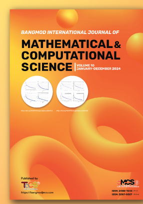
BANGMOD INTERNATIONAL JOURNAL OF MATHEMATICAL & COMPUTATIONAL SCIENCE
Volume 11 (2025)
Pages 109–134

BANGMOD INTERNATIONAL JOURNAL OF
MATHEMATICAL & COMPUTATIONAL SCIENCE
ISSN: 2408-154X (Print)
ISSN: 3057-0557 (Online)
https://bangmodjmcs.com



Strong Convergence Accelerated Alternated Inertial Relaxed Algorithm for Split Feasibilities with Applications in Breast Cancer Detection

Abdulwahab Ahmad[1,2], **Poom Kumam**[1,*]

[1] *Center of Excellence in Theoretical and Computational Science (TaCS-CoE) & Fixed Point Research Laboratory, Room SCL 802, Fixed Point Laboratory, Science Laboratory Building, Departments of Mathematics, Faculty of Science , King Mongkuts University of Technology Thonburi (KMUTT), 126 Pracha-Uthit Road Thung Khru, Bang Mod, 10140, Bangkok, Thailand*
E-mail: abumuhammaddm@gmail.com
[2] *Department of Mathematics, School of Secondary Education, Sciences, Federal College of Education, Katsina, 2041, Katsina, Nigeria*
E-mail:poom.kum@kmutt.ac.th

**Corresponding author.*

Received: 25 September 2024 / Accepted: 7 March 2025

Abstract In this article, we construct an accelerated relaxed algorithm with an alternating inertial extrapolation step. The proposed algorithm uses a three-term conjugate gradient-like direction, which helps to fasten the sequence of its iterates to a point in a solution set. The algorithm employs a self-adaptive step-length criterion that does not require any information related to the norm of the operator or the use of a line-search procedure. Moreover, we formulate and prove a strong convergence theorem for the algorithm to a minimum-norm solution of a split feasibility problem in infinite-dimensional real Hilbert spaces. Furthermore, we investigate its applications in breast cancer detection by solving classification problems for an interesting real-world breast cancer dataset, based on the extreme learning machine (ELM) with the ℓ_1-regularization approach (i.e., the Lasso model) and the ℓ_1-ℓ_2 hybrid regularization technique. The performance results of the experiments demonstrate that the proposed algorithm is robust, efficient, and achieves better generalization performance and stability than some existing algorithms in the literature.

MSC: 47J20, 47J25, 47J30, 65K15

Keywords: Relaxed $\mathcal{CQ}$ method; Alternated inertial method; Three-term Conjugate gradient method; Split feasibility problem; Classification problem

Published online: 30 April 2025

Published by Center of Excellence in Theoretical and Computational Science (TaCS-CoE)

Please cite this article as: A. Ahmad et al., Strong Convergence Accelerated Alternated Inertial Relaxed Algorithm for Split Feasibilities with Applications in Breast Cancer Detection, Bangmod Int. J. Math. & Comp. Sci., Vol. 11 (2025), 109–134. https://doi.org/10.58715/bangmodjmcs.2025.11.6

1. Introduction

In this work, $\mathcal{H}_1$ and $\mathcal{H}_2$ are two real Hilbert spaces, and $\mathcal{C} \subseteq \mathcal{H}_1$ and $\mathcal{Q} \subseteq \mathcal{H}_2$ are two nonempty, closed and convex sets. The split feasibility problem (SFP) is of finding a point $s^* \in \mathcal{C}$ if it exists, such that

$$\mathcal{A}s^* \in \mathcal{Q}, \tag{1.1}$$

where $\mathcal{A} : \mathcal{H}_1 \to \mathcal{H}_2$ is a bounded linear operator. The problem (1.1) was initially coined by Censor and Elfving [10] and has captured the attentions of many researchers, due to its vast area of applications arising from different real-world challenging problems, some of which include the problems in X-ray tomography [9, 32], machine learning [38], medical image reconstruction, signal processing, jointly constrained Nash equilibrium, and others, see for examples [13, 16, 34, 40].

In [10], the proposed algorithm requires the direct use of inverse of the operator $\mathcal{A}$, which consequently, turn out to become unrealistic in solving some large scale problems. In attempts to tackle this serious problem, extensive studies were carried out by various researchers. One of the well known and noticeable efforts in this regard was due to Byrne [6], who developed an algorithm based on the classical gradient projection method (GPM) called $\mathcal{CQ}$ algorithm, which is also a special case of the proximal forward-backward splitting method [12]. For any initial point $s_0 \in \mathcal{H}_1$, it is recursively defined by

$$s_{n+1} = P_{\mathcal{C}}\big(s_n - \tau\mathcal{A}^*(I - P_{\mathcal{Q}})\mathcal{A}s_n\big), \quad \forall n \geq 0, \tag{1.2}$$

where $P_{\mathcal{C}}$ and $P_{\mathcal{Q}}$ are metric projection operators from $\mathcal{H}_1$ and $\mathcal{H}_2$ onto $\mathcal{C}$ and $\mathcal{Q}$, respectively, $\mathcal{A}^*$ is the adjoint operator of $\mathcal{A}$ and $\tau \in (0, \frac{2}{||\mathcal{A}||^2})$ is the step length. Although Method (1.2) has the advantage that it does not require the direct use of the inverse of $\mathcal{A}$, making its implementations easier than the method in [10], it has been identified with certain challenges. The first is its requirement to compute the projections $P_{\mathcal{C}}$ and $P_{\mathcal{Q}}$ in each iteration, which depend on the geometry of the sets $\mathcal{C}$ and $\mathcal{Q}$. These are often very expensive operations or even not obtainable in some practical applications. The second is in the selection of the step-size that requires information of the norm of $\mathcal{A}$. This is also generally very hard to obtain in many practice.

By defining two sub level sets $\mathcal{C}$ and $\mathcal{Q}$ as follows:

$$\mathcal{C} = \{s \in \mathcal{H}_1 : l(s) \leq 0\} \quad \text{and} \quad \mathcal{Q} = \{t \in \mathcal{H}_2 : k(t) \leq 0\}, \tag{1.3}$$

$l : \mathcal{H}_1 \to \mathbb{R}$ and $k : \mathcal{H}_2 \to \mathbb{R}$ are weakly lower semicontinuous and convex functions and two half-spaces at a point s_n by

$$\mathcal{C}_n = \{s \in \mathcal{H}_1 : l(s_n) \leq \langle \phi_n, s_n - s\rangle\} \quad \textit{and} \quad \mathcal{Q}_n = \{t \in \mathcal{H}_2 : k(\mathcal{A}s_n) \leq \langle \varphi_n, \mathcal{A}s_n - t\rangle\}, \tag{1.4}$$

with $\phi_n \in \partial l(s_n)$, $\varphi_n \in \partial k(\mathcal{A}s_n)$, where $\mathcal{C} \subseteq \mathcal{C}_n$ and $\mathcal{Q} \subseteq \mathcal{Q}_n$ for every $n \geq 0$, Yang [44] introduced the relaxed version of method (1.2). He replaced the sets $\mathcal{C}$ and $\mathcal{Q}$ by the half-spaces $\mathcal{C}_n$ and $\mathcal{Q}_n$, respectively, so that the computations of $P_{\mathcal{C}_n}$ and $P_{\mathcal{Q}_n}$ become easier using their closed-form expressions.

On the other hand, the second challenge is still there even in the relaxed $\mathcal{CQ}$ algorithm in [44]. However, many attempts were made by some researchers to overcome it. One of the efficient results in this regard was that of Lopez et al. [29], in which they consider the following self-adaptively generated sequence of parameters $\{\tau_n\}$ to replace the fixed step-size τ.

$$\tau_n = \frac{\eta_n g(s_n)}{||\nabla g(s_n)||^2}, \tag{1.5}$$

where $0 < \eta_n < 4$ and g_n is a continuously differentiable function defined by

$$g_n(s) = \frac{1}{2}||(I - P_{\mathcal{Q}_n})\mathcal{A}s||^2 \tag{1.6}$$

and its Lipschitz continuous gradient ∇g_n given by

$$\nabla g_n(s) = \mathcal{A}^*(I - P_{\mathcal{Q}_n})\mathcal{A}s, \tag{1.7}$$

with Lipschitz constant $K = ||\mathcal{A}||^2$. Subsequently, many numerical algorithms with self-adaptive step-sizes were developed to analyze solutions to problem (1.1), see [14, 25, 27, 37].

Developing fast convergence algorithms become a flourishing research area, since they are mostly needed in many practical applications. One way to improve the speed of many iterative methods is by incorporating the Polyak's inertial step [33]. Since its inception, many algorithms were developed based on the inertial extrapolation techniques, with the numerical justifications of having faster convergence speed in several instances over their corresponding non inertial ones (see, e.g., [3, 17, 19, 21, 34, 38, 39]). However, it has been noticed that the loss in the monotonicity of the sequence produced by Polyak's inertial-type methods in relation to a point in the solution set of the problem is one of the greatest challenges associated with these methods, which in some instances, result in converging slower than their non inertial counterparts [5, 30]. In an attempt to curtail the situation, Mu and Peng [31] introduced a modified version called alternated inertial method. The identified advantage of the later inertial method over the earlier one is its ability to recover the monotonicity of the even subsequence associated with the solution set of the problem. The algorithm with alternated inertial step is defined in such away that the inertial effects are only added to odd iterations. Lately, many fast convergence algorithms were developed by various researchers based on the alternated inertial extrapolation techniques (see e.g., [2, 14, 28, 36, 37]).

Apart from the inertial approaches, it is immediately seen from (1.6) and (1.7) that all the methods mentioned above for solving problem (1.1), such as those in [6, 44, 45], are hybrid steepest descent-type methods with the directions $d_n = -\nabla g_n(s_n)$. However, as noted from [24], the accelerated versions of these methods may be obtained, when considered with a conjugate gradient-like direction or a three-term conjugate gradient-like direction, which are respectively defined by

$$d_n = -\nabla g_n(s_n) + \varsigma_n^{(1)} d_{n-1} \tag{1.8}$$

and

$$d_n = -\nabla g_n(s_n) + \varsigma_n^{(1)} d_{n-1} - \varsigma_n^{(2)} x_n, \tag{1.9}$$

where for each $i = 1, 2$, $\varsigma_n^{(i)} \in [0, \infty)$ and $x_n \in \mathcal{H}_1$ is an arbitrary point. In [24], the authors have numerically shown that the hybrid gradient method with the direction (1.9), converges faster than its variant with the direction (1.8) when $\lim_{n\to\infty} \varsigma_n^i = 0$, for $i = 1, 2$ and $\{x_n\}$ is bounded. In this direction, many iterative methods using conjugate gradient-like directions for solving different nonlinear problems were suggested, see [1, 15, 20, 26]. Very recently, by combining relaxed algorithm with Polyak's inertial term and conjugate gradient-like direction (1.8), Che et al. [11] introduced the following algorithm for problem (1.1).

