

RESEARCH

Open Access



# A proximal gradient method with double inertial steps for minimization problems involving demicontractive mappings

Thanasak Mouktonglang<sup>1</sup>, Wipawinee Chaiwino<sup>1</sup> and Raweerote Suparatulorn<sup>1,2\*</sup>

\*Correspondence:

[raweerote.s@gmail.com](mailto:raweerote.s@gmail.com)

<sup>1</sup>Department of Mathematics,  
Faculty of Science, Chiang Mai  
University, Chiang Mai 50200,  
Thailand

<sup>2</sup>Office of Research Administration,  
Chiang Mai University, Chiang Mai  
50200, Thailand

## Abstract

In this article, we present a novel proximal gradient method based on double inertial steps for solving fixed points of demicontractive mapping and minimization problems. We also establish a weak convergence theorem by applying this method. Additionally, we provide a numerical example related to a signal recovery problem.

**Mathematics Subject Classification:** 47H10; 65K05; 90C25

**Keywords:** Proximal gradient method; Mann iteration; Signal recovery; Process innovation

## 1 Introduction

Optimization and fixed point problems epitomize fundamental mathematical concepts with wide-ranging applications across various fields, including engineering, medicine, signal processing, and image processing. Engineers routinely grapple with the imperative to minimize expenditures, optimize designs, or maximize system efficiency, all of which can be aptly framed as optimization conundrums. In parallel, fixed point theorems assume profound significance in addressing particular engineering challenges, providing a robust mathematical framework for establishing the existence of solutions in diverse scenarios. Signal processing substantially reaps benefits from the incorporation of fixed point problems, particularly within the purview of optimization methodologies. These methodologies establish a resilient framework for effectively navigating the intricate complexities associated with denoising and restoration tasks. Notably, the least absolute shrinkage and selection operator (LASSO) [1] emerges as a pivotal optimization problem, assuming a critical role in the domain of signal reconstruction. Acknowledged for its remarkable efficacy within the compressed sensing paradigm, LASSO has enjoyed widespread recognition within the official discourse of signal processing. Within the realm of image processing, the application of optimization techniques and fixed point problem-solving takes on paramount importance, proving invaluable in the effective resolution of challenges such as image deblurring and image inpainting (refer to [2–7] for comprehensive information).

In 2014 (Jaggi [8]), an equivalence between the LASSO and support vector machines (SVMs) was shown in the following sense. Given any  $L_2$  norm loss function SVMs, a cor-

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

responding LASSO formulation has the same optimal solutions and vice versa. As a result, one can be re-translated into the other. From further investigation, the sparsity of a LASSO solution is equal to the number of support vectors for the corresponding SVMs. Many useful properties and sublinear time algorithms for SVMs naturally arise from LASSO properties. SVMs are commonly used for classification and regression tasks and have an extensive list of applications in natural language processing (NLP), particularly in the fields of information extraction and email phishing detection. SVMs are highly effective in information extraction tasks, such as named entity recognition, text categorization, and relation extraction. They are able to identify entities and patterns in unstructured text, as demonstrated in [9]. SVMs in email phishing detection [10, 11] utilize features such as sender addresses and content to perform binary classification, distinguishing between valid and questionable emails. SVMs can effectively detect anomalies in email traffic, but their success depends on the quality of features, data representation, and the training dataset. Utilizing ensemble approaches, which involve mixing SVMs with other models, improves the effectiveness of phishing detection. This emphasizes the significance of regularly updating the system to effectively respond to shifting phishing strategies. Moreover, in 2021, Afrin et al. [12] employed SVMs in conjunction with LASSO feature selection techniques to predict liver disease. More recently, Cholamjiak and Das [13] developed a modified projective forward-backward splitting algorithm for multiple models, including the LASSO, aimed at the prediction of Parkinson's disease through the application of the extreme learning machine.

Prominent optimization techniques for the minimization of the sum of a smooth function and a nonsmooth function encompass the proximal gradient algorithm, credited to its originator [14] (also referenced in [15]). This method involves the sequential application of gradient steps to the first function, followed by the proximity operator applied to the second function. It is widely recognized that the inclusion of inertia, alternatively referred to as Nesterov's acceleration [16], has the capacity to notably enhance both the theoretical and practical convergence rates of this approach. The recent surge in popularity of Nesterov's acceleration [16] has spurred the development of numerous variations, such as those detailed in references [17–19]. Particularly noteworthy is the fast iterative shrinkage-thresholding algorithm (FISTA), as introduced by Beck and Teboulle [17], which exhibits a significantly enhanced convergence rate akin to Nesterov's optimal gradient approach, specifically tailored for convex composite objective functions.

Throughout this article, denote by  $\mathcal{H}$  a real Hilbert space with the inner product  $\langle \cdot, \cdot \rangle$  and the associated norm  $\| \cdot \|$ . Let  $\mathbb{R}$  and  $\mathbb{N}$  be the sets of real numbers and nonnegative integers, respectively. We are interested in the following minimization problem:

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

where  $f : \mathcal{H} \rightarrow \mathbb{R}$  and  $g : \mathcal{H} \rightarrow (-\infty, +\infty]$  belong to the class of proper, lower semi-continuous (l.s.c.), and convex functions on  $\mathcal{H}$ . Furthermore, the function  $f$  is assumed to be differentiable with  $L$ -Lipschitz continuous gradient  $\nabla f$ . The set of minimizers of  $f + g$  is denoted by  $\arg \min(f + g)$ . It is well known that

$$\tilde{x} \in \arg \min(f + g) \iff 0 \in (\nabla f + \partial g)(\tilde{x}),$$

where  $\partial g$  is the subdifferential of  $g$ . Recently, Kesornprom and Cholamjiak [20] introduced a new proximal gradient method that integrates the inertial technique alongside adaptive step size, demonstrating its effectiveness in addressing the minimization problem defined in equation (1.1). This algorithm has been applied to solve X-ray image deblurring. Similarly, Kankam and Cholamjiak [21] investigated image restoration as a mathematical model using the minimization problem (1.1).

