of PGA as follows:

$$x_{n+1} = \mathbf{prox}_{\lambda\varphi}(x_n - \lambda_n \nabla \psi(x_n)), \quad \forall n \geq 0, \tag{4}$$

where $\lambda_n > 0$ is a step size. Recently, Guo and Cui [16] modified the PGA by using viscosity algorithm [17] as follows:

$$x_{n+1} = \alpha_n f(x_n) + (1 - \alpha_n) \text{prox}_{\lambda_n \varphi}(x_n - \lambda_n \nabla \psi(x_n)) + e(x_n), \quad \forall n \geq 0, \tag{5}$$

where $\{\alpha_n\}$ is a sequence in $[0, 1]$, $f : H \rightarrow H$ is a $\rho \in (0, 1)$ contractive operator, $L > 0$ a Lipschitz constant of ψ and $e : H \rightarrow H$ represents a perturbation operator satisfying

$$\sum_{n=0}^{\infty} \|e(x_n)\| < +\infty.$$

They proved that the sequence $\{x_n\}$ generated by (5) converges strongly to a point $x^* \in \Omega$, where x^* is the unique solution of the *variational inequality problem*:

$$\langle x^* - f(x^*), x - x^* \rangle \geq 0, \quad \forall x \in \Omega, \tag{6}$$

under the following conditions:

- (C1) $0 < a = \inf_n \lambda_n \leq \lambda_n < \frac{2}{L}$ and $\sum_{n=0}^{\infty} \|\lambda_{n+1} - \lambda_n\| < \infty$;
- (C2) $\{\alpha_n\} \subset (0, 1)$ and satisfying $\lim_{n \rightarrow \infty} \alpha_n = 0$;
- (C3) $\sum_{n=0}^{\infty} \alpha_n = \infty$ and $\sum_{n=0}^{\infty} \|\alpha_{n+1} - \alpha_n\| < \infty$.

The superiorization methodology (shortly, SM) was invented by Censor et al. [18] in attempt to check computation tractability of certain image recovery algorithms with limited computing resources. The SM is a heuristic approach with focus on time consumption [19, 20]. The base operation of SM requires that an algorithm which generates a convergent sequence of feasible solutions is bounded perturbation resilient, and that superiorization which uses perturbations proactively to get superior feasible solutions in the sense that the values of the objective function decrease is put in force. The technique is to find a point with a lower cost function value than other points through a simple algorithm is called the *basic algorithm*. This basic algorithm is usually pre-checked for perturbation resilience. The SM seems to be vast oversimplification, yet it represents a dramatic slight in approach to implementation of iterative algorithms. The SM has been investigated and studied by many researchers (see e.g. [6, 21–23]). Very recently, Guo and Cui showed that the sequence generated by (5) is perturbation resilient and prove strong convergence results which is also solution of variational inequality problem (6).

In optimization theory, the technique to increase efficiency of convergence rate, Polyak [24] pioneered the use of heavy ball method of the two-order time dynamical system which is a two-step iterative method for solving smooth convex minimization problem (i.e. a problem where all the constraints are convex functions, and the objective function is a smooth convex function). Later on, the development of this method was emphasized and used to increase the performance the convergence rate by Nesterov [25] as follows:

$$y_n = x_n + \zeta_n(x_n - x_{n-1}), \tag{7}$$

$$x_{n+1} = y_n - \lambda_n \nabla f(y_n), \quad \forall n \geq 1, \tag{8}$$

where $\zeta_n \in [0, 1)$ is an extrapolation factor, λ_n is a positive real sequence and ∇f is the gradient of a smooth convex function f . The method is effective and converges faster since the vector $(x_n - x_{n-1})$ acts as an impulsion term and ζ_n serves as a speed regulator of the momentum term $\zeta_n(x_n - x_{n-1})$ in (7). For more about inertial algorithms, we refer to [26–32] and the references therein.

Inspired and motivated by some contributions of authors mentioned above, we study a new method of approximation of solutions of the NSCOP (1) using a modified proximal gradient algorithm combined with inertial technique, and we then prove its strong convergence theorem of the sequence generated by the proposed algorithm under suitable conditions. Further, we verify that the proposed algorithm is of bounded perturbation resilient and apply it to image recovery problem. The effectiveness and the implementation of the proposed algorithm were illustrated by comparing it with previously known algorithms through numerical experiments in signal recovery.

The rest of this paper is organized as follows: basic notations, definitions and lemmas are recalled in Sect. 2. Our main results are presented in Sect. 3, that is, the inertial modified proximal gradient algorithm involving both perturbation and superiorization method are presented and some strong convergence results are proved. Next, Sect. 4 contains an application and numerical experiments to support our study. Finally, the conclusion of this research is given.

2 Preliminaries

The following notations, definitions and lemmas will play a crucial role in the sequel. Let $\{x_n\}$ be a sequence in Hilbert space H and $x \in H$.

- (1) $x_n \rightarrow x$ means $\{x_n\}$ converges strongly to x ;
- (2) $x_n \rightharpoonup x$ means $\{x_n\}$ converges weakly to x ;
- (3) A point z is said to be a cluster point of $\{x_n\}$ if there exists a subsequence $\{x_{n_j}\}$ which converges weakly to a point z . The set of all cluster points of $\{x_n\}$ is denoted by $\omega_w(x_n)$.
- (4) An operator $T : H \rightarrow H$ is said to be λ -averaged if $T = (1 - \lambda)I + \lambda S$, where $\lambda \in (0, 1)$ and $S : H \rightarrow H$ is nonexpansive;
- (5) $A : H \rightarrow H$ is said to be ν -inverse strongly monotone (ν -ism) with $\nu > 0$, if

$$\langle Ax - Ay, x - y \rangle \geq \|Ax - Ay\|^2 \quad \forall x, y \in H;$$

- (6) T is nonexpansive if and only if the complement $I - T$ is $\frac{1}{2}$ -ism [33];
- (7) If T is ν -ism, then for $\gamma > 0$, γT is $\frac{\nu}{\gamma}$ -ism;

- (8) T is averaged if and only if the complement $I - T$ is ν -ism for some $\nu > \frac{1}{2}$. Indeed, T is α -averaged if and only if $I - T$ is $\frac{1}{2\alpha}$ -ism, for $\alpha \in (0, 1)$;
- (9) The composite of finitely many averaged operators is averaged. That is, if for each $i = 1, \dots, N$, the operator T_i is averaged, then so is the composite operator $T_1 \dots T_N$. In particular, if T_1 is α_1 -averaged and T_2 is α_2 -averaged, where $\alpha_1, \alpha_2 \in (0, 1)$, then the composite $T_1 T_2$ is α -averaged, where $\alpha = \alpha_1 + \alpha_2 - \alpha_1 \alpha_2$.

Lemma 1 *The following results are well known.*

- (i) $\|x + y\|^2 = \|x\|^2 + 2\langle x, y \rangle + \|y\|^2, \quad \forall x, y \in H$;
- (ii) $\|x + y\|^2 \leq \|x\|^2 + 2\langle y, x + y \rangle, \quad \forall x, y \in H$.

Lemma 2 [15] *Let $\varphi \in \Gamma_0(H)$, $\lambda > 0$ and $\mu > 0$. Then*

$$\mathbf{prox}_{\lambda\varphi}(x) = \mathbf{prox}_{\mu\varphi}\left(\frac{\mu}{\lambda}x + \left(1 - \frac{\mu}{\lambda}\right)\mathbf{prox}_{\lambda\varphi}x\right).$$

Lemma 3 [34] *Let $\varphi \in \Gamma_0(H)$ and $\lambda > 0$. Then the proximal operator $\mathbf{prox}_{\lambda\varphi}$ is $\frac{1}{2}$ -average. In particular, it is nonexpansive:*

$$\|\mathbf{prox}_{\lambda\varphi}(x) - \mathbf{prox}_{\lambda\varphi}(y)\| \leq \|x - y\|, \quad \forall x, y \in H.$$

Lemma 4 [1] *Let $T : H \rightarrow H$ be a nonexpansive mapping with $\text{Fix}(T) \neq \emptyset$. If a sequence $\{x_n\}$ in H converges weakly to x and $\{(I - T)x_n\}$ converges strongly to y , then $(I - T)x = y$, where $\text{Fix}(T)$ is a fixed point set of T .*

Lemma 5 [21] *Assume that $\{a_n\}$ is a sequence of nonnegative real numbers satisfying*

$$a_{n+1} \leq (1 - \gamma_n)a_n + \gamma_n\delta_n + \beta_n$$

where $\{\gamma_n\}, \{\delta_n\}$, and $\{\beta_n\}$ satisfy the conditions:

- (i) $\{\gamma_n\} \subset [0, 1], \quad \sum_{n=0}^{\infty} \gamma_n = \infty$;
- (ii) $\limsup_{n \rightarrow \infty} \delta_n \leq 0$;
- (iii) $\beta_n \geq 0$ and $\sum_{n=0}^{\infty} \beta_n < \infty$. Then $\lim_{n \rightarrow \infty} a_n = 0$.