Algorithm 1 (CZWC)

Initialization: Choose c-contraction function $h : \mathbb{R}^k \to \mathbb{R}^k$ and strongly positive bounded linear operator $F : \mathbb{R}^k \to \mathbb{R}^k$ with coefficient $\delta > 0$. Take $\gamma \in (0, \frac{\delta}{c})$, ϑ, δ_n η_1, $\eta_n \in (0, 1)$, λ_1, $\lambda_n \in [0, 1)$ $\varsigma_n \in [0, \frac{1}{2})$, $\varepsilon > 0$ and choose s_0, $s_1 \in \mathbb{R}^k$.

Step 1. Set $u_1 = s_1 + \lambda_1(s_1 - s_0)$, If $||(I - P_{\mathcal{C}_1})u_1 + \mathcal{A}^T(I - P_{\mathcal{Q}_1})\mathcal{A}u_1|| \leq \varepsilon$, terminate; otherwise, set

$$d_0 = -\tau_1 \nabla g_1(u_1), \text{ with } \tau_1 = \frac{2\eta_1\left(||(I - P_{\mathcal{C}_1})u_1||^2 + ||(I - P_{\mathcal{Q}_1})\mathcal{A}u_1||^2\right)}{||(I - P_{\mathcal{C}_1})u_1 + \mathcal{A}^T(I - P_{\mathcal{Q}_1})\mathcal{A}u_1||^2}.$$

Set $n = 1$ and go to Step 2.

Step 2. Compute

$$\tau_n = \frac{2\eta_n\left(||(I - P_{\mathcal{C}_n})u_n||^2 + ||(I - P_{\mathcal{Q}_n})\mathcal{A}u_n||^2\right)}{||(I - P_{\mathcal{C}_n})u_n + \mathcal{A}^T(I - P_{\mathcal{Q}_n})\mathcal{A}u_n||^2},$$
$$d_n = -\tau_n \nabla g_n(u_n) + \mu\varsigma_n d_{n-1},$$
$$p_n = u_n + d_n.$$

Step 3. Compute

$$s_{n+1} = \delta_n \gamma h(p_n) + (1 - \delta_n F)P_{\mathcal{C}_n} p_n, \text{ set } n = n + 1 \text{ and go to Step 4.}$$

Step 4. Set $u_n = s_n + \lambda_n(s_n - s_{n-1})$. If $||(I - P_{\mathcal{C}_n})u_n + \mathcal{A}^T(I - P_{\mathcal{Q}_n})\mathcal{A}u_n|| \leq \varepsilon$, terminate; otherwise, go to step 2.

They formulated a theorem for Algorithm 1 and proved its strong convergence to a solution of problem (1.1) under the following conditions.

$(A1)$ $\lim_{n\to\infty} \delta_n = 0$, $\sum_{n=1}^{\infty} \delta_n = \infty$ and $\varsigma_n = \delta_n^2$;

$(A2)$ $\lim_{n\to\infty} \frac{\lambda_n}{\delta_n}||x_n - x_{n-1}|| = 0$ and $0 < \liminf_{n\to\infty} \eta_n < \limsup_{n\to\infty} \eta_n < 1$;

$(A3)$ $\{(I - P_{\mathcal{C}_n})u_n\}$ and $\{(I - P_{\mathcal{Q}_n})\mathcal{A}u_n\}$ are bounded.

Although the relaxed Algorithm 1 with Polyak's inertial step and conjugate gradient-like direction (1.8) has achieved some remarkable performance on signal and image restoration problems, see [11], it is natural to ask the question: *Can we modify Algorithm 1 and incorporate the modified version with the alternated inertial step and the three-term conjugate gradient-like direction* (1.9) *to improve its computational efficiency in infinite-dimensional Hilbert spaces*? Inspired and motivated by the results in [11, 24, 29, 31], we provide answer to this question in an affirmative. Moreover, we improve the choice of the inertial parameter λ_n such that the condition $(A2)$ is not needed and this simplify the implementations of the proposed algorithm. For the applications, we analyse the efficiency of our proposed algorithm in solving classification problems for an interesting real-world dataset based ℓ_1-regularization and ℓ_1-ℓ_2 hybrid regularization models.

2. Preliminaries

In this paper, we use $s_n \rightharpoonup s$ (resp., $s_n \to s$) to represents the weak (resp., strong) convergence of the sequence $\{s_n\}$ to s. Let $\mathcal{H}$ be a real Hilbert space. For all $s,\ t \in \mathcal{H}$ and $\lambda \in [0,1]$, we use the following identities:

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

and

$$||\lambda s + (1-\lambda)t||^2 = \lambda||s||^2 + (1-\lambda)||t||^2 - \lambda(1-\lambda)||s-t||^2. \tag{2.2}$$

Definition 2.1. (see [4]) Let $T : \mathcal{H} \to \mathcal{H}$ be a map, then T is called

- K - Lipschitz continuous with $K > 0$, if

$$||Ts - Tt|| \leq K||s-t||, \quad \forall s,\ t \in \mathcal{H}. \tag{2.3}$$

- Nonexpansive, if $K = 1$ in (2.3).

- Firmly nonexpansive, if

$$||Ts - Tt|| \leq \langle s-t, Ts-Tt\rangle, \quad \forall s,\ t \in \mathcal{H}. \tag{2.4}$$

Recall that for any $s \in \mathcal{H}$ and a nonempty, closed and convex set $\mathcal{C} \subset \mathcal{H}$, an element $P_{\mathcal{C}}s \in \mathcal{C}$ such that

$$||s - P_{\mathcal{C}}s|| \leq ||s-t||, \quad \forall t \in \mathcal{C}, \tag{2.5}$$

is uniquely determined, where $P_{\mathcal{C}}$ is termed as the metric operator defining a projection of $\mathcal{H}$ onto $\mathcal{C}$. Additionally, $\forall\ s \in \mathcal{H}$ and $t \in \mathcal{C}$, the following are some of the properties of the element $P_{\mathcal{C}}s$ (see [18]):

$$\langle s - P_{\mathcal{C}}s, t - P_{\mathcal{C}}s\rangle \leq 0. \tag{2.6}$$

Note that (2.6) is equivalent to

$$||s - P_{\mathcal{C}}s||^2 + ||t - P_{\mathcal{C}}s||^2 \leq ||s-t||^2. \tag{2.7}$$

Remark 2.2. It is commonly known that $I - P_{\mathcal{C}}$, where I is the identity operator, satisfies the inequality (2.4) (see [43]).

Definition 2.3. (see [4]) Let $g : \mathcal{H} \to (-\infty, +\infty]$ be a convex and proper function. Then:

- g is said to be (weakly) lower semicontinuous (w-lsc) if for any sequence $\{s_n\} \subset \mathcal{H}$ such that $(s_n \rightharpoonup s^*)$ $s_n \to s^*$ as $n \to \infty$, we have

$$\liminf_{n\to\infty} g(s_n) \geq g(s^*). \tag{2.8}$$

- The subdifferential of g denoted by $\partial g(s)$ at a point s is defined by

$$\partial g(s) := \{v \in \mathcal{H} :\ \langle v, t-s\rangle + g(s) \leq g(t)\ \forall t \in \mathcal{H}\}.$$

An element $v \in \partial g(s)$ is called a subgradient of g at s.

Lemma 2.4. *(see e.g., [8, 43]) Let $\tau > 0$ and $s^* \in \mathcal{H}$, then, the following statements are equivalent.*

- *s^* solves the problem (1.1)*
- *s^* solves the fixed point problem*

$$s^* = P_{\mathcal{C}}(s^* - \tau A^*(I - P_{\mathcal{Q}})As^*).$$

Lemma 2.5. *(see [22]) Let $\{s_n\}$ be a sequence of nonnegative real numbers, such that $\forall n \geq 1$,*

$$s_{n+1} \leq (1 - \beta_n)s_n + \beta_n \Gamma_n,$$
$$s_{n+1} \leq s_n - \chi_n + \omega_n,$$

where $\beta_n \in (0,1)$, $\chi_n \in [0, +\infty)$ and Γ_n, $\omega_n \in (-\infty, +\infty)$ such that (A1) $\sum_{n=1}^{\infty} \beta_n = \infty$, (A2) $\lim\limits_{n\to\infty} \omega_n = 0$ and (A3) $\lim\limits_{r\to\infty} \chi_{n_r} = 0$ implies that $\limsup\limits_{r\to\infty} \Gamma_{n_r} \leq 0$, for any subsequence $\{n_r\}$ of $\{n\}$. Then, $\lim\limits_{n\to\infty} s_n = 0$.

3. Iterative method and its convergence analysis

In this section, we provide the modified self-adaptive relaxed $\mathcal{CQ}$ algorithm with alternated inertial step and three-term conjugate gradient-like direction for a solution of problem (1.1) and analyze its strong convergence in real Hilbert spaces. The proposed algorithm in this paper is constructed by defining $\mathcal{C}$, $\mathcal{Q}$, $\mathcal{C}_n$, $\mathcal{Q}_n$, g_n and its gradient ∇g_n as in (1.3), (1.4), (1.6) and (1.7), respectively. Moreover, we make the following assumptions to analysze its convergence.

Assumption 3.1. (A1) The solutions' set of problem (1.1) is denoted by $\Omega \neq \emptyset$.
(A2) $l : \mathcal{H}_1 \to \mathbb{R}$ and $k : \mathcal{H}_2 \to \mathbb{R}$ are respectively convex, subdifferentiable and weakly lower semicontinuous functions on $\mathcal{H}_1$ and $\mathcal{H}_2$.
(A3) For any $s \in \mathcal{H}_1$ and $t \in \mathcal{H}_2$, at least one subgradient $\phi \in \partial l(s)$ and $\varphi \in \partial k(t)$ are obtainable and the subdifferential operators ∂l and ∂k are bounded on bounded sets.

Assumption 3.2. (B1) Let λ_n, α_n, δ_n, $\beta_n \in (0,1)$ such that $\lim\limits_{n\to\infty} \alpha_n, \beta_n = 0$, $\lim\limits_{n\to\infty} \delta_n \neq 0$ and $\sum_{n=1}^{\infty} \beta_n = \infty$ with $\lim\limits_{n\to\infty} \frac{\alpha_n}{\beta_n} = 0$, and $\lambda_n < \frac{(1-\delta_{n-1})}{\delta_{n-1}(1+\delta_{n-1})}$.
(B2) Let $\eta_n \in (0, \frac{2}{\varepsilon})$ for $\varepsilon > 0$ such that $\liminf\limits_{n\to\infty} \eta_n(2 - \eta_n) > 0$ and $\pi_n \in (0,4)$ such that $\liminf\limits_{n\to\infty} \pi_n(4 - \pi_n) > 0$.
(B3) Let $\vartheta > 0$ and $\varsigma_n^i \in [0, +\infty)$ for each $i = 1, 2$ such that $\lim\limits_{n\to\infty} \varsigma_n^i = 0$ and $\lim\limits_{n\to\infty} \frac{\varsigma_n^{(i)}}{\beta_n} = 0$.
(B4) $\{(I - P_{\mathcal{Q}})\mathcal{A}h_n\}$ and $\{A_{g_n} h_n - z\}$ for any $z \in \Omega$ are bounded.