Next, we consider the following fixed point problem:

$$\text{find } x \in \mathcal{H} \text{ such that } x = Tx, \tag{1.2}$$

where  $T : \mathcal{H} \rightarrow \mathcal{H}$  is a mapping. We denote by  $Fix(T)$  the fixed point set of  $T$ . The Mann iteration [22] is prominent among the frequently employed algorithms for solving the fixed point problem described in equation (1.2). In 2008, Maingé [23] introduced an algorithm that cleverly integrates the inertial technique with Mann iteration, customizing it to address the fixed point problem (1.2). It is noteworthy that, under certain conditions, the iterative sequence generated by this algorithm weakly converges to a fixed point of a non-expansive mapping. The general inertial Mann iteration for a nonexpansive mapping was introduced by Dong et al. in 2018 [24]. It is evident that the method in [23] is a specific instance of this general inertial Mann iteration. According to [24], the sequence obtained by the general inertial Mann iteration weakly converges to a fixed point under certain suitable circumstances.

Drawing upon the insights garnered from preceding research, this paper proposes a novel proximal gradient method that incorporates the general inertial Mann iteration to obtain a weak convergence theorem for solving both the minimization problem (1.1) and the fixed point problem (1.2) associated with a demicontractive mapping, subject to specified control conditions. Furthermore, the efficacy of our proposed algorithm is demonstrated by its application to a signal recovery problem, underscoring its practical utility.

## 2 Preliminaries

To establish our primary result, this section provides necessary definitions and lemmas. We use the symbol  $\rightarrow$  to represent strong convergence and denote weak convergence as  $\rightharpoonup$ . Let  $s, t \in \mathcal{H}$  and  $\eta \in \mathbb{R}$ . Then we have

$$\|s + t\|^2 = \|s\|^2 + 2\langle s, t \rangle + \|t\|^2, \tag{2.1}$$

$$\|s + t\|^2 \leq \|s\|^2 + 2\langle t, s + t \rangle, \tag{2.2}$$

and

$$\|\eta s + (1 - \eta)t\|^2 = \eta\|s\|^2 + (1 - \eta)\|t\|^2 - \eta(1 - \eta)\|s - t\|^2. \tag{2.3}$$

**Definition 2.1** Let  $h : \mathcal{H} \rightarrow (-\infty, +\infty]$  be proper, convex, and l.s.c. function and  $\tilde{c} > 0$ . The proximity operator of  $h$  of order  $\tilde{c}$  is defined by

$$\text{prox}_{\tilde{c}h}(s) := \arg \min_{t \in \mathcal{H}} \left\{ h(t) + \frac{1}{2\tilde{c}} \|s - t\|^2 \right\}$$

for all  $s \in \mathcal{H}$ .

Next, let  $T : \mathcal{H} \rightarrow \mathcal{H}$  be a mapping and  $G : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  be a multivalued mapping.

**Definition 2.2**  $T$  is said to be

- (i)  $\mu$ -demicontractive if  $Fix(T) \neq \emptyset$  and there is  $\mu \in [0, 1)$  such that for all  $s \in \mathcal{H}$  and all  $p \in Fix(T)$ ,

$$\|Ts - p\|^2 \leq \|s - p\|^2 + \mu \|s - Ts\|^2,$$

- (ii)  $L$ -Lipschitz continuous if there is  $L > 0$  such that

$$\|Ts - Tt\| \leq L \|s - t\|$$

for all  $s, t \in \mathcal{H}$ .

**Definition 2.3**  $G$  is said to be

- (i) monotone if for all  $(s, u), (t, v) \in graph(G)$  (the graph of mapping  $G$ ),

$$\langle u - v, s - t \rangle \geq 0,$$

- (ii) maximal monotone if for every  $(s, u) \in \mathcal{H} \times \mathcal{H}$ ,

$$(s, u) \in graph(G) \iff \langle u - v, s - t \rangle \geq 0 \text{ for all } (t, v) \in graph(G).$$

**Definition 2.4** [25] Suppose  $Fix(T) \neq \emptyset$ . Then  $I - T$  is demiclosed at zero if for any  $\{s_n\} \in \mathcal{H}$ , the following implication holds:

$$s_n \rightharpoonup s^* \text{ and } s_n - Ts_n \rightarrow 0 \implies s^* \in Fix(T).$$

**Lemma 2.5** [26] *If  $T$  is a Lipschitz continuous and monotone mapping and  $G$  is a maximal monotone mapping, then the mapping  $T + G$  is maximal monotone.*

**Lemma 2.6** [27] *Let  $\{x_n\}$  and  $\{\Lambda_n\}$  be nonnegative sequences of real numbers satisfying  $x_{n+1} \leq (1 + \Lambda_n)x_n + \Lambda_n x_{n-1}$ . Then  $x_{n+1} \leq K \cdot \prod_{j=1}^n (1 + 2\Lambda_j)$ , where  $K = \max\{x_1, x_2\}$ . Furthermore, if  $\sum_{n=1}^\infty \Lambda_n < \infty$ , then  $\{x_n\}$  is bounded.*

**Lemma 2.7** [28] *Let  $\{x_n\}$  and  $\{y_n\}$  be sequences of nonnegative real numbers such that  $\sum_{n=1}^\infty y_n < \infty$  and  $x_{n+1} \leq x_n + y_n$ . Then  $\{x_n\}$  is a convergent sequence.*

**Lemma 2.8** [29, Opial] *Let  $\{s_n\}$  be a sequence in  $\mathcal{H}$  and  $\Psi$  be a nonempty subset of  $\mathcal{H}$ . If, for every  $s^* \in \Psi$ ,  $\{\|s_n - s^*\|\}$  converges and every weak sequential cluster point of  $\{s_n\}$  belongs to  $\Psi$ , then  $\{s_n\}$  converges weakly to a point in  $\Psi$ .*

### 3 Main result

We first assume that the following conditions are satisfied for the convergence analysis of our algorithm:

*Condition 1.*  $f : \mathcal{H} \rightarrow \mathbb{R}$  and  $g : \mathcal{H} \rightarrow (-\infty, +\infty]$  are two proper, l.s.c., and convex functions.