3 Main results

In this section, using the proximal mapping, we demonstrate a prove of a strong convergence theorem for finding a solution of the non-smooth convex composite optimization problem (1).

Assumption 3

- (i) $f : H \rightarrow H$ is a ρ -contractive operator with $\rho \in (0, 1)$;
- (ii) $\psi : H \rightarrow (-\infty, \infty]$ is proper, lower semicontinuous, convex and $\nabla\psi$ is Lipschitz continuous with Lipschitz constant $L > 0$;
- (iii) $\varphi : H \rightarrow (-\infty, \infty]$ is proper, lower semicontinuous and convex;
- (iv) The set Ω of all solutions of problem 1 is nonempty.

Assumption 4 Suppose $\{\lambda_n\} \subset (0, \frac{2}{L})$, $\{\zeta_n\} \subset [0, 1)$ and $\{\alpha_n\}$ is a sequence in $(0, 1)$ satisfying the following conditions:

- (B1) $0 < a = \inf_n \lambda_n \leq \lambda_n < \frac{2}{L}$ and $\sum_{n=0}^\infty \|\lambda_{n+1} - \lambda_n\| < \infty$;
- (B2) $\lim_{n \rightarrow \infty} \alpha_n = 0$, and $\sum_{n=0}^\infty \alpha_n = \infty$
- (B3) $\lim_{n \rightarrow \infty} \frac{\zeta_n}{\alpha_n} \|x_n - x_{n-1}\| = 0$.

Algorithm 3.1 : Inertial viscosity proximal gradient algorithm with perturbation

Initialization: Take arbitrary $x_0, x_1 \in H$, and set $n = 1$. Choose sequences $\{\alpha_n\}$ and $\{\zeta_n\}$ such that the conditions in Assumption 4 hold.

Step 1: If $\|x_n - x_{n-1}\| = 0$. **STOP**, (x_n) is a solution of **NSCOP (1)**. Otherwise, go to Step 2.

Step 2: For a given current x_{n-1}, x_n and ζ_n . Compute θ_n as

$$\theta_n = x_n + \zeta_n(x_n - x_{n-1}).$$

Step 3: Compute the next iterate (x_{n+1}) as:

$$x_{n+1} = \alpha_n f(\theta_n) + (1 - \alpha_n) \text{prox}_{\lambda_n \varphi}(\theta_n - \lambda_n \nabla \psi(\theta_n)) + e(\theta_n), \quad n \geq 0.$$

where $e : H \rightarrow H$ is a perturbation operator satisfying the following condition

$$\sum_{n=0}^\infty \|e(\theta_n)\| < +\infty. \tag{9}$$

Last step: Update $n := n + 1$ and go to **Step1** .

If $\zeta_n = 0$ in Algorithm 3.1, then we obtain the viscosity proximal gradient algorithm with perturbation (5) of Guo and Cui [16].

Definition 1 (Bounded Perturbation Resilience)

Let $\{x_n\}$ be a sequence generated by the algorithm

$$\begin{cases} \text{select arbitrary starting point } x_0 \in H, \\ x_{n+1} = Ax_n, \quad n \geq 0, \end{cases}$$

where A is an algorithmic operator from H into H . Then A is said to be *bounded perturbation resilient* if

$$\lim_{n \rightarrow \infty} x_n = v \text{ implies } \lim_{n \rightarrow \infty} y_n = v, \text{ for any sequence } \{y_n\} \text{ in } H \text{ generated by}$$

$$\begin{cases} \text{select arbitrary starting point } y_0 \in H, \\ y_{n+1} = A(y_n + \beta_n v_n) \quad n \geq 0, \end{cases}$$

for a bounded vector sequence $\{v_n\}$ and scalar $\{\beta_n\}$ such that $\beta_n \geq 0 \forall n \geq 0$, and $\sum_{n=0}^\infty \beta_n < \infty$. This operator A is called the *basic algorithm*.

Algorithm 3.2: Inertial viscosity proximal gradient algorithm with superiorization method

Initialization: Take arbitrary $x_0, x_1 \in H$, and set $n = 1$. Choose sequences $\{v_n\}$ and scalar $\{\beta_n\}$ as in Definition 1. Choose $\{\alpha_n\}$ and $\{\zeta_n\}$ such that the conditions in Assumption 4 hold.

Step 1: If $\|x_n - x_{n-1}\| = 0$. **STOP** (x_n) is a solution of NSCOP (1). Otherwise, go to Step 2.

Step 2: For a given current x_{n-1}, x_n and ζ_n . Compute θ_n as

$$\theta_n = x_n + \zeta_n(x_n - x_{n-1}).$$

Step 3: Compute the next iterate (x_{n+1}) as:

$$x_{n+1} = \alpha_n f(\theta_n + \beta_n v_n) + (1 - \alpha_n) \mathbf{prox}_{\lambda_n \varphi}(\theta_n + \beta_n v_n - \lambda_n \nabla \psi(\theta_n + \beta_n v_n)), \quad n \geq 0.$$

Last step: Update $n := n + 1$ and go to **Step1**.

Theorem 1 Assume that Assumption 3 holds. Then the sequence $\{x_n\}$ generated by Algorithm 3.1 is bounded.

Proof Step 1: We prove that $\mathbf{prox}_{\lambda_n \varphi}(I - \lambda_n \nabla \psi)$ is nonexpansive for each n . Since $\nabla \psi$ is L -Lipschitzian then $\nabla \psi$ is $\frac{1}{L}$ -ism which implies that $\lambda_n \nabla \psi$ is $\frac{1}{\lambda_n L}$ -ism and by (8), the complement $(I - \lambda_n \nabla \psi)$ is $\frac{\lambda_n L}{2}$ -average as $\lambda_n \in (0, \frac{2}{L})$. This implies that $\mathbf{prox}_{\lambda_n \varphi}$ is $\frac{1}{2}$ -averaged. This further implies that the composition $\mathbf{prox}_{\lambda_n \varphi}(I - \lambda_n \nabla \psi)$ is $\frac{\lambda_n L + 2}{4}$ -averaged. Therefore the composition

$$\mathbf{prox}_{\lambda_n \varphi}(I - \lambda_n \nabla \psi) = \left(\frac{2 - \lambda_n L}{4}\right)I + \left(\frac{2 + \lambda_n L}{4}\right)T = (1 - \alpha)I + \alpha T \text{ is nonexpansive,}$$

for each n .

Step 2: We show that

$$\|\theta_n - x^*\| \leq \|x_n - x^*\| + \alpha_n M_1. \tag{10}$$

Let $x^* \in \Omega$, and $T_n = \mathbf{prox}_{\lambda_n \varphi}(I - \lambda_n \nabla \psi)$. From Algorithm 3.1, we compute

$$\begin{aligned} \|\theta_n - x^*\| &= \|x_n + \zeta_n(x_n - x_{n-1}) - x^*\| \\ &\leq \|x_n - x^*\| + \zeta_n \|x_n - x_{n-1}\| \\ &= \|x_n - x^*\| + \alpha_n \frac{\zeta_n}{\alpha_n} \|x_n - x_{n-1}\|. \end{aligned}$$

By condition (B3) in Assumption 4. Then there exists a constant $M_1 \geq 0$ such that

$$\frac{\zeta_n}{\alpha_n} \|x_n - x_{n-1}\| \leq M_1, \quad \forall n \geq 0.$$

Step 3: We show that

$$\|x_{n+1} - x^*\| \leq \max \left\{ \|x_0 - x^*\|, \frac{M_1 + \|fx^* - x^*\|}{1 - \rho} \right\} + \sum_{n=0}^{\infty} \|e(\theta_n)\|. \tag{11}$$