Algorithm 2 Strong Convergence Accelerated Alternated Inertial Relaxed $\mathcal{CQ}$ Algorithm (SCAAiRA)

Initialization: Take ϑ, ε, $\{\delta_n\}$, $\{\beta_n\}$, $\{\alpha_n\}$, $\{\lambda_n\}$, $\{\eta_n\}$, $\{\pi_n\}$ and for each $i = 1, 2$, $\{\varsigma_n^i\}$ such that the conditions of Assumption 3.2 hold. Choose a bounded sequence $\{x_n\} \subset \mathcal{H}_1$, $s_0,\ s_1 \in \mathcal{H}_1$ and $d_0 = -\nabla g_0(s_0)$. Set $n = 1$

Step 1. Compute

$$u_n = \begin{cases} s_n, & if\ n = even, \\ s_n + \lambda_n(s_n - s_{n-1}), & if\ n = odd. \end{cases} \tag{3.1}$$

Step 2. Compute $h_n = (1-\alpha_n)(u_n - \varepsilon\tau_n \nabla g_n(s_n))$, where the step size τ_n is obtain by the relations

$$\tau_n = \begin{cases} \frac{\eta_n g_n(s_n)}{||\nabla g_n(s_n)||^2}, & if\ ||\nabla g_n(s_n)|| \neq 0, \\ 0, & if\ ||\nabla g_n(s_n)|| = 0. \end{cases} \tag{3.2}$$

Step 3. Compute $d_{n+1} = \frac{1}{\vartheta}\left(A_{g_n} h_n - h_n\right) + \varsigma_n^1 d_n - \varsigma_n^2 x_n$ and $p_n = h_n + \vartheta d_{n+1}$, where

$$A_{g_n} = I - \theta_n \nabla g_n$$

and

$$\theta_n = \begin{cases} \frac{\pi_n g_n(h_n)}{||\nabla g_n(h_n)||^2}, & if\ ||\nabla g_n(h_n)|| \neq 0, \\ 0, & if\ ||\nabla g_n(h_n)|| = 0. \end{cases} \tag{3.3}$$

Step 4. Compute

$$m_n = P_{\mathcal{C}_n}(1-\beta_n)p_n \text{ and } s_{n+1} = (1-\delta_n)u_n + \delta_n m_n.$$

Set $n := n+1$ and go back to Step 1.

We first prove the following lemma.

Lemma 3.3. *Let $\{m_n\}$, $\{u_n\}$ and $\{s_n\}$ be the sequences produced by Algorithm 2. Then, for any $z \in \Omega$, we have*

$$\begin{aligned} ||m_n - z||^2 \leq{}& (1-\beta_n)||u_n - z||^2 + (\beta_n + (1-\beta_n)\alpha_n)||z||^2 + (1-\alpha_n)(1-\beta_n)||u_n - s_n||^2 \\ &+ \vartheta^2(1-\beta_n)||\varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n||^2 + 2\vartheta(1-\beta_n)\left\langle A_{g_n} h_n - z, \varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n \right\rangle \\ &- 2\varepsilon\eta_n(2-\varepsilon\eta_n)(1-\alpha_n)(1-\beta_n)\frac{g_n^2(s_n)}{||\nabla g_n(s_n)||^2} - \pi_n(4-\pi_n)(1-\beta_n)\frac{g_n^2(h_n)}{||\nabla g_n(h_n)||^2} \end{aligned}$$

and

$$\begin{aligned} ||m_n - z||^2 \leq{}& (1-\beta_n)||u_n - z||^2 + \left(\beta_n^2 + (1-\beta_n)\alpha_n^2\right)||z||^2 + 2\alpha_n(1-\alpha_n)(1-\beta_n)\Big(\langle u_n - z, -z\rangle \\ &+ \varepsilon\tau_n||\nabla g_n(s_n)||||z||\Big) + 2\beta_n(1-\beta_n)\Big(\langle h_n - z, -z\rangle + \theta_n||\nabla g_n(h_n)||||z|| + \vartheta||\varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n||||z||\Big) + \\ &(1-\beta_n)(1-\alpha_n)||u_n - s_n||^2 + \vartheta(1-\beta_n)\Big(2\left\langle A_{g_n} h_n - z, \varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n \right\rangle \\ &+ ||\varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n||^2\Big). \end{aligned}$$

Proof. Observe that

$$\begin{aligned} p_n &= h_n + \vartheta d_{n+1} \\ &= A_{g_n} h_n + \vartheta \varsigma_n^1 d_n - \vartheta \varsigma_n^2 x_n. \end{aligned} \tag{3.4}$$

Let $z \in \Omega$. Then, we have $\mathcal{A}z \in \mathcal{Q}_n$. Consequently, $\nabla g_n(z) = \mathcal{A}^*(I - P_{\mathcal{Q}_n})\mathcal{A}z = 0$. Therefore, together with the fact that $I - P_{\mathcal{Q}_n}$ satisfies (2.4), we have

$$\begin{aligned} \langle \nabla g_n(s_n), s_n - z \rangle &= \langle \nabla g_n(s_n) - \nabla g_n(z), s_n - z \rangle \\ &= \langle \mathcal{A}^*(I - P_{\mathcal{Q}_n})\mathcal{A}s_n - \mathcal{A}^*(I - P_{\mathcal{Q}_n})\mathcal{A}z, s_n - z \rangle \\ &= \langle (I - P_{\mathcal{Q}_n})\mathcal{A}s_n - (I - P_{\mathcal{Q}_n})\mathcal{A}z, \mathcal{A}s_n - \mathcal{A}z \rangle \\ &\geq ||(I - P_{\mathcal{Q}_n})\mathcal{A}s_n||^2 \\ &= 2g_n(s_n). \end{aligned} \tag{3.5}$$

Now, let $a_n = u_n - \varepsilon\tau_n \nabla g_n(s_n)$, $z \in \Omega$. Then the identity (2.1) and inequality (3.5) imply that

$$\begin{aligned} ||a_n - z||^2 &= ||u_n - \varepsilon\tau_n \nabla g_n(s_n) - z||^2 \\ &= ||u_n - z||^2 + \varepsilon^2\tau_n^2 ||\nabla g_n(s_n))||^2 - 2\varepsilon\tau_n \langle \nabla g_n(s_n), s_n - z \rangle \\ &\quad + 2\varepsilon\tau_n \langle \nabla g_n(s_n), s_n - u_n \rangle \\ &\leq ||u_n - z||^2 + 2\varepsilon^2\tau_n^2 ||\nabla g_n(s_n))||^2 - 4\varepsilon\tau_n g_n(s_n) + ||s_n - u_n||^2 \\ &= ||u_n - z||^2 + ||u_n - s_n||^2 - 2\varepsilon\eta_n(2 - \varepsilon\eta_n) \frac{g_n^2(s_n)}{||\nabla g_n(s_n)||^2}. \end{aligned} \tag{3.6}$$

Thanks to condition (B2) of Assumption 3.2, we obtain

$$||a_n - z||^2 \leq ||u_n - z||^2 + ||u_n - s_n||^2. \tag{3.7}$$

The convexity of $||.||^2$ and inequality (3.6) imply

$$\begin{aligned} ||h_n - z||^2 &= ||(1 - \alpha_n)a_n - z||^2 \\ &\leq \alpha_n ||z||^2 + (1 - \alpha_n)||a_n - z||^2 \\ &\leq \alpha_n ||z||^2 + (1 - \alpha_n)\Big(||u_n - z||^2 + ||u_n - s_n||^2 - 2\varepsilon\eta_n(2 - \varepsilon\eta_n) \frac{g_n^2(s_n)}{||\nabla g_n(s_n)||^2} \Big) \\ &= ||u_n - z||^2 + \alpha_n ||z||^2 - 2\varepsilon\eta_n(2 - \varepsilon\eta_n)(1 - \alpha_n) \frac{g_n^2(s_n)}{||\nabla g_n(s_n))||^2} \\ &\quad + (1 - \alpha_n)||u_n - s_n||^2. \end{aligned} \tag{3.8}$$

In a similar fashion, the identity (2.1) and inequality (3.7) provide

$$\begin{aligned} ||h_n - z||^2 &= \alpha_n^2 ||z||^2 + (1 - \alpha_n)^2 ||a_n - z||^2 + 2\alpha_n(1 - \alpha_n) \langle a_n - z, -z \rangle \\ &\leq \alpha_n^2 ||z||^2 + (1 - \alpha_n)^2 (||u_n - z||^2 + ||u_n - s_n||^2) \\ &\quad + 2\alpha_n(1 - \alpha_n) \langle u_n - \varepsilon\tau_n \nabla g_n(s_n) - z, -z \rangle \\ &= \alpha_n^2 ||z||^2 + (1 - \alpha_n)^2 (||u_n - z||^2 + ||u_n - s_n||^2) \\ &\quad + 2\alpha_n(1 - \alpha_n) \langle u_n - z, -z \rangle + 2\alpha_n(1 - \alpha_n) \langle \varepsilon\tau_n \nabla g_n(s_n), z \rangle \\ &\leq ||u_n - z||^2 + \alpha_n \Big(\alpha_n ||z||^2 + 2(1 - \alpha_n)\big(\langle u_n - z, -z \rangle \\ &\quad + \varepsilon\tau_n ||\nabla g_n(s_n)|| ||z||\big) \Big) + (1 - \alpha_n)||u_n - s_n||^2. \end{aligned} \tag{3.9}$$

In view of the definition of A_{g_n}, (3.4), (3.5), (3.8) and (3.9), one respectively finds that

$$\begin{aligned}
||p_n - z||^2 &= ||A_{g_n} h_n + \vartheta \varsigma_n^1 d_n - \vartheta \varsigma_n^2 x_n - z||^2 \\
&\leq ||h_n - z||^2 + \vartheta^2 ||\varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n||^2 \\
&\quad + 2\vartheta \left\langle A_{g_n} h_n - z, \varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n \right\rangle \\
&\quad - \pi_n (4 - \pi_n) \frac{g_n^2(h_n)}{||\nabla g_n(h_n)||^2} \\
&\leq ||u_n - z||^2 + \alpha_n ||z||^2 - 2\varepsilon\eta_n (2 - \varepsilon\eta_n)(1 - \alpha_n) \frac{g_n^2(s_n)}{||\nabla g_n(s_n)||^2} \\
&\quad + (1 - \alpha_n)||u_n - s_n||^2 + \vartheta^2 ||\varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n||^2 \\
&\quad + 2\vartheta \left\langle A_{g_n} h_n - z, \varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n \right\rangle \\
&\quad - \pi_n (4 - \pi_n) \frac{g_n^2(h_n)}{||\nabla g_n(h_n)||^2} \qquad (3.10)
\end{aligned}$$

and

$$\begin{aligned}
||p_n - z||^2 &\leq ||h_n - z||^2 + \vartheta^2 ||\varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n||^2 + 2\vartheta \left\langle A_{g_n} h_n - z, \varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n \right\rangle \\
&\leq \alpha_n \Big(\alpha_n ||z||^2 + 2(1 - \alpha_n)\big(\langle u_n - z, -z \rangle + \varepsilon\tau_n ||\nabla g_n(s_n)|| ||z|| \big) \Big) \\
&\quad + ||u_n - z||^2 + (1 - \alpha_n)||u_n - s_n||^2 + \vartheta^2 ||\varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n||^2 \\
&\quad + 2\vartheta \left\langle A_{g_n} h_n - z, \varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n \right\rangle. \qquad (3.11)
\end{aligned}$$