---

**Algorithm 1** Modified Proximal Gradient Algorithm

---

*Initialization:* Select arbitrary elements  $s_0, s_1 \in \mathcal{H}$ . Let  $\lambda \in (0, 1)$ ,  $\tau_1 \in (0, \infty)$ ,  $\{\eta_n\} \subset (0, 1)$ ,  $\{\theta_n\} \subset [0, \infty)$ ,  $\{\zeta_n\} \subset [0, \infty)$ ,  $\{p_n\} \subset [0, \infty)$ , and  $\{q_n\} \subset [1, \infty)$ .

*Iterative Steps:* Construct  $\{s_n\}$  by using the following steps:

*Step 1.* Compute

$$\begin{aligned} z_n &= s_n + \theta_n(s_n - s_{n-1}), \\ w_n &= z_n + \zeta_n(z_n - s_{n-1}), \\ y_n &= \text{prox}_{\tau_n g}(I - \tau_n \nabla f)w_n, \\ u_n &= y_n + \tau_n(\nabla f(w_n) - \nabla f(y_n)), \end{aligned}$$

and

$$s_{n+1} = (1 - \eta_n)u_n + \eta_n T u_n.$$

If  $w_n = y_n = u_n = T u_n$ , then stop and  $w_n \in \Psi$ . Otherwise, go to *Step 2*.

*Step 2.* Update

$$\tau_{n+1} = \begin{cases} \min\left\{\frac{\lambda q_n \|w_n - y_n\|}{\|\nabla f(w_n) - \nabla f(y_n)\|}, \tau_n + p_n\right\} & \text{if } \nabla f(w_n) \neq \nabla f(y_n); \\ \tau_n + p_n & \text{otherwise.} \end{cases}$$

Replace  $n$  with  $n + 1$  and then repeat *Step 1*.

---

*Condition 2.*  $f$  is differentiable and has an  $L$ -Lipschitz continuous gradient  $\nabla f$ .

*Condition 3.*  $T : \mathcal{H} \rightarrow \mathcal{H}$  is a  $\mu$ -demicontractive mapping such that  $I - T$  is demiclosed at zero.

*Condition 4.*  $\Psi := \arg \min(f + g) \cap \text{Fix}(T)$  is nonempty.

*Remark 3.1* It is known from [29] that  $\tilde{x} \in \arg \min(f + g)$  if and only if  $\tilde{x} = \text{prox}_{\tilde{c}g}(I - \tilde{c}\nabla f)\tilde{x}$ , where  $\tilde{c} > 0$ . If  $w_n = y_n = u_n = T u_n$  in Algorithm 1, then  $w_n \in \Psi$ .

We are now prepared for the main convergence theorem.

**Theorem 3.2** *Let  $\{s_n\}$  be generated by Algorithm 1. Assume that the following conditions hold:*

(C1)  $\sum_{n=1}^{\infty} p_n < \infty$ ; (C2)  $\lim_{n \rightarrow \infty} q_n = 1$ ; (C3)  $\bar{\eta} < \eta_n < 1 - \mu$  for some  $\bar{\eta} > 0$ ;

(C4)  $\sum_{n=1}^{\infty} \theta_n < \infty$ ; (C5)  $\sum_{n=1}^{\infty} \zeta_n < \infty$ .

*Then  $\{s_n\}$  converges weakly to a solution of  $\Psi$ .*

*Proof* Let  $\tilde{s} \in \Psi$ . Next, we prove all the following claims.

*Claim 1.*  $\lim_{n \rightarrow \infty} \varrho_n = \lambda$ , where  $\varrho_n = \frac{\lambda q_n \tau_n}{\tau_{n+1}}$ .

Since  $\nabla f$  is  $L$ -Lipschitz continuous mapping, if  $\nabla f(w_n) \neq \nabla f(y_n)$ , then

$$\frac{\lambda q_n \|w_n - y_n\|}{\|\nabla f(w_n) - \nabla f(y_n)\|} \geq \frac{\lambda q_n \|w_n - y_n\|}{L \|w_n - y_n\|} = \frac{\lambda q_n}{L} \geq \frac{\lambda}{L}.$$

By using the same technique as in the proof of [30, Lemma 3.1], we obtain

$$\lim_{n \rightarrow \infty} \tau_n = \tau \in \left[ \min \left\{ \tau_1, \frac{\lambda}{L} \right\}, \tau_1 + p \right],$$

where  $p = \sum_{n=1}^{\infty} p_n$ . It follows from (C2) that

$$\lim_{n \rightarrow \infty} \varrho_n = \lim_{n \rightarrow \infty} \frac{\lambda q_n \tau_n}{\tau_{n+1}} = \lambda.$$

*Claim 2.* For any  $n \in \mathbb{N}$ ,

$$\langle y_n - \tilde{s}, y_n - w_n + \tau_n (\nabla f(w_n) - \nabla f(y_n)) \rangle \leq 0.$$