It follows from Algorithm 3.1, we compute

$$\begin{aligned} \|x_{n+1} - x^*\| &= \|\alpha_n f(\theta_n) + (1 - \alpha_n)T_n(\theta_n) + e(\theta_n) - x^*\| \\ &\leq \alpha_n \|f(\theta_n) - x^*\| + (1 - \alpha_n) \|T_n(\theta_n) - x^*\| + \|e(\theta_n)\| \\ &\leq \alpha_n \|f(\theta_n) - f(x^*)\| + \alpha_n \|f(x^*) - x^*\| + (1 - \alpha_n) \|T_n(\theta_n) - x^*\| + \|e(\theta_n)\| \\ &\leq \rho \alpha_n \| \theta_n - x^* \| + \alpha_n \|f(x^*) - x^*\| + (1 - \alpha_n) \| \theta_n - x^* \| + \|e(\theta_n)\| \\ &= (1 - \alpha_n(1 - \rho)) \| \theta_n - x^* \| + \alpha_n(1 - \rho) \frac{\|fx^* - x^*\|}{1 - \rho} + \|e(\theta_n)\| \\ &= (1 - \alpha_n(1 - \rho)) (\|x_n - x^*\| + \alpha_n M_1) + \alpha_n(1 - \rho) \frac{\|fx^* - x^*\|}{1 - \rho} + \|e(\theta_n)\| \\ &= (1 - \alpha_n(1 - \rho)) \|x_n - x^*\| + \alpha_n(1 - \rho) \frac{(M_1 + \|fx^* - x^*\|)}{1 - \rho} + \|e(\theta_n)\| \\ &\leq \max \left\{ \|x_n - x^*\|, \frac{M_1 + \|fx^* - x^*\|}{1 - \rho} \right\} + \|e(\theta_n)\|. \end{aligned}$$

By mathematical induction, we derive relation (11). Using condition (9), we deduce that $\{x_n\}$ is bounded. □

Theorem 2 Assume that Assumption 3 holds. Then the sequence $\{x_n\}$ generated by Algorithm 3.1 converges strongly to a point $x^* \in \Omega$, satisfying

$$\langle x^* - f(x^*), x - x^* \rangle \geq 0, \quad \forall x \in \Omega. \tag{12}$$

Proof Step 1: We show that

$$\|\theta_n - \theta_{n-1}\| \leq \|x_n - x_{n-1}\| + (\alpha_n + \alpha_{n-1})M_1. \tag{13}$$

By using Algorithm 3.1, we get

$$\begin{aligned} \|\theta_n - \theta_{n-1}\| &= \|x_n + \zeta_n(x_n - x_{n-1}) - x_{n-1} - \zeta_{n-1}(x_{n-1} - x_{n-2})\| \\ &= \|x_n - x_{n-1} + \zeta_n(x_n - x_{n-1}) - \zeta_{n-1}(x_{n-1} - x_{n-2})\| \\ &\leq \|x_n - x_{n-1}\| + \zeta_n \|x_n - x_{n-1}\| + \zeta_{n-1} \|x_{n-1} - x_{n-2}\| \\ &\leq \|x_n - x_{n-1}\| + \alpha_n M_1 + \alpha_{n-1} M_1 \\ &= \|x_n - x_{n-1}\| + (\alpha_n + \alpha_{n-1}) M_1. \end{aligned}$$

Step 2: We show that

$$\|T_n(\theta_{n-1}) - T_{n-1}(\theta_{n-1})\| \leq \frac{|\lambda_n - \lambda_{n-1}|}{a} \|\theta_{n-1} - T_{n-1}(\theta_{n-1})\|, \quad \text{where } a = \inf_{n \in \mathbb{N}} \lambda_n > 0. \tag{14}$$

We compute as follows:

$$\begin{aligned} &\|T_n(\theta_{n-1}) - T_{n-1}(\theta_{n-1})\| \\ &= \|\mathbf{prox}_{\lambda_n \varphi}(\theta_n - \lambda_n \nabla \psi(\theta_n)) - \mathbf{prox}_{\lambda_{n-1} \varphi}(\theta_{n-1} - \lambda_{n-1} \nabla \psi(\theta_{n-1}))\| \\ &= \|\mathbf{prox}_{\lambda_n \varphi}(\theta_{n-1} - \lambda_n \nabla \psi(\theta_{n-1})) \\ &\quad - \mathbf{prox}_{\lambda_n \varphi}[\frac{\lambda_n}{\lambda_{n-1}}(\theta_{n-1} - \lambda_{n-1} \nabla \psi(\theta_{n-1})) \\ &\quad + (1 - \frac{\lambda_n}{\lambda_{n-1}})\mathbf{prox}_{\lambda_{n-1} \varphi}(\theta_{n-1} - \lambda_{n-1} \nabla \psi(\theta_{n-1}))]\| \\ &\leq \|(\theta_{n-1} - \lambda_n \nabla \psi(\theta_{n-1})) - \frac{\lambda_n}{\lambda_{n-1}}(\theta_{n-1} - \lambda_{n-1} \nabla \psi(\theta_{n-1})) \\ &\quad - (1 - \frac{\lambda_n}{\lambda_{n-1}})T_{n-1}(\theta_{n-1})\| \\ &= |1 - \frac{\lambda_n}{\lambda_{n-1}}| \|\theta_{n-1} - T_{n-1}(\theta_{n-1})\| \\ &\leq \frac{|\lambda_n - \lambda_{n-1}|}{a} \|\theta_{n-1} - T_{n-1}(\theta_{n-1})\|. \end{aligned}$$

Step 3: We show that

$$\lim_{n \rightarrow \infty} \|x_{n+1} - x_n\| = 0. \tag{15}$$

It follows from Algorithm 3.1, we compute

$$\begin{aligned} \|x_{n+1} - x_n\| &= \|\alpha_n f(\theta_n) + (1 - \alpha_n)T_n(\theta_n) + e(\theta_n) - \alpha_{n-1}f(\theta_{n-1}) \\ &\quad - (1 - \alpha_{n-1})T_{n-1}(\theta_{n-1}) - e(\theta_{n-1})\| \\ &\leq \|\alpha_n(f(\theta_n) - f(\theta_{n-1})) + (\alpha_n f(\theta_{n-1}) - \alpha_{n-1}f(\theta_{n-1}))\| \\ &\quad + \|(1 - \alpha_n)(T_n(\theta_n) - T_n(\theta_{n-1})) \\ &\quad + (1 - \alpha_n)T_n(\theta_{n-1}) - (1 - \alpha_{n-1})T_{n-1}(\theta_{n-1})\| \\ &\quad + \|e(\theta_n)\| + \|e(\theta_{n-1})\|. \end{aligned}$$

Since f is ρ -contractive, we obtain

$$\begin{aligned}
 \|x_{n+1} - x_n\| &\leq \rho\alpha_n\|\theta_n - \theta_{n-1}\| + |\alpha_n - \alpha_{n-1}| \|f(\theta_{n-1})\| \\
 &\quad + (1 - \alpha_n)\|T_n(\theta_n) - T_n(\theta_{n-1})\| + (1 - \alpha_n)\|T_n(\theta_{n-1}) - T_{n-1}(\theta_{n-1})\| \\
 &\quad + |\alpha_n - \alpha_{n-1}| \|T_{n-1}(\theta_{n-1})\| + \|e(\theta_n)\| + \|e(\theta_{n-1})\| \\
 &\leq \rho\alpha_n\|\theta_n - \theta_{n-1}\| + |\alpha_n - \alpha_{n-1}| \|f(\theta_{n-1})\| + (1 - \alpha_n)\|\theta_n - \theta_{n-1}\| \\
 &\quad + (1 - \alpha_n)\|T_n(\theta_{n-1}) - T_{n-1}(\theta_{n-1})\| + |\alpha_n - \alpha_{n-1}| \|T_{n-1}(\theta_{n-1})\| \\
 &\quad + \|e(\theta_n)\| + \|e(\theta_{n-1})\| \\
 &= (1 - \alpha_n(1 - \rho))\|\theta_n - \theta_{n-1}\| + |\alpha_n - \alpha_{n-1}| (\|f(\theta_{n-1})\| + \|T_{n-1}(\theta_{n-1})\|) \\
 &\quad + (1 - \alpha_n)\|T_n(\theta_{n-1}) - T_{n-1}(\theta_{n-1})\| + \|e(\theta_n)\| + \|e(\theta_{n-1})\|.
 \end{aligned} \tag{16}$$

Substituting (14) into (16), we obtain

$$\begin{aligned}
 \|x_{n+1} - x_n\| &\leq (1 - \alpha_n(1 - \rho))\|\theta_n - \theta_{n-1}\| \\
 &\quad + |\alpha_n - \alpha_{n-1}| (\|f(\theta_{n-1})\| + \|T_{n-1}(\theta_{n-1})\|) \\
 &\quad + (1 - \alpha_n) \frac{|\lambda_n - \lambda_{n-1}| \|\theta_{n-1} - T_{n-1}(\theta_{n-1})\|}{a} + \|e(\theta_n)\| + \|e(\theta_{n-1})\| \\
 &\leq (1 - \alpha_n(1 - \rho))\|\theta_n - \theta_{n-1}\| + \frac{|\lambda_n - \lambda_{n-1}| (\|\theta_{n-1}\| + \|T_{n-1}(\theta_{n-1})\|)}{a} \\
 &\quad - \alpha_n(1 - \rho) \frac{|\lambda_n - \lambda_{n-1}| (\|\theta_{n-1}\| + \|T_{n-1}(\theta_{n-1})\|)}{a(1 - \rho)} \\
 &\quad + |\alpha_n - \alpha_{n-1}| (\|f(\theta_{n-1})\| + \|T_{n-1}(\theta_{n-1})\|) + \|e(\theta_n)\| + \|e(\theta_{n-1})\|.
 \end{aligned} \tag{17}$$