The convexity of $||.||^2$, identity (2.7) and inequality (3.10) lead to obtain that

$$\begin{aligned}
||m_n - z||^2 &\leq (1 - \beta_n)||u_n - z||^2 + (\beta_n + (1 - \beta_n)\alpha_n)||z||^2 \\
&\quad + (1 - \alpha_n)(1 - \beta_n)||u_n - s_n||^2 + \vartheta^2 (1 - \beta_n)||\varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n||^2 \\
&\quad + 2\vartheta(1 - \beta_n) \left\langle A_{g_n} h_n - z, \varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n \right\rangle \\
&\quad - 2\varepsilon\eta_n (2 - \varepsilon\eta_n)(1 - \alpha_n)(1 - \beta_n) \frac{g_n^2(s_n)}{||\nabla g_n(s_n)||^2} \\
&\quad - \pi_n (4 - \pi_n)(1 - \beta_n) \frac{g_n^2(h_n)}{||\nabla g_n(h_n)||^2}. \qquad (3.12)
\end{aligned}$$

We equivalently see from (2.1) and (3.11) that

$$\begin{aligned}
||m_n - z||^2 &\leq (1 - \beta_n)||u_n - z||^2 + \big(\beta_n^2 + (1 - \beta_n)\alpha_n^2\big)||z||^2 \\
&\quad + 2\alpha_n (1 - \alpha_n)(1 - \beta_n)\Big(\langle u_n - z, -z \rangle + \varepsilon\tau_n ||\nabla g_n(s_n)|| ||z|| \Big) \\
&\quad + 2\beta_n (1 - \beta_n)\Big(\langle h_n - z, -z \rangle + \theta_n ||\nabla g_n(h_n)|| ||z|| \\
&\quad + \vartheta ||\varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n|| ||z|| \Big) + (1 - \beta_n)(1 - \alpha_n)||u_n - s_n||^2 \\
&\quad + \vartheta(1 - \beta_n)\Big(2 \left\langle A_{g_n} h_n - z, \varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n \right\rangle \\
&\quad + \vartheta ||\varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n||^2 \Big). \qquad (3.13)
\end{aligned}$$

∎

In the next lemma, we establish the boundedness of the sequence $\{d_n\}$ produced by Algorithm 2.

Lemma 3.4. *Let $\{d_n\}$ be the sequence formed by Algorithm 2, then it is bounded.*

Proof. It is not difficult to see from (B3) of Assumption 3.2 that for each $i = 1, 2$, $\lim_{n\to\infty} \varsigma_n^{(i)} = 0$, thus the existence of $n_0 \in \mathbb{N}$ such that $\varsigma_n^{(i)} \leq \frac{1}{4}$, $\forall\ n \geq n_0$ is guaranteed. Similarly, from (B4) of Assumption 3.2, one finds that $\{\theta_n \nabla g_n(h_n)\}$ is bounded. Let us define K as follows:

$$K = \max\left\{\max_{1\leq j\leq n_0} ||d_j||,\ \sup_{n\in\mathbb{N}} ||x_n||,\ \frac{2}{\vartheta}\sup_{n\in\mathbb{N}} \theta_n ||\nabla g_n(h_n)||\right\}.$$

Then, combining with the fact that $\{x_n\}$ is bounded, we find that $K < \infty$. Now, assume that $||d_n|| \leq K$ for some $n \geq n_0$, then

$$\begin{aligned} ||d_{n+1}|| &= ||\frac{1}{\vartheta}(A_{g_n} h_n - h_n + \varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n|| \\ &\leq \frac{1}{2}\big(\frac{2}{\vartheta}\theta_n ||\nabla g_n(h_n)||\big) + \varsigma_n^{(1)} ||d_n|| + \varsigma_n^{(2)} ||x_n|| \\ &\leq K. \end{aligned} \tag{3.14}$$

This implies that

$$||d_n|| \leq K,\ n \geq 0,$$

so, we conclude that $\{d_n\}$ is bounded. ∎

In what follows, we prove that the even subsequence of the sequence $\{s_n\}$ produced by Algorithm 2 is bounded.

Lemma 3.5. *Let $\{s_n\}$ be the sequence generated by Algorithm 2. Then, for any point $z \in \Omega$, the even subsequence $\{||s_{2n} - z||\}$ of $\{||s_n - z||\}$ is bounded.*

Proof. We observe from Algorithm 2 that for any $z \in \Omega$, we obtain from (3.12) that

$$\begin{aligned} ||s_{n+1} - z||^2 \leq\ & (1 - \delta_n\beta_n)||u_n - z||^2 + \delta_n(\beta_n + (1-\beta_n)\alpha_n)||z||^2 - \delta_n(1-\delta_n)||u_n - m_n||^2 \\ & + \delta_n(1-\beta_n)(1-\alpha_n)\Big(||u_n - s_n||^2 - 2\varepsilon\eta_n(2 - \varepsilon\eta_n)\frac{g_n^2(s_n)}{||\nabla g_n(s_n))||^2}\Big) \\ & + \delta_n(1-\beta_n)\Big(\vartheta^2||\varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n||^2 - \pi_n(4-\pi_n)\frac{g_n^2(h_n)}{||\nabla g_n(h_n)||^2}\Big) \\ & + \delta_n 2(1-\beta_n)\vartheta\left\langle A_{g_n} h_n - z, \varsigma_n^{(1)} d_n - \varsigma_n^{(2)} x_n\right\rangle. \end{aligned} \tag{3.15}$$

In view of (3.1), one sees from (3.15) that

$$\begin{aligned}||s_{2n+1}-z||^2 \leq\ & (1-\delta_{2n}\beta_{2n})||s_{2n}-z||^2+\delta_{2n}(\beta_{2n}+(1-\beta_{2n})\alpha_{2n})||z||^2\\ &+\delta_{2n}\vartheta^2(1-\beta_{2n})||\varsigma_{2n}^{(1)}d_{2n}-\varsigma_{2n}^{(2)}x_{2n}||^2\\ &+2\delta_{2n}\vartheta(1-\beta_{2n})\left\langle A_{g_{2n}}h_{2n}-z,\varsigma_{2n}^{(1)}d_{2n}-\varsigma_{2n}^{(2)}x_{2n}\right\rangle\\ &-\pi_{2n}(4-\pi_{2n})\delta_{2n}(1-\beta_{2n})\frac{g_{2n}^2(h_{2n})}{||\nabla g_{2n}(h_{2n})||^2}\\ &-2\varepsilon\eta_{2n}(2-\varepsilon\eta_{2n})\delta_{2n}(1-\beta_{2n})(1-\alpha_{2n})\frac{g_{2n}^2(s_{2n})}{||\nabla g_{2n}(s_{2n}))||^2}\\ &-\delta_{2n}(1-\delta_{2n})||s_{2n}-m_{2n}||^2. \end{aligned} \tag{3.16}$$

From the definition of s_{2n+1} and that of u_{2n} in (3.1), we have

$$\begin{aligned}||s_{2n+1}-s_{2n}||^2 &= ||(1-\delta_{2n})(u_{2n}-s_{2n})+\delta_{2n}(m_{2n}-s_{2n}||^2\\ &= \delta_{2n}^2||m_{2n}-s_{2n}||^2.\end{aligned} \tag{3.17}$$

Combining (3.16), (3.17) and the definition of u_{2n+1} in (3.1), we equivalently have

$$\begin{aligned}||u_{2n+1}-z||^2 \leq\ & (1-\delta_{2n}\beta_{2n})||s_{2n}-z||^2+\delta_{2n}(1+\lambda_{2n+1})\Bigg((\beta_{2n}+(1-\beta_{2n})\alpha_{2n})||z||^2\\ &+\vartheta(1-\beta_{2n})\Big(\vartheta||\varsigma_{2n}^{(1)}d_{2n}-\varsigma_{2n}^{(2)}x_{2n}||^2-\pi_{2n}(4-\pi_{2n})\frac{g_{2n}^2(h_{2n})}{||\nabla g_{2n}(h_{2n})||^2}\Big)\\ &+2(1-\beta_{2n})\left\langle A_{g_{2n}}h_{2n}-z,\varsigma_{2n}^{(1)}d_{2n}-\varsigma_{2n}^{(2)}x_{2n}\right\rangle\\ &-\big((1-\delta_{2n})-\lambda_{2n+1}\delta_{2n}\big)||m_{2n}-s_{2n}||^2\\ &-2\varepsilon\eta_{2n}(2-\varepsilon\eta_{2n})(1-\beta_{2n})(1-\alpha_{2n})\frac{g_{2n}^2(s_{2n})}{||\nabla g_{2n}(s_{2n}))||^2}\Bigg).\end{aligned} \tag{3.18}$$

In view of (3.17) and the definition of u_{2n+1} in (3.1), we find that

$$||u_{2n+1}-s_{2n+1}||^2\leq\lambda_{2n+1}(1+\lambda_{2n+1})\delta_{2n}^2||m_{2n}-s_{2n}||^2. \tag{3.19}$$

Using (3.15), (3.18), (3.19) and the fact that $\beta_n,\ \delta_n\in(0,1)\ \forall n\geq 1$, we get

$$\begin{aligned}||s_{2n+2}-z||^2 \leq\ & (1-\delta_{2n}\beta_{2n})||s_{2n}-z||^2+\delta_{2n}(1+\lambda_{2n+1})\big(2(\beta_{2n}+\alpha_{2n})||z||^2+\Phi_{2n}\big)\\ &-2\delta_{2n}(1+\lambda_{2n+1})\varepsilon\eta_{2n}(2-\varepsilon\eta_{2n})(1-\beta_{2n})(1-\alpha_{2n})\frac{g_{2n}^2(s_{2n})}{||\nabla g_{2n}(s_{2n}))||^2}\\ &-2\varepsilon\eta_{2n+1}(2-\varepsilon\eta_{2n+1})\delta_{2n+1}(1-\beta_{2n+1})(1-\alpha_{2n+1})\frac{g_{2n+1}^2(s_{2n+1})}{||\nabla g_{2n+1}(s_{2n+1}))||^2}\\ &-\delta_{2n}(1+\lambda_{2n+1})\pi_{2n}(4-\pi_{2n})(1-\beta_{2n})\frac{g_{2n}^2(h_{2n})}{||\nabla g_{2n}(h_{2n})||^2}\\ &-\pi_{2n+1}(4-\pi_{2n+1})\delta_{2n+1}(1-\beta_{2n+1})\frac{g_{2n+1}^2(h_{2n+1})}{||\nabla g_{2n+1}(h_{2n+1})||^2}\\ &-\delta_{2n}(1+\lambda_{2n+1})\big((1-\delta_{2n})-\lambda_{2n+1}\delta_{2n}(1+\delta_{2n})\big)||m_{2n}-s_{2n}||^2,\end{aligned} \tag{3.20}$$

where $\Phi_{2n} = \vartheta^2\big(||\varsigma_{2n}^{(1)}d_{2n} - \varsigma_{2n}^{(2)}x_{2n}||^2 + ||\varsigma_{2n+1}^{(1)}d_{2n+1} - \varsigma_{2n+1}^{(2)}x_{2n+1}||^2\big) + 2\vartheta||\varsigma_{2n+1}^{(1)}d_{2n+1} - \varsigma_{2n+1}^{(2)}x_{2n+1}||||A_{g_{2n+1}}h_{2n+1} - z|| + 2\vartheta(1-\beta_{2n})\left\langle A_{g_{2n}}h_{2n} - z, \varsigma_{2n}^{(1)}d_{2n} - \varsigma_{2n}^{(2)}x_{2n}\right\rangle$.