By using the definition of  $y_n$ , we have

$$(I - \tau_n \nabla f)w_n \in (I + \tau_n \partial g)y_n.$$

Thus, we can write

$$c_n = \frac{w_n - y_n}{\tau_n} - \nabla f(w_n),$$

where  $c_n \in \partial g(y_n)$ . By Lemma 2.5, we have that the mapping  $\nabla f + \partial g$  is maximal monotone. This leads to

$$\langle y_n - \tilde{s}, \nabla f(y_n) + c_n \rangle \geq 0,$$

and thus

$$\langle y_n - \tilde{s}, y_n - w_n + \tau_n (\nabla f(w_n) - \nabla f(y_n)) \rangle \leq 0.$$

*Claim 3.* For any  $n \in \mathbb{N}$ ,

$$\|s_{n+1} - \tilde{s}\|^2 \leq \|w_n - \tilde{s}\|^2 - (1 - \varrho_n^2) \|w_n - y_n\|^2 - \eta_n (1 - \mu - \eta_n) \|Tu_n - u_n\|^2.$$

From (2.1), we have

$$\begin{aligned} \|u_n - \tilde{s}\|^2 &= \|y_n - \tilde{s} + \tau_n (\nabla f(w_n) - \nabla f(y_n))\|^2 \\ &= \|(y_n - w_n) + (w_n - \tilde{s})\|^2 + \tau_n^2 \|\nabla f(w_n) - \nabla f(y_n)\|^2 \\ &\quad + 2\tau_n \langle y_n - \tilde{s}, \nabla f(w_n) - \nabla f(y_n) \rangle \\ &= \|y_n - w_n\|^2 + \|w_n - \tilde{s}\|^2 + \tau_n^2 \|\nabla f(w_n) - \nabla f(y_n)\|^2 + 2\langle y_n - w_n, w_n - \tilde{s} \rangle \\ &\quad + 2\tau_n \langle y_n - \tilde{s}, \nabla f(w_n) - \nabla f(y_n) \rangle \\ &= \|w_n - \tilde{s}\|^2 - \|y_n - w_n\|^2 + \tau_n^2 \|\nabla f(w_n) - \nabla f(y_n)\|^2 + 2\langle y_n - w_n, y_n - \tilde{s} \rangle \\ &\quad + 2\tau_n \langle y_n - \tilde{s}, \nabla f(w_n) - \nabla f(y_n) \rangle \end{aligned}$$

$$\begin{aligned}
 &= \|w_n - \tilde{s}\|^2 - \|y_n - w_n\|^2 + \tau_n^2 \|\nabla f(w_n) - \nabla f(y_n)\|^2 \\
 &\quad + 2\langle y_n - \tilde{s}, y_n - w_n + \tau_n(\nabla f(w_n) - \nabla f(y_n)) \rangle.
 \end{aligned}$$

Using Claim 2, we get

$$\|u_n - \tilde{s}\|^2 \leq \|w_n - \tilde{s}\|^2 - \|y_n - w_n\|^2 + \tau_n^2 \|\nabla f(w_n) - \nabla f(y_n)\|^2. \tag{3.1}$$

By the definitions of  $u_n$  and  $\tau_n$ , we have

$$\|u_n - y_n\| = \tau_n \|\nabla f(w_n) - \nabla f(y_n)\| \leq \varrho_n \|w_n - y_n\|. \tag{3.2}$$

This together with (3.1) implies that

$$\|u_n - \tilde{s}\|^2 \leq \|w_n - \tilde{s}\|^2 - (1 - \varrho_n^2) \|w_n - y_n\|^2.$$

Applying this to (2.3) and the demicontractiveness of  $T$ , we derive

$$\begin{aligned}
 \|s_{n+1} - \tilde{s}\|^2 &= (1 - \eta_n) \|u_n - \tilde{s}\|^2 + \eta_n \|Tu_n - \tilde{s}\|^2 - \eta_n(1 - \eta_n) \|Tu_n - u_n\|^2 \\
 &\leq (1 - \eta_n) \|u_n - \tilde{s}\|^2 + \eta_n \|u_n - \tilde{s}\|^2 + \eta_n \mu \|Tu_n - u_n\|^2 \\
 &\quad - \eta_n(1 - \eta_n) \|Tu_n - u_n\|^2 \\
 &= \|u_n - \tilde{s}\|^2 - \eta_n(1 - \mu - \eta_n) \|Tu_n - u_n\|^2 \\
 &\leq \|w_n - \tilde{s}\|^2 - (1 - \varrho_n^2) \|w_n - y_n\|^2 - \eta_n(1 - \mu - \eta_n) \|Tu_n - u_n\|^2.
 \end{aligned}$$

*Claim 4.*  $\lim_{n \rightarrow \infty} \|s_n - \tilde{s}\|$  exists.

From Claim 1, we immediately get that  $\lim_{n \rightarrow \infty} (1 - \varrho_n^2) = 1 - \lambda^2 > 0$ , and so we can find  $n_0 \in \mathbb{N}$  such that  $1 - \varrho_n^2 > 0$  for all  $n \geq n_0$ . By the definitions of  $w_n, z_n$ , and using Claim 3, we have, for all  $n \geq n_0$ ,

$$\begin{aligned}
 \|s_{n+1} - \tilde{s}\| &\leq \|w_n - \tilde{s}\| \\
 &\leq \|z_n - \tilde{s}\| + \zeta_n \|z_n - s_{n-1}\| \\
 &\leq \|s_n - \tilde{s}\| + \theta_n \|s_n - s_{n-1}\| + \zeta_n(1 + \theta_n) \|s_n - s_{n-1}\| \\
 &\leq \|s_n - \tilde{s}\| + \Lambda_n \|s_n - s_{n-1}\| \tag{3.3}
 \end{aligned}$$

$$\leq (1 + \Lambda_n) \|s_n - \tilde{s}\| + \Lambda_n \|s_{n-1} - \tilde{s}\|, \tag{3.4}$$

where  $\Lambda_n = \theta_n + \zeta_n(1 + \theta_n)$ . By (C4) and (C5), we have  $\sum_{n=1}^{\infty} \Lambda_n < \infty$ , which together with Lemma 2.6 and (3.4) conclude that  $\{s_n\}$  is bounded. This implies that

$$\sum_{n=1}^{\infty} \Lambda_n \|s_n - s_{n-1}\| < \infty, \tag{3.5}$$

and so

$$\lim_{n \rightarrow \infty} \|s_n - w_n\| = \lim_{n \rightarrow \infty} \Lambda_n \|s_n - s_{n-1}\| = 0. \tag{3.6}$$