Substituting (13) into (17), we obtain

$$\begin{aligned}
 \|x_{n+1} - x_n\| &\leq (1 - \alpha_n(1 - \rho))(\|x_n - x_{n-1}\| + (\alpha_n + \alpha_{n-1})M_1) \\
 &\quad - \alpha_n(1 - \rho) \frac{|\lambda_n - \lambda_{n-1}| (\|\theta_{n-1}\| + \|T_{n-1}(\theta_{n-1})\|)}{a(1 - \rho)} \\
 &\quad + M_2(|\alpha_n - \alpha_{n-1}| + |\lambda_n - \lambda_{n-1}|) + \|e(\theta_n)\| + \|e(\theta_{n-1})\|,
 \end{aligned}$$

where $M_2 = \sup_{n \in \mathbb{N}} \left\{ \|f(\theta_{n-1})\| + \|T_{n-1}\theta_{n-1}\|, \frac{\|\theta_{n-1}\| + \|T_{n-1}(\theta_{n-1})\|}{a} \right\}$. This implies

$$\begin{aligned}
 \|x_{n+1} - x_n\| &\leq (1 - \alpha_n(1 - \rho))\|x_n - x_{n-1}\| \\
 &\quad + (-1)\alpha_n(1 - \rho) \frac{|\lambda_n - \lambda_{n-1}| (\|\theta_{n-1}\| + \|T_{n-1}(\theta_{n-1})\|)}{a(1 - \rho)} \\
 &\quad + M_2(|\alpha_n - \alpha_{n-1}| + |\lambda_n - \lambda_{n-1}|) + (1 - \alpha_n(1 - \rho))(\alpha_n + \alpha_{n-1})M_1 \\
 &\quad + \|e(\theta_n)\| + \|e(\theta_{n-1})\|.
 \end{aligned} \tag{18}$$

Taking $\beta_n = M_2(|\alpha_n - \alpha_{n-1}| + |\lambda_n - \lambda_{n-1}|) + (1 - \alpha_n(1 - \rho))(\alpha_n + \alpha_{n-1})M_1 + \|e(\theta_n)\| + \|e(\theta_{n-1})\|$, $\gamma_n = \alpha_n(1 - \rho)$ and $\delta_n = \frac{(-1)|\lambda_n - \lambda_{n-1}|(\|\theta_{n-1}\| + \|T_{n-1}(\theta_{n-1})\|)}{a(1 - \rho)}$ in equation (18), we have that

$$\|x_{n+1} - x_n\| \leq (1 - \gamma_n)\|x_n - x_{n-1}\| + \gamma_n\delta_n + \beta_n. \tag{19}$$

Applying Lemma 5 on (19), we deduce that $\lim_{n \rightarrow \infty} \|x_{n+1} - x_n\| = 0$.

Step 4: We show that

$$\lim_{j \rightarrow \infty} \|T_{n_j}(\theta_{n_j}) - T(\theta_{n_j})\| = 0. \tag{20}$$

We apply Lemma 3 to obtain

$$\begin{aligned} & \|T_{n_j}\theta_{n_j} - T\theta_{n_j}\| \\ &= \|\mathbf{prox}_{\lambda_{n_j}\varphi}(\theta_{n_j} - \lambda_{n_j}\nabla\psi(\theta_{n_j})) - \mathbf{prox}_{\lambda\varphi}(\theta_{n_j} - \lambda\nabla\psi(\theta_{n_j}))\| \\ &= \|\mathbf{prox}_{\lambda_n\varphi}\left(\frac{\lambda}{\lambda_{n_j}}(\theta_{n_j} - \lambda_{n_j}\nabla\psi(\theta_{n_j})) + (1 - \frac{\lambda}{\lambda_{n_j}})\mathbf{prox}_{\lambda_{n_j}\varphi}(\theta_{n_j} - \lambda_{n_j}\nabla\psi(\theta_{n_j}))\right) \\ &\quad - \mathbf{prox}_{\lambda\varphi}(\theta_{n_j} - \lambda\nabla\psi(\theta_{n_j}))\| \\ &\leq \|\frac{\lambda}{\lambda_{n_j}}(\theta_{n_j} - \lambda_{n_j}\nabla\psi(\theta_{n_j})) \\ &\quad + (1 - \frac{\lambda}{\lambda_{n_j}})\mathbf{prox}_{\lambda_{n_j}\varphi}(\theta_{n_j} - \lambda_{n_j}\nabla\psi(\theta_{n_j})) - (\theta_{n_j} - \lambda\nabla\psi(\theta_{n_j}))\| \\ &= \|\left(\frac{\lambda}{\lambda_{n_j}} - 1\right)\theta_{n_j} + (1 - \frac{\lambda}{\lambda_{n_j}})\mathbf{prox}_{\lambda_{n_j}\varphi}(\theta_{n_j} - \lambda_{n_j}\nabla\psi(\theta_{n_j}))\| \\ &= |1 - \frac{\lambda}{\lambda_{n_j}}| \|\theta_{n_j} - T_{n_j}\theta_{n_j}\| \\ &\leq |1 - \frac{\lambda}{\lambda_{n_j}}| (\|\theta_{n_j}\| + \|T_{n_j}\theta_{n_j}\|). \end{aligned} \tag{21}$$

Letting $\lambda_{n_j} \rightarrow \lambda$ as $j \rightarrow \infty$ in (21), we derive (20).

Step 5: We show that

$$\lim_{j \rightarrow \infty} \|x_{n_j} - Tx_{n_j}\| = 0. \tag{22}$$

We compute

$$\begin{aligned} \|x_{n_j} - Tx_{n_j}\| &= \|x_{n_j} - \mathbf{prox}_{\lambda\varphi}(x_{n_j} - \lambda\nabla\psi(x_{n_j}))\| \\ &\leq \|x_{n_j} - x_{n_{j+1}}\| + \|x_{n_{j+1}} - \mathbf{prox}_{\lambda\varphi}(x_{n_j} - \lambda\nabla\psi(x_{n_j}))\| \\ &= \|x_{n_j} - x_{n_{j+1}}\| + \|\alpha_{n_j}f(\theta_{n_j}) - (1 - \alpha_{n_j})T_{n_j}\theta_{n_j} + e(\theta_{n_j}) \\ &\quad - \mathbf{prox}_{\lambda\varphi}(\theta_{n_j} - \lambda\nabla\psi(\theta_{n_j}))\| \\ &\leq \|x_{n_j} - x_{n_{j+1}}\| + \alpha_{n_j}\|f(\theta_{n_j}) - T(\theta_{n_j})\| + (1 - \alpha_{n_j})\|T_{n_j}(\theta_{n_j}) - T(\theta_{n_j})\| \\ &\quad + \|e(\theta_{n_j})\|. \end{aligned} \tag{23}$$

Since $\{x_n\}$ is bounded, then there exists a subsequence $\{x_{n_j}\}$ such that $x_{n_j} \rightarrow z$ as $j \rightarrow \infty$. That is, z is a cluster point of $\{x_n\}$. Taking limit on both sides of (23) using (16), (20) and the hypothesis on $\{\alpha_n\}$. We obtain (22). Since $x_{n_j} \rightarrow z$ as $j \rightarrow \infty$ and $(I - T)x_{n_j} \rightarrow 0$, then by Lemma 4, we have that $(I - T)z = 0$ which implies that $z \in \Omega$. Hence $\omega_W(x_n) \subset \Omega$.