Taking $M = \sup\limits_{n\geq 1}(1+\lambda_{2n+1})\big(2(1+\frac{\alpha_{2n}}{\beta_{2n}})||z||^2 + \frac{\Phi_{2n}}{\beta_{2n}}\big)$, then, together with Assumption 3.2 and inequality (3.20), we obtain

$$\begin{aligned} |s_{2n+2} - z||^2 &\leq (1-\delta_{2n}\beta_{2n})||s_{2n} - z||^2 + \delta_{2n}\beta_{2n}M \\ &\leq \max\left\{||s_{2n} - z||^2, M\right\} \\ &\vdots \\ &\leq \max\left\{||s_0 - z||^2, M\right\}. \end{aligned} \tag{3.21}$$

From conditions (B1), (B3) and inequality (3.21), we obtain that the even subsequence $\{||s_{2n} - z||\}$ with respect to a point $z \in \Omega$ is bounded. So that the boundedness of an even subsequence $\{s_{2n}\}$ of $\{s_n\}$ generated by Algorithm 2 is obtained. Consequently, from the inequality (3.16), one easily sees that an odd subsequence $\{s_{2n+1}\}$ of the same sequence is also bounded. ∎

Lemma 3.6. *Let $\{s_{2n}\}$ be an even subsequence of $\{s_n\}$ generated by Algorithm 2. Then, for any point $z \in \Omega$, the following holds:*

$$||s_{2n+2} - z||^2 \leq (1-\delta_{2n}\beta_{2n})||s_{2n} - z||^2 + \delta_{2n}\beta_{2n}\Gamma_{2n}. \tag{3.22}$$

Proof. Let $z \in \Omega$. Then, it follows from inequality (3.13) and Algorithm 2 that

$$\begin{aligned} ||s_{n+1} - z||^2 \leq\ & (1-\delta_n\beta_n)||u_n - z||^2 + \delta_n\Bigg(\big(\beta_n^2 + (1-\beta_n)\alpha_n^2\big)||z||^2 - (1-\delta_n)||u_n - m_n||^2 \\ & + 2\alpha_n(1-\alpha_n)(1-\beta_n)\Big(\langle u_n - z, -z\rangle + \varepsilon\tau_n||\nabla g_n(s_n)||||z||\Big) \\ & + 2\beta_n(1-\beta_n)\Big(\langle h_n - z, -z\rangle + \theta_n||\nabla g_n(h_n)||||z|| \\ & + \vartheta||\varsigma_n^{(1)}d_n - \varsigma_n^{(2)}x_n||||z||\Big) + \vartheta(1-\beta_n)\Big(2\left\langle A_{g_n}h_n - z, \varsigma_n^{(1)}d_n - \varsigma_n^{(2)}x_n\right\rangle \\ & + ||\varsigma_n^{(1)}d_n - \varsigma_n^{(2)}x_n||^2\Big) + (1-\alpha_n)(1-\beta_n)||u_n - s_n||^2\Bigg). \end{aligned} \tag{3.23}$$

Combining (3.23) and the same arguments used in deducing (3.16), we find that

$$\begin{aligned} ||s_{2n+1} - z||^2 \leq\ & (1-\delta_{2n}\beta_{2n})||s_{2n} - z||^2 + \delta_{2n}\Bigg(\big(\beta_{2n}^2 + (1-\beta_{2n})\alpha_{2n}^2\big)||z||^2 \\ & + 2\alpha_{2n}(1-\alpha_{2n})(1-\beta_{2n})\Big(\langle s_{2n} - z, -z\rangle + \varepsilon\tau_{2n}||\nabla g_{2n}(s_{2n})||||z||\Big) \\ & + 2\beta_{2n}(1-\beta_{2n})\Big(\langle h_{2n} - z, -z\rangle + \theta_{2n}||\nabla g_{2n}(h_{2n})||||z|| \\ & + \vartheta||\varsigma_{2n}^{(1)}d_{2n} - \varsigma_{2n}^{(2)}x_{2n}||||z||\Big) - (1-\delta_{2n})||s_{2n} - m_{2n}||^2 \\ & + \vartheta(1-\beta_{2n})\Big(2\left\langle A_{g_{2n}}h_{2n} - z, \varsigma_{2n}^{(1)}d_{2n} - \varsigma_{2n}^{(2)}x_{2n}\right\rangle \\ & + ||\varsigma_{2n}^{(1)}d_{2n} - \varsigma_{2n}^{(2)}x_{2n}||^2\Big)\Bigg). \end{aligned} \tag{3.24}$$

Following the same lines of proof of (3.18), we equivalently obtain from (3.17) and (3.24) that

$$\begin{aligned}||u_{2n+1}-z||^2 \quad\leq\quad & (1-\delta_{2n}\beta_{2n})||s_{2n}-z||^2-\delta_{2n}(1+\lambda_{2n+1})\Big((1-\delta_{2n})\\ & -\lambda_{2n+1}\delta_{2n}\Big)||s_{2n}-m_{2n}||^2+\delta_{2n}(1+\lambda_{2n+1})\Bigg((\beta_{2n}^2+(1-\beta_{2n})\alpha_{2n}^2)||z||^2\\ & +2\alpha_{2n}(1-\alpha_{2n})(1-\beta_{2n})\Big(\langle s_{2n}-z,-z\rangle+\varepsilon\tau_{2n}||\nabla g_{2n}(s_{2n})||||z||\Big)\\ & +2\beta_{2n}(1-\beta_{2n})\Big(\langle h_{2n}-z,-z\rangle+\theta_{2n}||\nabla g_{2n}(h_{2n})||||z||\\ & +\vartheta||\varsigma_{2n}^{(1)}d_{2n}-\varsigma_{2n}^{(2)}x_{2n}||||z||\Big)\\ & +\vartheta(1-\beta_{2n})\Big(2\left\langle A_{g_{2n}}h_{2n}-z,\varsigma_{2n}^{(1)}d_{2n}-\varsigma_{2n}^{(2)}x_{2n}\right\rangle\\ & +||\varsigma_{2n}^{(1)}d_{2n}-\varsigma_{2n}^{(2)}x_{2n}||^2\Big)\Bigg). \end{aligned}\tag{3.25}$$

We similarly obtain from (3.19), (3.23) (3.24) and (3.25) that

$$\begin{aligned}||s_{2n+2}-z||^2 \quad\leq\quad & (1-\delta_{2n}\beta_{2n})||s_{2n}-z||^2-\delta_{2n}(1+\lambda_{2n+1})\Big((1-\delta_{2n})\\ & -\lambda_{2n+1}\delta_{2n}(1+\delta_{2n})\Big)||s_{2n}-m_{2n}||^2+\delta_{2n}(1+\lambda_{2n+1})\Bigg(2\big(\beta_{2n}^2\\ & +(1-\beta_{2n})\alpha_{2n}^2\big)||z||^2+2\alpha_{2n}\Big((1-\alpha_{2n})(1-\beta_{2n})\big(\langle s_{2n}-z,-z\rangle\\ & +\varepsilon\tau_{2n}||\nabla g_{2n}(s_{2n})||||z||\big)\\ & +\big(||u_{2n+1}-z||^2+\varepsilon\tau_{2n+1}||\nabla g_{2n+1}(s_{2n+1})||\big)||z||^2\Big)\\ & +2\beta_{2n}\Big((1-\beta_{2n})\big(\langle h_{2n}-z,-z\rangle+\theta_{2n}||\nabla g_{2n}(h_{2n})||||z||\\ & +\vartheta||\varsigma_{2n}^{(1)}d_{2n}-\varsigma_{2n}^{(2)}x_{2n}||||z||\big)+\theta_{2n+1}||\nabla g_{2n+1}(h_{2n+1})||||z||\\ & +\vartheta||\varsigma_{2n+1}^{(1)}d_{2n+1}-\varsigma_{2n+1}^{(2)}x_{2n+1}||||z||\Big)\\ & +\vartheta\Big((1-\beta_{2n})\Big(2\left\langle A_{g_{2n}}h_{2n}-z,\varsigma_{2n}^{(1)}d_{2n}-\varsigma_{2n}^{(2)}x_{2n}\right\rangle\\ & +||\varsigma_{2n}^{(1)}d_{2n}-\varsigma_{2n}^{(2)}x_{2n}||^2\Big)+||\varsigma_{2n+1}^{(1)}d_{2n+1}-\varsigma_{2n+1}^{(2)}x_{2n+1}||^2\\ & +2\left\langle A_{g_{2n+1}}h_{2n+1}-z,\varsigma_{2n+1}^{(1)}d_{2n+1}-\varsigma_{2n+1}^{(2)}x_{2n+1}\right\rangle\Big)\Bigg)\\ & +2\delta_{2n+1}\beta_{2n+1}(1-\beta_{2n+1})\left\langle h_{2n+1}-z,-z\right\rangle\\ \leq\quad & (1-\delta_{2n}\beta_{2n})||s_{2n}-z||^2+\delta_{2n}\beta_{2n}\Gamma_{2n}, \end{aligned}\tag{3.26}$$

where
$\Gamma_{2n}=\frac{(1+\lambda_{2n+1})}{\beta_n}\Bigg(2\big(\beta_{2n}^2+(1-\beta_{2n})\alpha_{2n}^2\big)||z||^2+2\alpha_{2n}\Big((1-\alpha_{2n})(1-\beta_{2n})\ \big(\langle s_{2n}-z,-z\rangle+$ $\varepsilon\tau_{2n}||\nabla g_{2n}(s_{2n})||||z||\big)+\big(||u_{2n+1}-z||^2+\varepsilon\tau_{2n+1}||\nabla g_{2n+1}(s_{2n+1})||\big)||z||\Big)$ $+2\beta_{2n}\Big((1-\beta_{2n})$ $\big(\langle h_{2n}-z,-z\rangle+\theta_{2n}||\nabla g_{2n}(h_{2n})||||z||+\vartheta||\varsigma_{2n}^{(1)}d_{2n}-\varsigma_{2n}^{(2)}x_{2n}||||z||\big)$ $+\theta_{2n+1}||\nabla g_{2n+1}(h_{2n+1})||$ $||z||+\vartheta||\varsigma_{2n+1}^{(1)}d_{2n+1}-\varsigma_{2n+1}^{(2)}x_{2n+1}||||z||\Big)+\vartheta\Big((1-\beta_{2n})\ \Big(2\left\langle A_{g_{2n}}h_{2n}-z,\varsigma_{2n}^{(1)}d_{2n}-\varsigma_{2n}^{(2)}x_{2n}\right\rangle+$

$$||\varsigma_{2n}^{(1)}d_{2n}-\varsigma_{2n}^{(2)}x_{2n}||^2\Big)+2\left\langle A_{g_{2n+1}}h_{2n+1}-z,\varsigma_{2n+1}^{(1)}d_{2n+1}-\varsigma_{2n+1}^{(2)}x_{2n+1}\right\rangle$$
$$+||\varsigma_{2n+1}^{(1)}d_{2n+1}-\varsigma_{2n+1}^{(2)}x_{2n+1}||^2\Big)\Big)+2\delta_{2n+1}\tfrac{\beta_{2n+1}}{\beta_{2n}}(1-\beta_{2n+1})\left\langle h_{2n+1}-z,-z\right\rangle. \qquad \blacksquare$$