It follows that  $\{\|s_n - \tilde{s}\|\}$  is convergent because of using Lemma 2.7, (3.3), and (3.5).

*Claim 5.*  $\lim_{n \rightarrow \infty} \|s_n - u_n\| = 0$ .

Indeed, applying Claim 3 and (2.2), we have

$$\begin{aligned} \|s_{n+1} - \tilde{s}\|^2 &\leq \|z_n - \tilde{s}\|^2 + 2\zeta_n \langle z_n - s_{n-1}, w_n - \tilde{s} \rangle \\ &\quad - (1 - \varrho_n^2) \|w_n - y_n\|^2 - \eta_n(1 - \mu - \eta_n) \|Tu_n - u_n\|^2 \\ &\leq \|s_n - \tilde{s}\|^2 + 2\theta_n \langle s_n - s_{n-1}, z_n - \tilde{s} \rangle + 2\zeta_n(1 + \theta_n) \langle s_n - s_{n-1}, w_n - \tilde{s} \rangle \\ &\quad - (1 - \varrho_n^2) \|w_n - y_n\|^2 - \eta_n(1 - \mu - \eta_n) \|Tu_n - u_n\|^2 \\ &\leq \|s_n - \tilde{s}\|^2 + 2\theta_n \|s_n - s_{n-1}\| \|z_n - \tilde{s}\| + 2\zeta_n(1 + \theta_n) \|s_n - s_{n-1}\| \|w_n - \tilde{s}\| \\ &\quad - (1 - \varrho_n^2) \|w_n - y_n\|^2 - \eta_n(1 - \mu - \eta_n) \|Tu_n - u_n\|^2 \\ &\leq \|s_n - \tilde{s}\|^2 + 2M\Lambda_n \|s_n - s_{n-1}\| \\ &\quad - (1 - \varrho_n^2) \|w_n - y_n\|^2 - \eta_n(1 - \mu - \eta_n) \|Tu_n - u_n\|^2, \end{aligned}$$

where  $M := \sup_{n \in \mathbb{N}} \{\|z_n - \tilde{s}\|, \|w_n - \tilde{s}\|\}$ . It follows that

$$\begin{aligned} (1 - \varrho_n^2) \|w_n - y_n\|^2 &\leq \|s_n - \tilde{s}\|^2 - \|s_{n+1} - \tilde{s}\|^2 + 2M\Lambda_n \|s_n - s_{n-1}\| \\ &\quad - \eta_n(1 - \mu - \eta_n) \|Tu_n - u_n\|^2. \end{aligned}$$

From Claim 4, (C3), (3.6), and  $\lim_{n \rightarrow \infty} (1 - \varrho_n^2) > 0$ , we obtain

$$\lim_{n \rightarrow \infty} \|w_n - y_n\| = 0, \tag{3.7}$$

implying that

$$\lim_{n \rightarrow \infty} \|Tu_n - u_n\| = 0. \tag{3.8}$$

Using  $\lim_{n \rightarrow \infty} \varrho_n = \lambda$ , (3.2), and (3.7), we deduce

$$\lim_{n \rightarrow \infty} \|u_n - y_n\| = 0. \tag{3.9}$$

This implies in view of (3.6) and (3.7) that

$$\|s_n - u_n\| \leq \|s_n - w_n\| + \|w_n - y_n\| + \|y_n - u_n\| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

*Claim 6.* Every weak sequential cluster point of  $\{s_n\}$  belongs to  $\Psi$ .

Let  $s^*$  be a weak sequential cluster point of  $\{s_n\}$ , meaning that  $s_{n_k} \rightharpoonup s^*$  as  $k \rightarrow \infty$  for some subsequence  $\{s_{n_k}\}$  of  $\{s_n\}$ . This implies by Claim 5 that  $u_{n_k} \rightharpoonup s^*$  as  $k \rightarrow \infty$ . This together with (3.8), by the demiclosedness at zero of  $I - T$ ,  $s^* \in \text{Fix}(T)$ . Next, we show that  $s^* \in \arg \min(f + g)$ . Let  $(v, u) \in \text{graph}(\nabla f + \partial g)$ , that is,  $u - \nabla f(v) \in \partial g(v)$ . It implies by the definition of  $y_n$  that  $\frac{w_{n_k} - y_{n_k} - \tau_{n_k} \nabla f(w_{n_k})}{\tau_{n_k}} \in \partial g(y_{n_k})$ . By the maximal monotonicity of  $\partial g$ , we have

$$\left\langle v - y_{n_k}, u - \nabla f(v) - \frac{w_{n_k} - y_{n_k} - \tau_{n_k} \nabla f(w_{n_k})}{\tau_{n_k}} \right\rangle \geq 0.$$