Step 6: We show that

$$\limsup_{n \rightarrow \infty} \langle fx^* - x^*, \theta_n - x^* \rangle \leq 0. \tag{24}$$

To see this, let $\{\theta_{n_j}\}$ be arbitrary subsequence of $\{\theta_n\}$ such that

$$\limsup_{n \rightarrow \infty} \langle fx^* - x^*, \theta_n - x^* \rangle = \lim_{j \rightarrow \infty} \langle fx^* - x^*, \theta_{n_j} - x^* \rangle.$$

Since $\{\theta_{n_j}\}$ is bounded, it has a weakly convergent subsequence. Without loss of generality, choose $\{\theta_{n_j}\}$ has the subsequence and assume that $\theta_{n_j} \rightarrow u \in \Omega$. Then, by our hypothesis, we have

$$\langle fx^* - x^*, u - x^* \rangle \leq 0.$$

Hence,

$$\limsup_{n \rightarrow \infty} \langle fx^* - x^*, \theta_n - x^* \rangle = \lim_{j \rightarrow \infty} \langle fx^* - x^*, \theta_{n_j} - x^* \rangle \leq \langle fx^* - x^*, u - x^* \rangle \leq 0.$$

Step 7: We show that

$$\lim_{n \rightarrow \infty} x_n = x^*. \tag{25}$$

It follows from Algorithm 3.1, we compute

$$\begin{aligned} \|x_{n+1} - x^*\|^2 &= \|\alpha_n f(\theta_n) - (1 - \alpha_n)T_n \theta_n + e(\theta_n) - x^*\|^2 \\ &\leq \|\alpha_n(f(\theta_n) - x^*) + (1 - \alpha_n)(T_n \theta_n - x^*) + e(\theta_n)\|^2 \\ &\leq \|\alpha_n(f(\theta_n) - x^*) + (1 - \alpha_n)(T_n \theta_n - x^*)\|^2 + 2\langle \theta_{n+1} - x^*, +e(\theta_n) \rangle \\ &= \|\alpha_n(f(\theta_n) - x^*) + (1 - \alpha_n)(T_n \theta_n - x^*)\|^2 + 2\|\theta_{n+1} - x^*\| \|e(\theta_n)\| \\ &= \|\alpha_n(f(\theta_n) - fx^*) + (1 - \alpha_n)(T_n \theta_n - x^*) + \alpha_n(f(x^*) - x^*)\|^2 \\ &\quad + 2\|\theta_{n+1} - x^*\| \|e(\theta_n)\| \\ &\leq \|\alpha_n(f(\theta_n) - fx^*) + (1 - \alpha_n)(T_n \theta_n - x^*)\|^2 \\ &\quad + 2\langle \alpha_n(fx^* - x^*), \theta_{n+1} - e(\theta_n) - x^* \rangle \\ &\quad + 2\|\theta_{n+1} - x^*\| \|e(\theta_n)\|. \end{aligned}$$

Applying Lemma 1, we get

$$\begin{aligned}
 \|x_{n+1} - x^*\|^2 &\leq \alpha_n \|f(\theta_n) - fx^*\|^2 + (1 - \alpha_n) \|T_n \theta_n - x^*\|^2 \\
 &\quad + 2\langle \alpha_n (fx^* - x^*), \theta_{n+1} - e(\theta_n) - x^* \rangle + 2\|\theta_{n+1} - x^*\| \|e(\theta_n)\| \\
 &\leq \rho^2 \alpha_n \|\theta_n - x^*\|^2 + (1 - \alpha_n) \|\theta_n - x^*\|^2 \\
 &\quad + 2\alpha_n \langle fx^* - x^*, \theta_{n+1} - e(\theta_n) - x^* \rangle + 2\|\theta_{n+1} - x^*\| \|e(\theta_n)\| \\
 &= (1 - \alpha_n(1 - \rho^2)) \|\theta_n - x^*\|^2 + 2\alpha_n \langle fx^* - x^*, \theta_{n+1} - x^* \rangle \\
 &\quad + 2\alpha_n \langle fx^* - x^*, e(\theta_n) \rangle + 2\|\theta_{n+1} - x^*\| \|e(\theta_n)\| \\
 &\leq (1 - \alpha_n(1 - \rho^2)) \|\theta_n - x^*\|^2 + 2\alpha_n \langle fx^* - x^*, \theta_{n+1} - x^* \rangle \\
 &\quad + 2(\alpha_n \|fx^* - x^*\| + \|\theta_{n+1} - x^*\|) \|e(\theta_n)\| \\
 &\leq (1 - \alpha_n(1 - \rho^2)) \|\theta_n - x^*\|^2 + 2\alpha_n \langle fx^* - x^*, \theta_{n+1} - x^* \rangle \\
 &\quad + 2(\alpha_n \|fx^* - x^*\| + \|\theta_{n+1} - x^*\|) \|e(\theta_n)\|.
 \end{aligned}
 \tag{26}$$

Substituting (13) into (26) yields

$$\begin{aligned}
 \|x_{n+1} - x^*\|^2 &\leq (1 - \alpha_n(1 - \rho^2)) (\|x_n - x^*\| + \alpha_n M_1)^2 + 2\alpha_n \langle fx^* - x^*, \theta_{n+1} - x^* \rangle \\
 &\quad + 2(\alpha_n \|fx^* - x^*\| + \|\theta_{n+1} - x^*\|) \|e(\theta_n)\| \\
 &\leq (1 - \alpha_n(1 - \rho^2)) \|x_n - x^*\|^2 + 2\alpha_n \langle fx^* - x^*, \theta_{n+1} - x^* \rangle \\
 &\quad + M_3 \|e(\theta_n)\|,
 \end{aligned}$$

where $M_3 = \sup_n \{2(\alpha_n \|fx^* - x^*\| + \|\theta_{n+1} - x^*\|)\} < \infty$. Setting $\gamma_n = \alpha_n(1 - \rho^2)$, $\beta_n = M_3 \|e(\theta_n)\|$ and $\delta_n = \frac{2}{1 - \rho^2} \langle fx^* - x^*, \theta_{n+1} - x^* \rangle$. We have

$$\|x_{n+1} - x^*\|^2 \leq (1 - \gamma_n) \|x_n - x^*\|^2 + \gamma_n \delta_n + \beta_n.
 \tag{27}$$

Hence by Lemma 5, we get $\lim_{n \rightarrow \infty} \|x_n - x^*\| = 0$. This completes the proof. □

Theorem 3 *Under the same condition of Theorem 2. Let $\{\beta_n\}$ and $\{v_n\}$ satisfy the assumption in Definition 1. Then Algorithm 3.1 is bounded perturbation resilient. That is, the sequence generated by Algorithm 3.2 converges strongly to $x^* \in \Omega$.*

Proof Algorithm 3.2 can be rewritten as

$$\begin{cases} \theta_n = x_n + \zeta_n(x_n - x_{n-1}), \\ x_{n+1} = \alpha_n f(\theta_n) + (1 - \alpha_n) \mathbf{prox}_{\lambda_n \varphi}(\theta_n - \lambda_n \nabla \psi(\theta_n)) + e(\theta_n), \end{cases} \quad n \geq 0,$$

where

$$\begin{aligned}
 e(\theta_n) &= \alpha_n (f(\theta_n + \beta_n v_n) - f(\theta_n)) \\
 &\quad + (1 - \alpha_n) (\mathbf{prox}_{\lambda_n \varphi}(\theta_n + \beta_n v_n - \lambda_n \nabla \psi(\theta_n + \beta_n v_n)) - \mathbf{prox}_{\lambda_n \varphi}(\theta_n - \lambda_n \nabla \psi(\theta_n))).
 \end{aligned}$$

We check that $\sum_{n=0}^\infty \|e(\theta_n)\| < +\infty$. Applying Lemma 3, we get

$$\begin{aligned}
 \|e(\theta_n)\| &\leq \alpha_n \|f(\theta_n + \beta_n v_n) - f(\theta_n)\| \\
 &\quad + (1 - \alpha_n) \|\mathbf{prox}_{\lambda_n \varphi}(\theta_n + \beta_n v_n - \lambda_n \nabla \psi(\theta_n + \beta_n v_n)) \\
 &\quad - \mathbf{prox}_{\lambda_n \varphi}(\theta_n - \lambda_n \nabla \psi(\theta_n))\| \\
 &\leq \alpha_n \rho \|(\theta_n + \beta_n v_n) - (\theta_n)\| + (1 - \alpha_n) \|(\theta_n + \beta_n v_n - \lambda_n \nabla \psi(\theta_n + \beta_n v_n)) \\
 &\quad - (\theta_n - \lambda_n \nabla \psi(\theta_n))\| \\
 &\leq \alpha_n \|\beta_n v_n\| + (1 - \alpha_n) \|\beta_n v_n + \lambda_n (\nabla \psi(\theta_n) - \nabla \psi(\theta_n + \beta_n v_n))\| \\
 &\leq \alpha_n \|\beta_n v_n\| + (1 - \alpha_n) \|\beta_n v_n\| + (1 - \alpha_n) \lambda_n L \|\theta_n - (\theta_n + \beta_n v_n)\| \\
 &= (1 + (1 - \alpha_n) \lambda_n L) \|\beta_n v_n\|.
 \end{aligned}$$

This means that $\sum_{n=0}^{\infty} \|e(\theta_n)\| < +\infty$. Therefore, the conclusion follows from Theorem 2. □

Remark 1 Typically, the inertial scheme speeds up (boosts) the iterative sequence towards the solution set. It provides a major advantage over the execution time and the number of iterations for large-scale problems.