Next, we apply Lemma 2.5 to establish the strong convergence of an even subsequence $\{s_{2n}\}$ of a sequence $\{s_n\}$ produced by Algorithm 2 to the minimum-norm solution of problem (1.1) (i.e., $z^*=P_\Omega 0$).

Theorem 3.7. *Let Assumptions 3.1-3.2 hold, and $\{s_{2n}\}$ be an even subsequence of the sequence generated by Algorithm 2. Then $\{s_{2n}\}$ converges strongly to a point $z^*\in\Omega$, where $z^*=P_\Omega 0$.*

Proof. Without loss of generality, using Assumption 3.2, we can assume the existence of $p,\ q,\ t>0$ such that $\forall\ n\geq 1$,

$$2\varepsilon\eta_n(2-\varepsilon\eta_n)\delta_n(1-\beta_n)(1-\alpha_n)\geq p,$$
$$\pi_n(4-\pi_n)\delta_n(1-\beta_n)\geq q$$

and

$$\delta_n(1+\lambda_{n+1})((1-\delta_n)-\lambda_{n+1}\delta_n(1+\delta_n))\geq t.$$

Consequently, we respectively obtain from (3.20) and (3.26) that

$$||s_{2n+2}-z^*||^2\leq||s_{2n}-z^*||^2-\chi_{2n}+\omega_{2n},$$

and

$$||s_{2n+2}-z^*||^2\leq(1-\delta_{2n}\beta_{2n})||s_{2n}-z^*||^2+\delta_{2n}\beta_{2n}\Gamma_{2n}, \tag{3.27}$$

where

$$\chi_{2n}=p\left(\frac{g_{2n}^2(s_{2n})}{||\nabla g_{2n}(s_{2n})||^2}+\frac{g_{2n+1}^2(s_{2n+1})}{||\nabla g_{2n+1}(s_{2n+1})||^2}\right)+t||m_{2n}-s_{2n}||^2$$
$$+q\left(\frac{g_{2n}^2(h_{2n})}{||\nabla g_{2n}(h_{2n})||^2}+\frac{g_{2n+1}^2(h_{2n+1})}{||\nabla g_{2n+1}(h_{2n+1})||^2}\right).$$

and

$$\omega_{2n}=\delta_{2n}(1+\lambda_{2n+1})\big(2(\beta_{2n}+\alpha_{2n})||z||^2+\Phi_{2n}\big).$$

It is easily seen from conditions (B3) and (B4) of Assumption 3.2 that $\lim_{n\to\infty}\omega_{2n}=0$. It therefore remains to prove that for every subsequence $\{\chi_{2n_r}\}$ of $\{\chi_{2n}\}$, the following holds.

$$\lim_{r\to\infty}\chi_{2n_r}=0\Rightarrow\limsup_{r\to\infty}\Gamma_{2n_r}\leq 0.$$

Now, suppose that $\{\chi_{2n_r}\}$ is a subsequence of $\{\chi_{2n}\}$ such that $\lim_{r\to\infty}\chi_{2n_r}=0$, then, the conditions of Assumption 3.2 imply that

$$\lim_{r\to\infty}\frac{g_{2n_r}^2(s_{2n_r})}{||\nabla g_{2n_r}(s_{2n_r}))||^2}=0,\qquad\lim_{r\to\infty}\frac{g_{2n_r+1}^2(s_{2n_r+1})}{||\nabla g_{2n_r+1}(s_{2n_r+1}))||^2}=0,$$

$$\lim_{r\to\infty}\frac{g_{2n_r}^2(h_{2n_r})}{||\nabla g_{2n_r}(h_{2n_r}))||^2}=0,\qquad\lim_{r\to\infty}\frac{g_{2n_r+1}^2(h_{2n_r+1})}{||\nabla g_{2n_r+1}(h_{2n_r+1}))||^2}=0,$$

$$\lim_{r\to\infty}||m_{2n_r}-s_{2n_r}||=0. \tag{3.28}$$

Now, since ∇g_{2n_r} and ∇g_{2n_r+1} are K-Lipschitz continuous with $K = ||\mathcal{A}||^2$, then for any $z \in \Omega$ and following similar argument on $\nabla g_{2n_r}(z)$ and $\nabla g_{2n_r+1}(z)$ as in (3.5), one obtains that

$$||\nabla g_{2n_r}(s_{2n_r})|| \leq ||\mathcal{A}||^2||s_{2n_r} - z||, \quad ||\nabla g_{2n_r+1}(s_{2n_r+1})|| \leq ||\mathcal{A}||^2||s_{2n_r+1} - z||,$$

$$||\nabla g_{2n_r}(h_{2n_r})|| \leq ||\mathcal{A}||^2||h_{2n_r} - z||, \quad ||\nabla g_{2n_r+1}(h_{2n_r+1})|| \leq ||\mathcal{A}||^2||h_{2n_r+1} - z||. \tag{3.29}$$

Thus, in view of the boundedness of $\{||s_{2n_r} - z||\}$ and (3.28), we obtain that $\{||s_{2n_r+1} - z||\}$, $\{||h_{2n_r} - z||\}$ and $\{||h_{2n_r+1} - z||\}$ are bounded. Consequently, the inequalities (3.29) imply that the subsequences $\{\nabla g_{2n_r}(s_{2n_r})\}$, $\{\nabla g_{2n_r}(h_{2n_r})\}$, $\{\nabla g_{2n_r+1}(s_{2n_r+1})\}$ and $\{\nabla g_{2n_r+1}(h_{2n_r+1})\}$ are bounded. Therefore, combining with (3.28), we deduce that

$$\lim_{r\to\infty} g_{2n_r}(s_{2n_r}) = 0 \Leftrightarrow \lim_{r\to\infty} ||(I - P_{\mathcal{Q}_{2n_r}})\mathcal{A}s_{2n_r}||^2 = 0,$$

$$\lim_{r\to\infty} g_{2n_r+1}(s_{2n_r+1}) = 0 \Leftrightarrow \lim_{r\to\infty} ||(I - P_{\mathcal{Q}_{2n_r+1}})\mathcal{A}s_{2n_r+1}||^2 = 0,$$

$$\lim_{r\to\infty} g_{2n_r}(h_{2n_r}) = 0 \Leftrightarrow \lim_{r\to\infty} ||(I - P_{\mathcal{Q}_{2n_r}})\mathcal{A}h_{2n_r}||^2 = 0$$

and

$$\lim_{r\to\infty} g_{2n_r+1}(h_{2n_r+1}) = 0 \Leftrightarrow \lim_{r\to\infty} ||(I - P_{\mathcal{Q}_{2n_r+1}})\mathcal{A}h_{2n_r+1}||^2 = 0. \tag{3.30}$$

Now, based on the fact that the even subsequence $\{s_{2n}\}$ of $\{s_n\}$ is bounded, the existence of a subsequence $\{s_{2n_r}\}$ of $\{s_{2n}\}$ converging weakly to a point s^* is guaranteed. The condition (A3) of Assumption 3.1 implies the existence of a constant $\rho > 0$ such that $||\varphi_{2n_r}|| \leq \rho$. Consequently, by the definition of $\mathcal{Q}_{2n_r}$, $P_{\mathcal{Q}_{2n_r}}\mathcal{A}s_{2n_r} \in \mathcal{Q}_{2n_r}$, and (3.30), we have

$$\begin{aligned} k(\mathcal{A}s_{2n_r}) &\leq \langle \varphi_{2n_r}, \mathcal{A}s_{2n_r} - P_{\mathcal{Q}_{2n_r}}\mathcal{A}s_{2n_r} \rangle \\ &\leq ||\varphi_{2n_r}||||\mathcal{A}s_{2n_r} - P_{\mathcal{Q}_{2n_r}}\mathcal{A}s_{2n_r}|| \\ &\leq \rho||(I - P_{\mathcal{Q}_{2n_r}})\mathcal{A}s_{2n_r}||^2 \to 0 \; as \; r \to \infty. \end{aligned} \tag{3.31}$$

It is therefore not difficult to see from the weakly lower semicontuinity of k and inequality (3.31) that

$$k(\mathcal{A}s^*) \leq \liminf_{r\to\infty} k(\mathcal{A}s_{2n_r}) \leq 0, \tag{3.32}$$

implying that $\mathcal{A}s^* \in \mathcal{Q}$.

In a similar fashion, the boundedness of ∂l on bounded sets also implies the existence of $\sigma > 0$, such that $||\phi_{2n_r}|| \leq \sigma$. From the definition of $\mathcal{C}_{2n_r}$, $m_{2n_r} \in \mathcal{C}_{2n_r}$ and (3.28), we see that

$$\begin{aligned} l(s_{2n_r}) &\leq \langle \phi_{2n_r}, s_{2n_r} - m_{2n_r} \rangle \\ &\leq ||\phi_{2n_r}||||s_{2n_r} - m_{2n_r}|| \\ &\leq \sigma||s_{2n_r} - m_{2n_r}|| \to 0 \; as \; r \to \infty. \end{aligned} \tag{3.33}$$

Using similar technique followed in establishing (3.32), one sees that $l(s^*) \leq 0$, showing that $s^* \in \mathcal{C}$. Then, the conclusion that $s^* \in \Omega$ is reached. Which generally implies that $\omega_w(s_{2n}) \in \Omega$ since the choice of s^* was arbitrarily.