Thus, by the monotonicity of  $\nabla f$ , we get

$$\begin{aligned} \langle v - y_{n_k}, u \rangle &\geq \left\langle v - y_{n_k}, \nabla f(v) + \frac{w_{n_k} - y_{n_k} - \tau_{n_k} \nabla f(w_{n_k})}{\tau_{n_k}} \right\rangle \\ &= \langle v - y_{n_k}, \nabla f(v) - \nabla f(y_{n_k}) \rangle + \langle v - y_{n_k}, \nabla f(y_{n_k}) - \nabla f(w_{n_k}) \rangle \\ &\quad + \left\langle v - y_{n_k}, \frac{w_{n_k} - y_{n_k}}{\tau_{n_k}} \right\rangle \\ &\geq \langle v - y_{n_k}, \nabla f(y_{n_k}) - \nabla f(w_{n_k}) \rangle + \frac{1}{\tau_{n_k}} \langle v - y_{n_k}, w_{n_k} - y_{n_k} \rangle. \end{aligned}$$

It follows from the Lipschitz continuity of  $\nabla f$ ,  $\lim_{k \rightarrow \infty} \frac{1}{\tau_{n_k}} > 0$ , (3.7), and (3.9) that

$$\langle v - s^*, u \rangle = \lim_{k \rightarrow \infty} \langle v - y_{n_k}, u \rangle \geq 0,$$

from which, together with the maximal monotonicity of  $\nabla f + \partial g$ , we get that  $s^* \in \arg \min(f + g)$ . Therefore,  $s^* \in \Psi$ . Finally, by Lemma 2.8, we can conclude that  $\{s_n\}$  converges weakly to a solution of  $\Psi$ . □

#### 4 Signal recovery problem

We consider the signal recovery problem using the linear equation shown below:

$$b = Ax^* + \varepsilon,$$

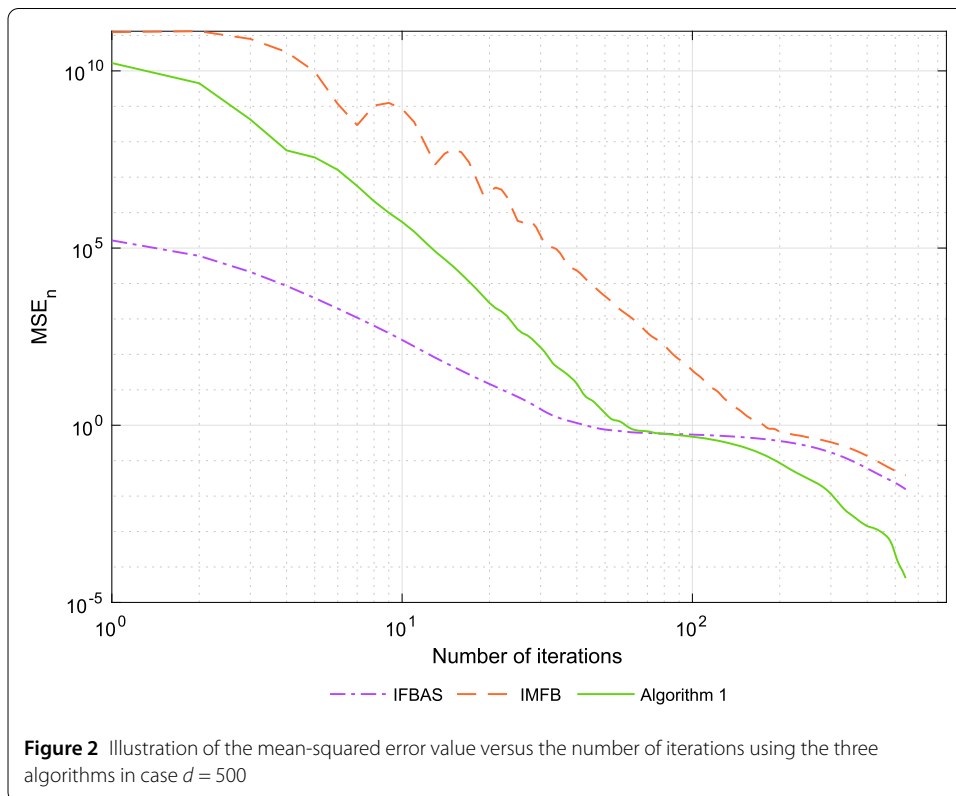
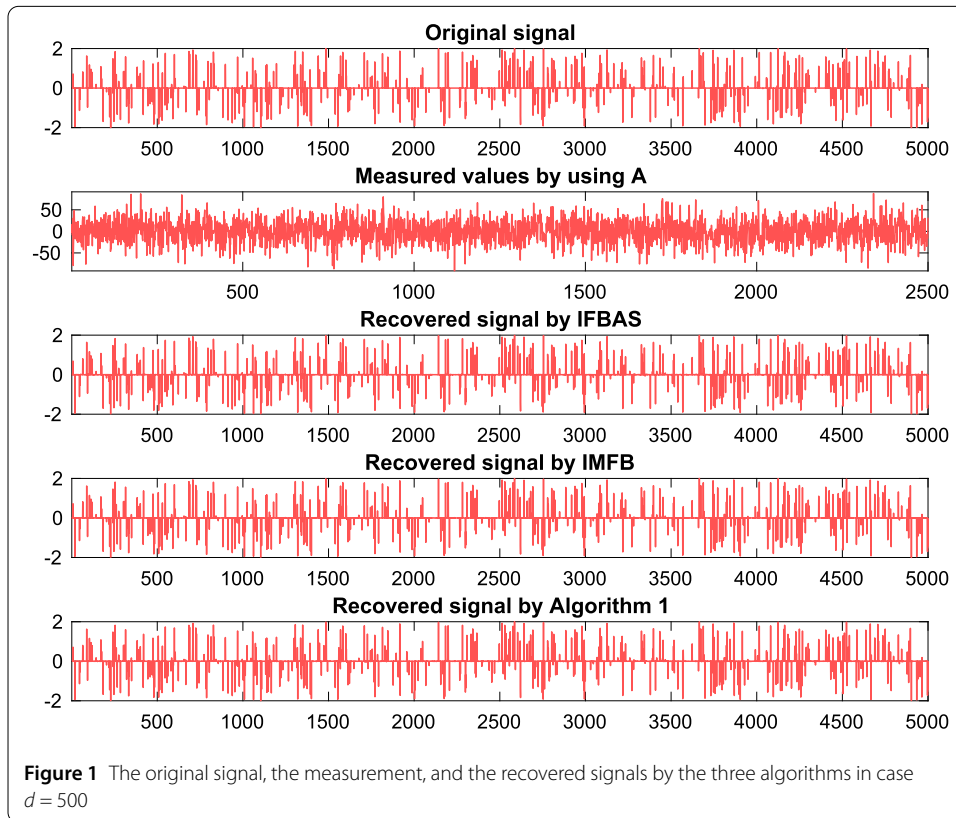
where the original signal is  $x^* \in \mathbb{R}^N$ ,  $b \in \mathbb{R}^M$  is the observed signal with noise  $\varepsilon$ , and  $A \in \mathbb{R}^{M \times N}$  ( $M < N$ ) is a filter matrix. It is generally known that resolving this linear equation is equivalent to determining the LASSO problem:

$$\min_{x \in \mathbb{R}^N} \frac{1}{2} \|Ax - b\|_2^2 + \|x\|_1.$$

So we can apply our algorithm to solve this problem in case  $f(\cdot) = \frac{1}{2} \|A(\cdot) - b\|_2^2$ ,  $g(\cdot) = \|\cdot\|_1$  and  $T = \text{prox}_{\tilde{c}g}(I - \tilde{c}\nabla f)$ , where  $0 < \tilde{c} < \frac{2}{\|A\|_2^2}$ . We present the numerical comparison of Algorithm 1 with Algorithm 3.1 in [20] (IMFB) and Algorithm 2.1 in [21] (IFBAS). All calculations are performed in Matlab R2021a on an iMac (Apple M1 chip with 16GB of RAM). Select the original signal  $x^*$  generated by the uniform distribution in  $[-2, 2]$  with  $d$  nonzero components, and  $A$  is the Gaussian matrix generated by command `randn(M, N)`, where the signal size is set to be  $N = 5000$  and  $M = 2500$ . The observation  $b$  is generated via the addition of white Gaussian noise  $\varepsilon$  with variance  $\sigma^2 = 0.01$  and the initial points being randomly generated. Let  $t_0 = 1$  and  $t_n = \frac{1 + \sqrt{1 + 4t_{n-1}^2}}{2}$  for all  $n \in \mathbb{N}$ . The control parameters of each algorithm are defined in the following manner:

(i) IFBAS:  $\alpha_1 = 0.09$ ,  $\delta = 0.6$ , and

$$\theta_n = \begin{cases} \frac{1}{n^2} & \text{if } n > 1500; \\ \frac{t_{n-1}-1}{t_n} & \text{otherwise;} \end{cases}$$



**Table 1** The numerical comparison of the three algorithms

$d$	Number of iterations			CPU time (s)		
	IFBAS	IMFB	Algorithm 1	IFBAS	IMFB	Algorithm 1
100	1053	1320	488	8.3065	10.4608	5.7580
180	1091	1381	501	8.6204	10.9159	5.8884
260	1093	1393	521	8.7759	11.0372	6.0982
340	1117	1399	531	8.8687	11.0945	6.2135
420	1158	1413	537	9.1763	11.1996	6.3128
500	1197	1456	543	9.7094	11.5568	6.3541

(ii) IMFB:  $\lambda_1 = 0.09$ ,  $\delta = 0.6$ , and

$$\theta_n = \begin{cases} \frac{1}{n^2} & \text{if } n > 1500; \\ \frac{t_{n-1}-1}{t_n} & \text{otherwise;} \end{cases}$$

(iii) Algorithm 1:  $\tilde{c} = \frac{1}{\|A\|_2^2}$ ,  $\tau_1 = 0.09$ ,  $\lambda = 0.6$ ,  $\eta_n = 0.9$ ,  $q_n = 1 + \frac{1}{n+1}$ ,  $p_n = \zeta_n = \frac{1}{(5n+2)^2}$ , and

$$\theta_n = \begin{cases} \frac{1}{n^2} & \text{if } n > 1500; \\ \frac{t_{n-1}-1}{t_n} & \text{otherwise.} \end{cases}$$

We measure the accuracy of the signal recovery using the mean-squared error, which is defined as  $MSE_n = \frac{1}{N} \|v_n - x^*\|_2^2 < 5 \times 10^{-5}$ , where  $\{v_n\}$  is the sequence to be measured. The numerical results are illustrated next.

When  $d = 500$ , Fig. 1 illustrates the original signal, measurement, and signals recovered by each algorithm. Figure 2 displays the mean-squared error for the results obtained from all three algorithms in the same scenario. As shown in Table 1, our algorithm improves CPU time and reduces the number of iterations compared to IFBAS and IMFB. This indicates that the new algorithm outperforms the other two.

**Acknowledgements**

This research work was partially supported by the CMU Proactive Researcher, Chiang Mai University [grant number 818/2566] and the NSRF via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation [grant number B05F650018].

**Author contributions**

Conceptualization: R.S.; Writing-original draft: T.M., R.S.; Software: W.C. All authors have read and approved the final version of the manuscript.

**Funding**

This research work was partially supported by the CMU Proactive Researcher, Chiang Mai University [grant number 818/2566] and the NSRF via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation [grant number B05F650018].

**Data Availability**

No datasets were generated or analysed during the current study.

**Declarations**

**Competing interests**

The authors declare no competing interests.