Remark 2 If $\zeta_n = 0$ in Algorithm 3.2, then we obtain the modified proximal gradient algorithm with superiorization method of Guo and Cui [16] as follows:

$$x_{n+1} = \alpha_n f(x_n + \beta_n v_n) + (1 - \alpha_n) \mathbf{prox}_{\lambda_n \varphi}(x_n + \beta_n v_n - \lambda_n \nabla \psi(x_n + \beta_n v_n)), \quad n \geq 0, \tag{28}$$

Remark 3 [28] The condition (B3) is easily satisfied in numerical computation because the valued of $\|x_n - x_{n-1}\|$ is known before choosing ζ_n . Indeed, the parameter ζ_n can be chosen such that $0 \leq \zeta_n \leq \bar{\zeta}_n$, where

$$\bar{\zeta}_n = \begin{cases} \min\{\frac{\tau_n}{\|x_n - x_{n-1}\|}, \tau\} & \text{if } x_n \neq x_{n-1}, \\ \tau, & \text{otherwise,} \end{cases}$$

where $\{\tau_n\}$ is a positive sequence such that $\tau_n = o(\alpha_n)$ and $\zeta_n \in [0, \tau] \subset [0, 1)$.

3.1 Application to image recovery problems

In this section, we illustrate the relevance of our theorem with a linear inverse problem called LASSO (Least Absolute Shrinkage and Selection Operator) in compressed sensing (see in [35–38]).

Let $A : H \rightarrow H$ be a bounded linear operator on a real Hilbert space. The image recovery model can be formulated as:

$$Ax = b + \omega \tag{29}$$

where ω is an unknown noise vector, x is an unknown image, and b is the image (degraded observation) measurement. Assume problem (29) is consistent, then it can be recast as a least-squares problem:

$$\min_{x \in H} \left\{ \frac{1}{2} \|Ax - b\|^2 + \mu \|x\| \right\}, \tag{30}$$

where $\mu > 0$ is a regularization parameter. We now apply our Theorem 3 to approximate the solutions of problem (30). Setting $\psi(x) = \frac{1}{2} \|Ax - b\|^2$, $\varphi(x) = \mu \|x\|$. We see that both φ and ψ are convex, lower semi continuous and proper functions with $\nabla\psi(x) = A^*(Ax - b)$ for each $x \in H$. Also $\nabla\psi$ is Lipschitz continuous with $L = \|A\|^2$. Indeed, the composite $\psi(x) + \varphi(x) := \Phi$ is coercive and the subdifferential of φ is

$$\partial\|\cdot\|(x) = \begin{cases} \frac{x}{\|x\|}, & x \neq 0, \\ B(0, 1), & x = 0. \end{cases}$$

While

$$\begin{aligned} \|\nabla\psi(x) - \nabla\psi(y)\| &= \|A^*(Ax - b) - A^*(Ay - b)\| \\ &= \|A^*A(x - y)\| \\ &\leq \|A\|^2 \|x - y\| = L\|x - y\|. \end{aligned}$$

The conclusion of our results in Theorem 3 thus applies with suitable standard assumption on the sequence parameters.

4 Numerical experiments

In this section, we present the numerical results of proposed algorithm and compare the performance with algorithm (27) of Guo and Cui [16] in signal recovery. For clarity sake, we reformulate problem (30) in $H = \mathbb{R}^N$ as follows:

$$\min_{x \in \mathbb{R}^N} \left\{ \frac{1}{2} \|Ax - b\|_2^2 + \mu \|x\|_1 \right\}, \tag{31}$$

where $A : \mathbb{R}^N \rightarrow \mathbb{R}^M$ ($M < N$) is a linear and bounded matrix operator and $b \in \mathbb{R}^M$ is the observed or measured data with a noise ω . The positive scalar μ called the regularization parameter and $x \in \mathbb{R}^N$ is a sparse vector with m nonzero components to be recovered. We claim that $\psi(x) = \frac{1}{2} \|Ax - b\|_2^2$ and $\varphi(x) = \mu \|x\|_1$ be an indicator function with proximal map as

$$\mathbf{prox}_{\mu\|\cdot\|_1}(x_n) = (\mathbf{prox}_{\mu|\cdot|_1}(x_n^1), \dots, \mathbf{prox}_{\mu|\cdot|_1}(x_n^k))^T = (s_n^1, \dots, s_n^k)^T,$$

where $s_n^k = \mathbf{sgn}(x_n^k) \cdot \max\{|x_n^k| - \mu, 0\}$ for $k = 1, 2, \dots, N$. It is worth noting that the subdifferential $\partial\varphi$ at x_n is

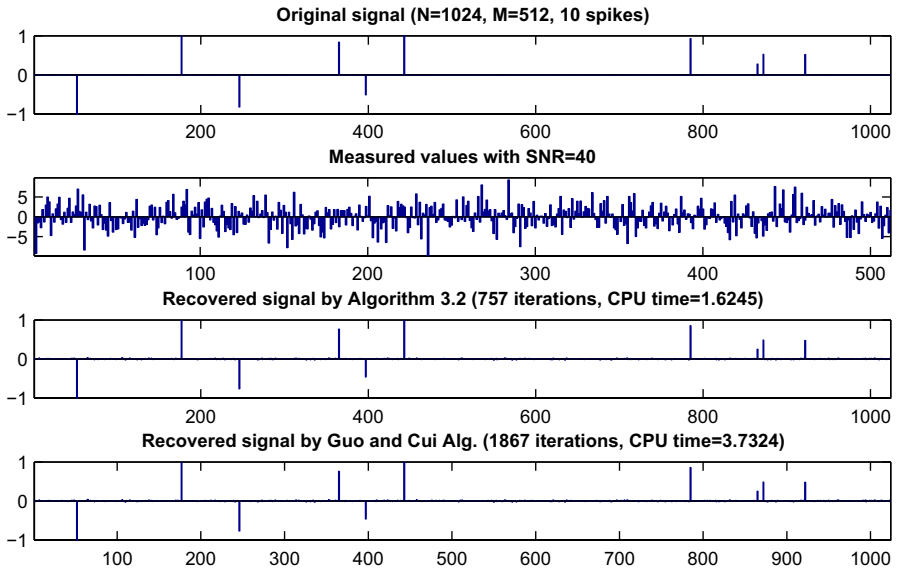


Fig. 1 The numerical experiments of Test 1

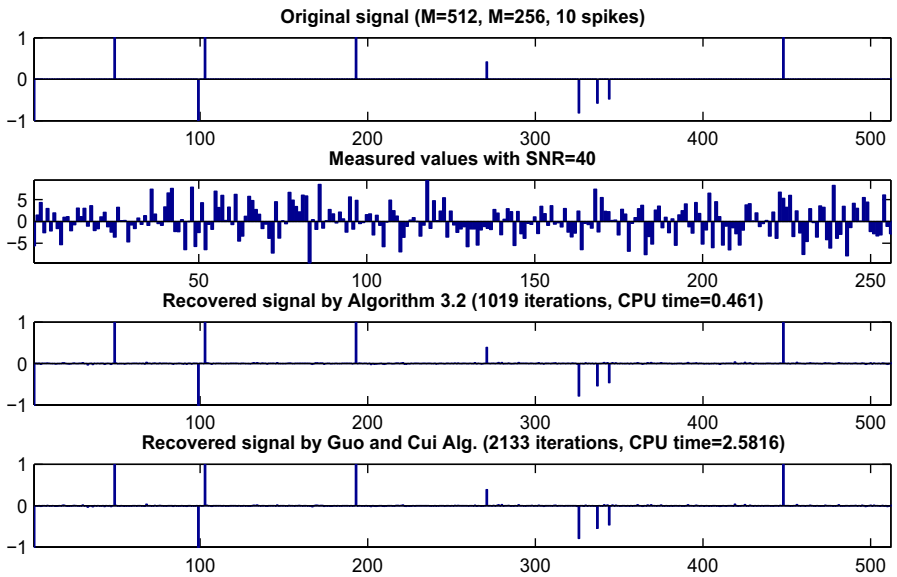


Fig. 2 The numerical experiments of Test 2

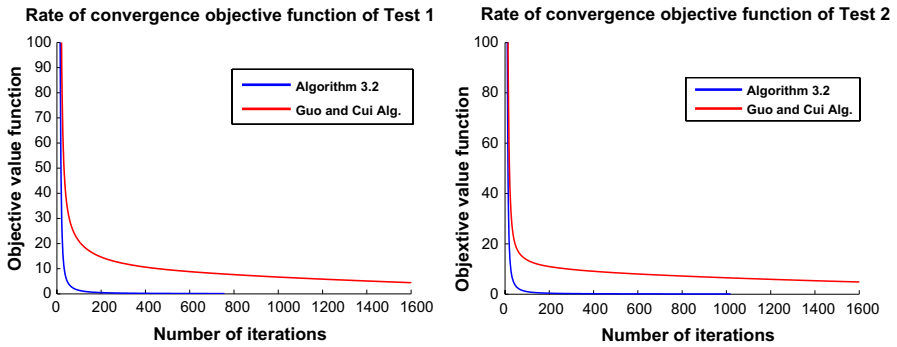


Fig. 3 The objective function value versus number of iterations of Test 1 and Test 2, respectively