The following is also obtainable by combining (3.17), (3.28) and condition (B1)

$$\lim_{r\to\infty} ||s_{2n_r+1} - s_{2n_r}|| = 0. \tag{3.34}$$

In a similar fashion, combining the condition (B1), (3.2), (3.28) and (3.30), one sees that

$$||h_{2n_r} - s_{2n_r}||^2 \leq \alpha_{2n_r}||s_{2n_r}||^2 + \varepsilon^2\tau_{2n_r}^2||\nabla g_{2n_r}(s_{2n_r})||^2 \to 0 \ as \ r \to \infty \tag{3.35}$$

and

$$\begin{aligned} ||h_{2n_r+1} - s_{2n_r+1}||^2 &\leq (1-\alpha_{2n_r+1})||u_{2n_r+1} - s_{2n_r+1} - \varepsilon\tau_{2n_r+1}\nabla g_{2n_r+1}(s_{2n_r+1})||^2 \\ &\quad + \alpha_{2n_r+1}||s_{2n_r+1}||^2 \\ &\leq \alpha_{2n_r+1}||s_{2n_r+1}||^2 + 2\varepsilon^2\eta_{2n_r+1}^2||\nabla g_{2n_r+1}(s_{2n_r+1})||^2 \\ &\quad + 2\delta_{2n_r}^2\lambda_{2n_r+1}^2||m_{2n_r} - s_{2n_r}||^2 \to 0 \ as \ n \to \infty. \end{aligned} \tag{3.36}$$

By (3.34), (3.35), (3.36) and the metric projection property described in (2.6), we obtain that

$$\limsup_{r\to\infty} \langle s_{2n_r} - z, -z\rangle = \max_{s^*\in\omega_w(s_{2n_r})} \langle s^* - z, -z\rangle \leq 0,$$
$$\limsup_{r\to\infty} \langle h_{2n_r} - z, -z\rangle = \max_{s^*\in\omega_w(s_{2n})} \langle s^* - z, -z\rangle \leq 0$$

and

$$\limsup_{r\to\infty} \langle h_{2n_r+1} - z, -z\rangle = \max_{s^*\in\omega_w(s_{2n})} \langle s^* - z, -z\rangle \leq 0. \tag{3.37}$$

Combining the results in (3.28), (3.37) and the conditions of Assumption 3.2, we have $\limsup_{r\to\infty} \Gamma_{2n_r} \leq 0$. We therefore observe from Lemma 2.5 that $\lim_{n\to\infty} ||s_{2n} - z^*|| = 0$ and hence, $s_{2n} \to z^* = P_\Omega 0 \ as \ n \to \infty$.

Finally, using the fact that $\lim_{n\to\infty} ||s_{2n} - z^*|| = 0$ and (3.34), it is not difficult to see that $\lim_{n\to\infty} ||s_{2n+1} - z^*|| = 0$. We therefore conclude that the odd subsequence $\{s_{2n+1}\}$ of $\{s_n\}$ produced by Algorithm 2 converges strongly to the same point $z^* \in \Omega$. Hence the whole sequence $\{s_n\}$ generated by Algorithm 2 converges strongly to $z^* \in \Omega$. This complete the proof. ∎

4. Applications

In this part, we perform some experiments on Wisconsin breast cancer dataset, which is a real-world classification data set to study the performance of the algorithm, that is SCAAiRA. We used a Matlab R2023b in a PC with 12th Gen Intel(R) Core(TM)i5-124P 1.70 GHz processor and 16.0GB RAM for all the experiments.

For these experiments, we consider an efficient learning algorithm called extreme learning machine ELM for single-hidden layer feedforward neural networks SLFNs [23], and set $\mathcal{M} = \{(s_j, y_j) \in \mathbb{R}^k \times \mathbb{R}^m, \ j = 1, 2, \ldots, \mathcal{K}\}$ as a $\mathcal{K}$ distinct training data points set, where for each input point $s_j = \left[s_{j1}, s_{j2}, \ldots, s_{jk}\right]^T$, $y_j = \left[y_{j1}, y_{j2}, \ldots, y_{jm}\right]^T$ is its corresponding target. The following is the output function of a SLFNs with $\mathcal{N}$ number of nodes in the hidden layer.

$$g_j = \sum_{i=1}^{\mathcal{N}} \beta_i f_i(s_j), \text{ for } j = 1, 2, \ldots, \mathcal{K}, \tag{4.1}$$

where $f_i(s_j) = \mathcal{F}\Big(\langle \omega_i, s_j\rangle + b_i\Big)$, $\mathcal{F}$ is an activation function, $\omega_i = (\omega_{i1}, \omega_{i2} \ldots, \omega_{ik})^T$ and $\beta_i = (\beta_{i1}, \beta_{i2}, \ldots, \beta_{im})^T$ are the input and output weight vectors, respectively and b_i is a bias. To train a SLFNs is to solve the linear system

$$\mathsf{H}\beta = \mathsf{T}, \tag{4.2}$$

where the hidden layer output matrix H of order $\mathcal{K} \times \mathcal{N}$ is given by

$$\mathsf{H} = \left[f_1(s), f_2(s), \ldots, f_{\mathcal{N}}(s)\right],$$

$\beta = (\beta_1, \beta_2, \ldots, \beta_{\mathcal{N}})^T$ and $\mathsf{T} = (y_1, y_2, \ldots, y_{\mathcal{K}})^T$ are the output weights and the target data matrices, respectively and the i^{th} column of H is the i^{th} hidden node output based on $s_1, s_2 \ldots, s_{\mathcal{K}}$, which is defined by $f_i(s) = \big[f_i(s_1), f_i(s_2), \ldots, f_i(s_{\mathcal{K}})\big]^T$. To use ELM to solve problem (4.2) is nothing but to find an optimal output weight $\hat{\beta} = \mathsf{H}^{\dagger}\mathsf{T}$, where $\mathsf{H}^{\dagger}$ represents the Moore-Penrose generalized inverse of H [35].

In view of the sparsity of the output weight parameter β for some high-dimensional data, Cao et al. [7] introduced an ℓ_1-regularization approach to solve problem (4.2) based on the following Lasso model [41].

$$\min_{\beta \in \mathbb{R}^{\mathcal{N}\times m}} \Big\{\frac{1}{2}||\mathsf{T} - \mathsf{H}\beta||_2^2 \;:\; ||\beta||_1 \le \mu\Big\}, \tag{4.3}$$

where $\mu > 0$ is the regularization parameter. Ye et al. [46] unified the ℓ_1 and ℓ_2 penalties and introduced a model called the ℓ_1-ℓ_2 hybrid regularization model, which improved the accuracy, sparsity and stability in the prediction process. Their model is given by

$$\min_{\beta \in \mathbb{R}^{\mathcal{N}\times m}} \Big\{\frac{1}{2}||\mathsf{T} - \mathsf{H}\beta||_2^2 \;:\; \lambda||\beta||_1 + \gamma||\beta||_2^2 \le \mu\Big\}, \tag{4.4}$$

where $\lambda,\ \gamma \ge 0$ and $\mu > 0$ are the regularization parameters. Recently, Suantai et al. [38] considered the problem (4.3) in the form of problem (1.1) by defining $\mathcal{C} = \{\beta \in \mathbb{R}^{\mathcal{N}\times m} : ||\beta||_1 \le \mu\}$, $\mathcal{Q} = \{\mathsf{T}\} \subseteq \mathbb{R}^{\mathcal{K}\times m}$, $c(\beta) = ||\beta||_1 - \mu$, $q(x) = \frac{1}{2}||x - \mathsf{T}||^2$ and define $\mathcal{C}_n$, $\mathcal{Q}_n$ and g_n as in (1.4) and (1.6), respectively and applied their inertial relaxed $\mathcal{CQ}$ algorithm to solve the problem (4.2) based on the model (4.3).

Inspired by the sparsity, stability and generalization performance of (4.4), we similarly transform the problem (4.4) into problem (1.1) by taking $\mathcal{C} = \{\beta \in \mathbb{R}^{\mathcal{N}\times m} : \lambda||\beta||_1 + \gamma||\beta||_2^2 \le \mu\}$, $\mathcal{Q} = \{\mathsf{T}\} \subseteq \mathbb{R}^{\mathcal{K}\times m}$ and $q(x) = \frac{1}{2}||x - \mathsf{T}||^2$. Moreover, it is not difficult to see that

the function $c(\beta) = \lambda||\beta||_1 + \gamma||\beta||_2^2 - \mu$ is convex. We consider $\mathcal{C}_n$, $\mathcal{Q}_n$, g_n and its gradient ∇g_n as described in (1.4) and (1.6), respectively. So, our algorithm SCAAiRA can be used to solve problem (4.2) based on the both models (4.3) and (4.4).

To find out the performance of SCAAiRA, we used it to solve problem (4.2) based on the models (4.3) and (4.4), which we respectively abbreviated as ℓ_1-SCAAiRA and ℓ_1-ℓ_2-SCAAiRA. We compare their results with Algorithm 1 in [11], abbreviated in this work as ℓ_1-CZWC and ℓ_1-ℓ_2-CZWC. We carried out the experiments on an interesting type of real-world dataset called Wisconsin breast cancer dataset [42]. The instances of the dataset are classified into two, which include 357 benign instances and 212 malignant instances. We find more details of this dataset from [42].

We conducted all the experiments after normalizing the original data by taking $\bar{s}_{ji} = \frac{s_{ji}-s_i^{\min}}{s_i^{\max}-s_i^{\min}}$, where $s_i^{\max} = \max\limits_{j=1,2,\ldots,\mathcal{M}}(s_{ji})$ and $s_i^{\min} = \min\limits_{j=1,2,\ldots,\mathcal{M}}(s_{ji})$ represent the maximum and minimum of i^{th} attributes over all the input data points s_j respectively and $\bar{s}_{ji}$ represents the normalized value of s_{ji}. We use 70% of the dataset for training, 30% for testing and two activation functions, namely, the Sigmoid and Radbas avtivation functions for $\mathcal{N} = 100,\ 300,\ 500$ and 700. We set $s_0 = s_1 = \text{randn}(\mathcal{N}, m),\ x_n = \text{ones}(\mathcal{N}, m)$ and the following for the parameters.

- For ℓ_1-SCAAiRA and ℓ_1-ℓ_2-SCAAiRA, we set $\varepsilon = 1.7,\ \vartheta = 3,\ \lambda_n = \frac{2n+1}{10^5 n^5+1},\ \beta_n = \frac{1}{10^4 n+1},\ \alpha_n = \frac{1}{(n+1)^5},\ \delta_n = \frac{10n}{10n+1},\ \pi_n = 2,\ \eta_n = \frac{2}{\varepsilon} - 0.001$ and $\varsigma_n^{(i)} = \frac{1}{10n^5+1}$, for $i = 1, 2$.
- For ℓ_1-CZWC and ℓ_1-ℓ_2-CZWC, we set $\eta_1 = 0.01,\ \gamma = 0.8,\ \varepsilon_1 = 0.3,\ \delta_n = \frac{1}{n+1},\ \varepsilon_n = \frac{1}{n^3},\ \eta_n = 0.7,\ \mu = 0.6,\ \varsigma_n = 0.7\delta^2$ and $f(x) = 0.9x$.