## References

1. Tibshirani, R.: Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* **58**, 267–288 (1996)
2. Deep, A., Abbas, S., Singh, B., Alharthi, M.R., Nisar, K.S.: Solvability of functional stochastic integral equations via Darbo's fixed point theorem. *Alex. Eng. J.* **60**(6), 5631–5636 (2021)
3. Qu, Z., Liu, C., Zhu, J., Zhang, Y., Zhou, Y., Wang, L.: Two-step proximal gradient descent algorithm for photoacoustic signal unmixing. *J. Photoacoust.* **27**, 100379 (2022)
4. Jiang, X., Zeng, X., Sun, J., Chen, J.: Distributed proximal gradient algorithm for non-convex optimization over time-varying networks. *IEEE Trans. Control Netw. Syst.* **10**(2), 1005–1017 (2023)
5. Khowaja, S.A., Lee, I.H., Dev, K., Jarwar, M.A., Qureshi, N.M.F.: Get your foes fooled: proximal gradient split learning for defense against model inversion attacks on iomt data. *IEEE Trans. Netw. Sci. Eng.* **10**(5), 2607–2616 (2023)
6. Mouktonglang, T., Pochinapan, K., Suparatulatorn, R.: A parallel method for common variational inclusion and common fixed point problems with applications. *Carpath. J. Math.* **39**(1), 189–200 (2023)
7. Suantai, S., Inkrong, P., Cholamjiak, P.: Forward–backward–forward algorithms involving two inertial terms for monotone inclusions. *Comput. Appl. Math.* **42**(6), 255 (2023)
8. Jaggi, M.: An equivalence between the Lasso and support vector machines. In: Suykens, J.A.K., Signoretto, M., Argyriou, A. (eds.) *Regularization, Optimization, Kernels, and Support Vector Machines*, pp. 1–26. Chapman and Hall/CRC, Boca Raton (2014)
9. Li, Y., Bontcheva, K., Cunningham, H.: Adapting SVM for data sparseness and imbalance: a case study in information extraction. *Nat. Lang. Eng.* **15**(2), 241–271 (2009)
10. Kumar, A., Chatterjee, J.M., Díaz, V.G.: A novel hybrid approach of SVM combined with NLP and probabilistic neural network for email phishing. *Int. J. Electr. Comput. Syst. Eng.* **10**(1), 486 (2020)
11. Salloum, S., Gaber, T., Vadera, S., Shaalan, K.: A systematic literature review on phishing email detection using natural language processing techniques. *IEEE Access* **10**, 65703–65727 (2022)
12. Afrin, S., Shamrat, F.J.M., Nibir, T.I., Muntasim, M.F., Moharram, M.S., Imran, M.M., Abdulla, M.: Supervised machine learning based liver disease prediction approach with LASSO feature selection. *Bull. Electr. Inform.* **10**(6), 3369–3376 (2021)
13. Cholamjiak, W., Das, S.: A modified projective forward-backward splitting algorithm for variational inclusion problems to predict Parkinson's disease. *Appl. Math. Sci. Eng.* **32**(1), 2314650 (2024)
14. Passty, G.: Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *J. Math. Anal. Appl.* **72**(2), 383–390 (1979)
15. Polyak, B.T.: *Introduction to Optimization*. Optim. Softw. Inc., New York (1987)
16. Nesterov, Y.E.: A method for solving the convex programming problem with convergence rate  $O(\frac{1}{k^2})$ . *Sov. Math. Dokl.* **27**(2), 372–376 (1983)
17. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**(1), 183–202 (2009)
18. Ross, I.M.: Generating Nesterov's accelerated gradient algorithm by using optimal control theory for optimization. *J. Comput. Appl. Math.* **423**, 114968 (2023)
19. Oka, T., Misawa, R., Yamada, T.: Nesterov's acceleration for level set-based topology optimization using reaction-diffusion equations. *Appl. Math. Model.* **120**, 57–78 (2023)
20. Kesornprom, S., Cholamjiak, P.: A modified inertial proximal gradient method for minimization problems and applications. *AIMS Math.* **7**(5), 8147–8161 (2022)
21. Kankam, K., Cholamjiak, P.: Inertial proximal gradient method using adaptive stepsize for convex minimization problems. *Thai J. Math.* **21**(2), 277–287 (2023)
22. Mann, W.R.: Mean value methods in iteration. *Proc. Am. Math. Soc.* **4**(3), 506–510 (1953)
23. Maingé, P.E.: Convergence theorems for inertial KM-type algorithms. *J. Comput. Appl. Math.* **219**, 223–236 (2008)
24. Dong, Q.L., Cho, Y.J., Rassias, T.M.: General inertial Mann algorithms and their convergence analysis for nonexpansive mappings. In: Rassias, T.M. (ed.) *Applications of Nonlinear Analysis*, pp. 175–191 (2018)
25. Zhou, H., Qin, X.: *Fixed Points of Nonlinear Operators. Iterative Methods*. de Gruyter, Berlin (2020)
26. Brézis, H.: *Opérateurs Maximaux Monotones et Semi-groupes de Contractions dans les Espaces de Hilbert*. Math. Studies, vol. 5. North-Holland, Amsterdam (1973)
27. Hanjing, A., Suantai, S.: A fast image restoration algorithm based on a fixed point and optimization method. *Mathematics* **8**(3), 378 (2020)
28. Auslender, A., Teboulle, M., Ben-Tiba, S.: A logarithmic-quadratic proximal method for variational inequalities. *Comput. Optim. Appl.* **12**, 31–40 (1999)
29. Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer, New York (2011)
30. Liu, H., Yang, J.: Weak convergence of iterative methods for solving quasimonotone variational inequalities. *Comput. Optim. Appl.* **77**, 491–508 (2020)

## Publisher's Note

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