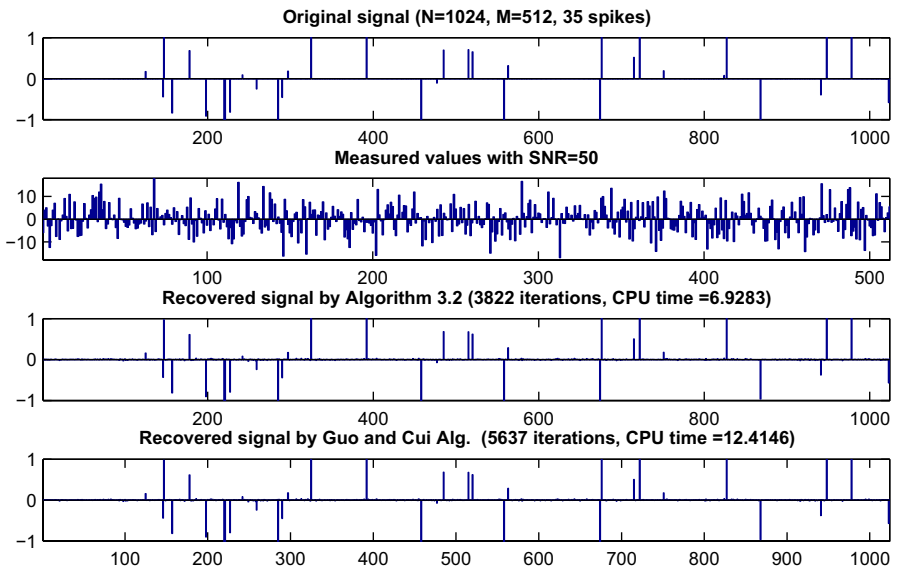


Fig. 4 The numerical experiments of Test 3

$$\partial\varphi(x_n) = \begin{cases} 1, & \text{if } x_n > 0, \\ [-1, 1], & \text{if } x_n = 0, \\ -1, & \text{if } x_n < 0. \end{cases}$$

Then the bounded sequence $\{v_n\}$ defined by

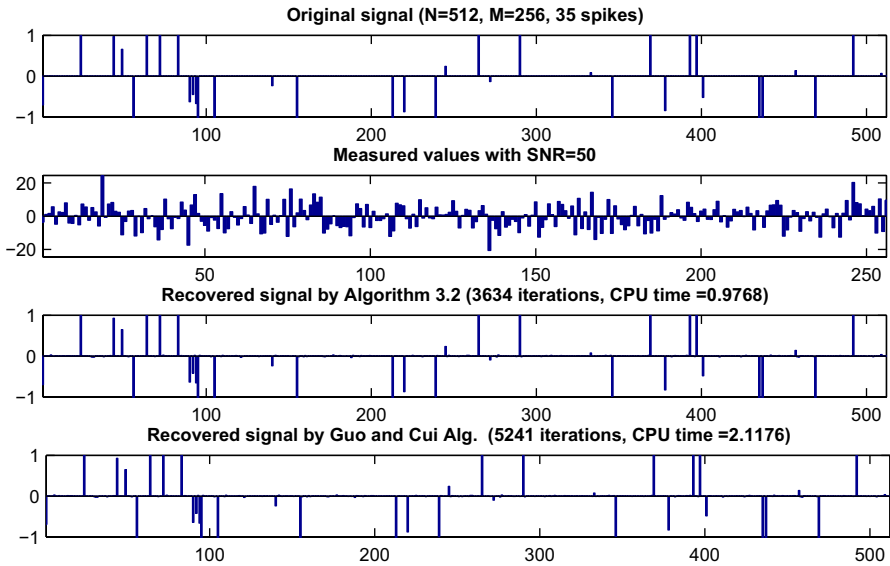


Fig. 5 The numerical experiments of Test 4

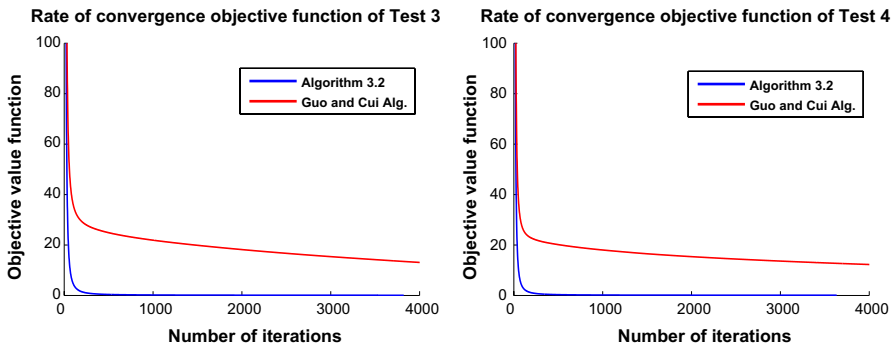


Fig. 6 The objective function value versus number of iterations of Test 3 and Test 4, respectively

$$v_n = \begin{cases} -\frac{d_n}{\|d_n\|}, & 0 \neq d_n \in \partial\varphi(x_n), \\ 0, & 0 \in \partial\varphi(x_n), \end{cases}$$

and the summarizable positive real sequence $\beta_n = \frac{1}{5n}$. We set $f(x) = \frac{x}{2}$ and control parameters $\alpha_n = \frac{1}{n+1}$, $\lambda_n = \frac{0.3}{\|A\|^2}$, $\mu = \frac{1}{L}$. Let $\tau = 0.5$, $\tau_n = \frac{1}{n^2}$ for each n and define ζ_n as in Remark 3

In experiment 1, the matrix $A^{M \times N}$ entries are sampled independently from a normal distribution of zero mean and unit variance. The observation vector b is generated from a Gaussian noise with signal-to-noise ratio SNR = 40. The sparse vector x is generated from a uniform distribution in the interval $[-2, 2]$ with m nonzero

elements. The initial point x_0, x_1 is picked randomly. The restoration accuracy is measured by the mean square error as $E_n = \frac{1}{N} \|x_n - x^*\| < 10^{-4}$, where x^* is the estimated signal of x . The numerical results are reported as follows:

Test 1: Set $N = 1024$, $M = 512$ and $m = 10$.

Test 2: Set $N = 512$, $M = 256$ and $m = 10$.

In experiment 2, the matrix $A^{M \times N}$ entries are sampled independently from a Gaussian distribution of zero mean and unit variance. The observation vector b is generated from a Gaussian noise with signal-to-noise ratio $\text{SNR} = 50$. The vector b is generated from a uniformly distribution in the interval $[-2, 2]$ with m nonzero elements. The restoration accuracy is measured by the mean square error as $E_n = \frac{1}{N} \|x_n - x^*\| < 10^{-4}$, where x^* is the estimated signal of x . The numerical results are reported as follows:

Test 3: Set $N = 1024$, $M = 512$ and $m = 35$.

Test 4: Set $N = 512$, $M = 256$ and $m = 35$.

From Figs. 1, 2, 4 and 5, it is guarantee that our algorithm can recover the original signal faster than Guo and Cui algorithm (28) in the compressed sensing. Moreover, Figs. 3 and 6 report that objective function values generated by Algorithm 3.2 decrease faster than Guo and Cui algorithm (28) in all of numerical tests.

5 Conclusions

In this article, we suggest an inertial viscosity proximal gradient algorithm which is proved effective to handle a non-smooth composite optimization problem with large datasets in real Hilbert spaces. We investigate its convergence and bounded perturbation resilience properties. Our results are demonstrated, and its performance is compared with Algorithm (28) of Guo and Cui [16] with respect to the signal recovery problem. The advantage of our proposed algorithm over some previous works is the reduction in the computational cost and the CPU time in computing numerical results. Numerical studies show that inertial effects usually improve the performance of the iterative sequence in terms of convergence in this context.

Acknowledgements This work was completed while the first author, Nuttapol Pakkaranang, was visiting North University Center at Baia Mare, Technical University of Cluj-Napoca on 24 February–24 May, 2019. He thanks Professor Vasile Berinde for their hospitality and support by ERAMUS+ project and also thanks the Thailand Research Fund (TRF) support through the Royal Golden Jubilee Ph.D. (RGJ-PHD) Program (Grant No. PHD/0205/2561). The second author, Poom Kumam, was supported by the Thailand Research Fund and the King Mongkut's University of Technology Thonburi (KMUTT) under the TRF Research Scholar (Grant No. RSA6080047). The fourth author, Yusuf I. Suleiman, thanks King Mongkut's University of Technology Thonburi (KMUTT) for visiting research 6 months (October 2018–March 2019). Furthermore, this project was supported by the Center of Excellence in Theoretical and Computational Science (TaCS-CoE), KMUTT.

References

1. Bauschke HH, Combettes PL (2011) Convex analysis and monotone operator theory in Hilbert spaces. Springer, New York
2. Engl HW, Hanke M, Neubauer A (1996) Regularization of inverse problems. Kluwer academic, Dordrecht
3. Bach F, Jenatton R, Mairal J, Obozinski G (2012) Optimization with sparsity-inducing penalties. *Found Trends Mach Learn* 4(1):1–106
4. Beck A, Teboulle M (2009) Gradient-based algorithms with applications to signal-recovery problems. In: Palomar D, Eldar Y (eds) Convex optimization in signal processing and communications. Cambridge University Press, Cambridge, pp 42–88. <https://doi.org/10.1017/CBO9780511804458.003>
5. Figueiredo M, Nowak R (2003) An EM algorithm for wavelet-based image restoration. *IEEE Trans Image Process* 12(8):906–916
6. Nikazad T, Davidi R, Herman GT (2012) Accelerated perturbation-resilient block-iterative projection methods with application to image reconstruction. *Inverse Probl* 28(3):035005. <https://doi.org/10.1088/0266-5611/28/3/035005>
7. Helou ES, Zibetti MVW, Miqueles EX (2017) Superiorization of incremental optimization algorithms for statistical tomographic image reconstruction. *Inverse Probl* 33(4):044010. <https://doi.org/10.1088/1361-6420/33/4/044010>
8. Davidi R, Herman GT, Censor Y (2009) Perturbation-resilient block-iterative projection methods with application to image reconstruction from projections. *Int Trans Oper Res* 16(4):505–524
9. Combettes PL, Pesquet JC (2011) Proximal splitting methods in signal processing. In: Bauschke HH, Burachik RS, Combettes PL, Elser V, Russell Luke D, Wolkowicz H (eds) Fixed-point algorithms for inverse problems in science and engineering, vol 49. Springer, New York, pp 185–212
10. Dunn JC (1976) Convexity, monotonicity, and gradient processes in Hilbert space. *J Math Anal Appl* 53(1):145–158
11. Wang C, Xiu N (2000) Convergence of the gradient projection method for generalized convex minimization. *Comput Optim Appl* 16(2):111–120
12. Güler O (1991) On the convergence of the proximal algorithm for convex minimization. *SIAM J Control Optim* 29(2):403–419
13. Parikh N, Boyd S (2014) Proximal algorithm. *Found Trends Optim* 1(3):123–231
14. Rockafellar RT (1976) Monotone operators and the proximal point algorithm. *SIAM J Control Optim* 14(5):877–898
15. Xu HK (2014) Properties and iterative methods for the lasso and its variants. *Chin Ann Math Ser B* 35(3):501–518
16. Guo Y, Cui W (2018) Strong convergence and bounded perturbation resilience of a modified proximal gradient algorithm. *J Inequal Appl*. <https://doi.org/10.1186/s13660-018-1695-x>
17. Moudafi A (2000) Viscosity approximation methods for fixed-points problems. *J Math Anal Appl* 241(1):46–55
18. Censor Y, Davidi R, Herman GT (2010) Perturbation resilience and superiorization of iterative algorithms. *Inverse Probl* 26(6):065008. <https://doi.org/10.1088/0266-5611/26/6/065008>
19. Davidi R, Censor Y, Schulte RW, Geneser S, Xing L (2015) Feasibility-seeking and superiorization algorithm applied to inverse treatment planning in radiation therapy. *Contemp Math* 636:83–92
20. Schrapp MJ, Herman GT (2014) Data fusion in X-ray computed tomography using a superiorization approach. *Rev Sci Instrum* 85:053701. <https://doi.org/10.1063/1.4872378>
21. Xu HK (2017) Bounded perturbation resilience and superiorization techniques for the projected scaled gradient method. *Inverse Probl* 33(4):044008. <https://doi.org/10.1088/1361-6420/33/4/044008>
22. He H, Xu HK (2017) Perturbation resilience and superiorization methodology of averaged mappings. *Inverse Probl* 33(4):044007. <https://doi.org/10.1088/1361-6420/33/4/044007>
23. Censor Y, Zaslavski AJ (2013) Convergence and perturbation resilience of dynamic string-averaging projection methods. *Comput Optim Appl* 54(1):65–74
24. Polyak BT (1964) Some methods of speeding up the convergence of iterative methods. *Ž Vyčisl Mat i Mat Fiz* 4:1–17
25. Nesterov Y (1983) A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Dokl. Akad. Nauk SSSR* 269:543–547

26. Dang Y, Sun J, Xu H (2017) Inertial accelerated algorithms for solving a split feasibility problem. *J Ind Manag Optim* 13(3):1383–1394
27. Dong QL, Yuan HB, Cho YJ, Rassias ThM (2016) Modified inertial Mann algorithm and inertial CQ-algorithm for nonexpansive mappings. *Optim Lett* 12(1):87–102
28. Suantai S, Pholasa N, Cholamjiak P (2018) The modified inertial relaxed CQ algorithm for solving the split feasibility problems. *J Ind Manag Optim* 14:1595–1615
29. Cholamjiak W, Cholamjiak P, Suantai S (2018) An inertial forward–backward splitting method for solving inclusion problems in Hilbert spaces. *J Fixed Point Theory Appl* 20(1):42. <https://doi.org/10.1007/s11784-018-0526-5>
30. Kankam K, Pholasa N, Cholamjiak P (2019) On convergence and complexity of the modified forward–backward method involving new line searches for convex minimization. *Math Methods Appl Sci* 42(5):1352–1362
31. Suantai S, Pholasa N, Cholamjiak P (2019) Relaxed CQ algorithms involving the inertial technique for multiple-sets split feasibility problems. *Rev R Acad Cienc Exactas Fis Nat Ser A Mat* 113(2):1081–1099
32. Kesornprom S, Cholamjiak P (2019) Proximal type algorithms involving linesearch and inertial technique for split variational inclusion problem in Hilbert spaces with applications. *Optimization* 68(12):2365–2391
33. Byrne C (2004) A unified treatment of some algorithms in signal processing and image construction. *Inverse Probl* 20:103–120
34. Combettes PL, Wajs VR (2005) Signal recovery by proximal forward–backward splitting. *Multi-scale Model Simul* 4(4):1168–1200
35. Abubakar AB, Kumam P, Awwal AM (2019) Global convergence via descent modified three-term conjugate gradient projection algorithm with applications to signal recovery. *Results Appl Math*. <https://doi.org/10.1016/j.rinam.2019.100069>
36. Abubakar AB, Kumam P, Awwal AM, Thounthong P (2019) A modified self-adaptive conjugate gradient method for solving convex constrained monotone nonlinear equations for signal recovery problems. *Mathematics* 7(8):693. <https://doi.org/10.3390/math7080693>
37. Padcharoen A, Kumam P, Chaipunya P, Shehu Y (2020) Convergence of inertial modified Krasnoselskii–Mann iteration with application to image recovery. *Thai J Math* 18(1):126–142
38. Cholamjiak P, Kankam K, Srinet P, Pholasa N (2020) A double forward–backward algorithm using linesearches for minimization problem. *Thai J Math* 18(1):63–76

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Nuttapol Pakkaranang^{1,2} · Poom Kumam^{1,2,3}  · Vasile Berinde⁴ · Yusuf I. Suleiman⁵

Nuttapol Pakkaranang
nuttapol.pakka@gmail.com

Vasile Berinde
vasile.berinde@gmail.com

Yusuf I. Suleiman
maaarbuy@hotmail.com

¹ Center of Excellence in Theoretical and Computational Science (TaCS-CoE), KMUTT-Fixed Point Theory and Applications Research Group (KMUTT-FPTA), Science Laboratory Building, King Mongkut's University of Technology Thonburi (KMUTT), 126 Pracha-Uthit Road, Bang Mod, Thrung Khru, Bangkok 10140, Thailand

- ² KMUTT Fixed Point Research Laboratory, Department of Mathematics, Room SCL 802 Fixed Point Laboratory, Science Laboratory Building, Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT), 126 Pracha-Uthit Road, Bang Mod, Thrung Khru, Bangkok 10140, Thailand
- ³ Department of Medical Research, China Medical University Hospital, China Medical University, Taichung 40402, Taiwan
- ⁴ Department of Mathematics and Computer Science, North University Center at Baia Mare, Technical University of Cluj-Napoca, Victoriei 76, 430072 Baia-Mare, Romania
- ⁵ Department of Mathematics, Kano University of Science and Technology, P.M.B. 3042, Wudil, Kano, Nigeria