We compute the prediction accuracy for each algorithm by the following relation.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \times 100\%, \tag{4.5}$$

where TP := True positive, TN := True negative, FP = False positive and FN = False negative, and estimate their averages as well as their standard deviations (SDs). These metrics, together with the number of iterations denoted by "Iter." and the execution times in seconds denoted by "Time" are used to investigate the performance of each algorithm. In all the experiments, we set $||x_{n+1} - x_n|| < 10^{-5}$ and the maximum number of iterations as 500 to terminate the iterations for all the algorithms. As shown in Tables 1 and 2, we set the parameters μ, λ and γ according to the number of hidden neurons and the activation function. The training and testing accuracies and times as well as the number of iterations of all the algorithms are reported in Tables 1 and 2. Additionally, we displayed the corresponding results of all the algorithms in Figure 1. We further display the average accuracies and SDs of the accuracies of all the algorithms for each activation function in Table 3.

TABLE 1. Performance results of all the algorithms for $\mathcal{N}$ = 100, 300, 500 and 700 using Sigmoid activation function.

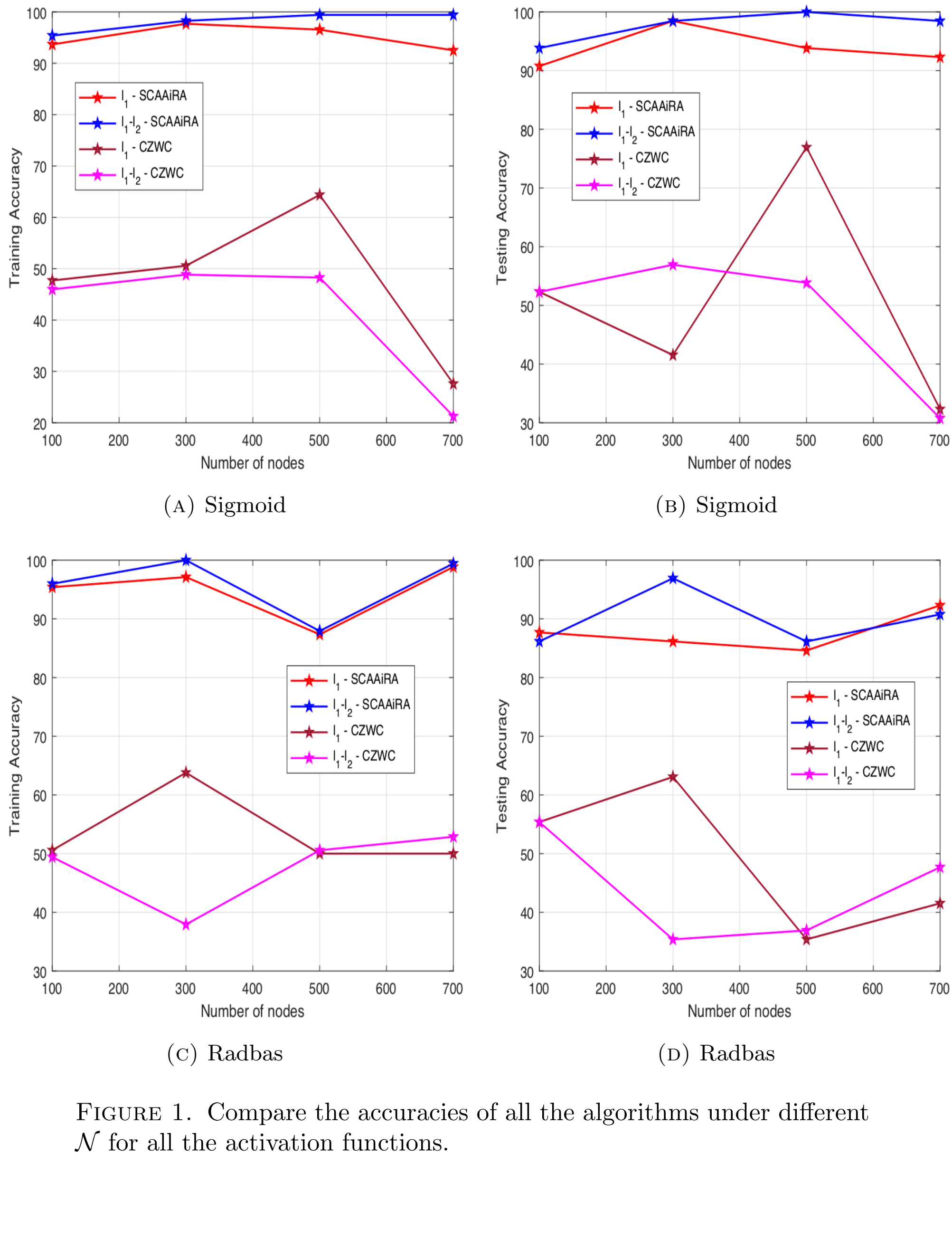
Act. Funct.		Sigmoid					
				Time (s)		Accuracy (%)	
Nodes	Algorithms	μ, λ, γ	Iter.	Training	Testing	Training	Testing
$\mathcal{N}=100$	ℓ_1 - SCAAiRA	5, 1.0001, 0.00505	51	0.2347	0.0017	93.6782	90.7692
	ℓ_1-ℓ_2 - SCAAiRA		55	0.3041	0.0017	**95.4023**	**93.8462**
	ℓ_1 - CZWC		77	0.3245	0.0022	47.7011	52.3077
	ℓ_1-ℓ_2 - CZWC		78	0.3572	0.0021	45.977	52.3077
$\mathcal{N}=300$	ℓ_1 - SCAAiRA	5, 1.001, 0.0001	28	0.1664	0.0021	97.7011	**98.4615**
	ℓ_1-ℓ_2 - SCAAiRA		30	0.1952	0.0016	**98.2759**	**98.4615**
	ℓ_1 - CZWC		73	0.3739	0.0021	50.5747	41.5385
	ℓ_1-ℓ_2 - CZWC		86	0.4662	0.0024	48.8506	56.9231
$\mathcal{N}=500$	ℓ_1 - SCAAiRA	21, 1.001, 0.002	43	0.2296	0.003	96.5517	93.8462
	ℓ_1-ℓ_2 - SCAAiRA		39	0.2712	0.0018	**99.4253**	**100**
	ℓ_1 - CZWC		78	0.5209	0.0026	64.3678	76.9231
	ℓ_1-ℓ_2 - CZWC		66	0.4583	0.0019	48.2759	53.8462
$\mathcal{N}=700$	ℓ_1 - SCAAiRA	12.2, 0.999, 0.1	60	0.4073	0.0033	92.5287	92.3077
	ℓ_1-ℓ_2 - SCAAiRA		44	0.3915	0.0026	**99.4253**	**98.4615**
	ℓ_1 - CZWC		500	3.1394	0.0026	27.5862	32.3077
	ℓ_1-ℓ_2 - CZWC		500	3.2336	0.0041	21.2644	30.7692

TABLE 2. Performance results of all the algorithms for $\mathcal{N}$ = 100, 300, 500 and 700 using Radbas activation function.

Act. Funct.		Radbas					
				Time (s)		Accuracy (%)	
Nodes	Algorithms	μ, λ, γ	Iter.	Training	Testing	Training	Testing
$\mathcal{N} = 100$	ℓ_1 - SCAAiRA	12.2, 0.9999, 0.0001	199	0.7491	0.0017	<u>95.4023</u>	**87.6923**
	ℓ_1-ℓ_2 - SCAAiRA		193	0.9254	0.0016	**95.977**	<u>86.1538</u>
	ℓ_1 - CZWC		89	0.4271	0.0018	50.5747	55.3846
	ℓ_1-ℓ_2 - CZWC		86	0.4579	0.0016	49.4253	55.3846
$\mathcal{N} = 300$	ℓ_1 - SCAAiRA	12.2, 0.9991, 0.0001	31	0.2396	0.0018	<u>97.1264</u>	<u>86.1538</u>
	ℓ_1-ℓ_2 - SCAAiRA		47	0.2653	0.0018	**100**	**96.9231**
	ℓ_1 - CZWC		63	0.3407	0.0028	63.7931	63.0769
	ℓ_1-ℓ_2 - CZWC		58	0.3592	0.0019	37.931	35.3846
$\mathcal{N} = 500$	ℓ_1 - SCAAiRA	37.9651, 1.001, 0.003	30	0.2025	0.002	<u>87.3563</u>	<u>84.6154</u>
	ℓ_1-ℓ_2 - SCAAiRA		17	0.1739	0.0023	**87.931**	**86.1538**
	ℓ_1 - CZWC		111	0.5944	0.0019	50	35.3846
	ℓ_1-ℓ_2 - CZWC		112	0.6184	0.0021	50.5747	36.9231
$\mathcal{N} = 700$	ℓ_1 - SCAAiRA	15.5, 0.94, 0.0001	55	0.4513	0.0029	<u>98.8506</u>	**92.3077**
	ℓ_1-ℓ_2 - SCAAiRA		80	0.5654	0.0029	**99.4253**	<u>90.7692</u>
	ℓ_1 - CZWC		500	3.3131	0.0049	50	41.5385
	ℓ_1-ℓ_2 - CZWC		500	3.4549	0.0035	52.8736	47.6923

TABLE 3. Averages and SDs of the accuracies of all the algorithms for all the activation functions.

		Aver. Accuracy (%)		SDs	
Act. Funct.	Algorithms	Training	Testing	Training	Testing
Sigmoid	ℓ_1-SCAAiRA	<u>95.1149</u>	<u>93.8462</u>	<u>2.4156</u>	<u>3.3234</u>
	ℓ_1-ℓ_2 - SCAAiRA	**98.1322**	**97.6923**	**1.8989**	**2.6647**
	ℓ_1 - CZWC	47.5574	50.7692	15.1719	19.2564
	ℓ_1-ℓ_2 - CZWC	41.092	48.4616	13.2766	11.95
Radbas	ℓ_1-SCAAiRA	<u>94.6839</u>	<u>87.6923</u>	**5.0839**	**3.3235**
	ℓ_1-ℓ_2 - SCAAiRA	**95.8333**	**90.00**	<u>5.5597</u>	<u>5.1025</u>
	ℓ_1 - CZWC	53.5919	48.8462	6.8062	12.6475
	ℓ_1-ℓ_2 - CZWC	47.7012	43.8461	6.6693	9.442



(A) Sigmoid (B) Sigmoid

(C) Radbas (D) Radbas

FIGURE 1. Compare the accuracies of all the algorithms under different $\mathcal{N}$ for all the activation functions.

Remark 4.1. Based on the reported performance results of all the algorithms in Tables 1, 2 and 3, and Figure 1, we conclude the following:

(i) It is easy to see that the proposed algorithms ℓ_1-SCAAiRA and ℓ_1-ℓ_2-SCAAiRA considerably achieve higher training and testing accuracies in all the experiments with fewer iterations and shorter training times in most cases than ℓ_1-CZWC and ℓ_1-ℓ_2-CZWC. Additionally, the SDs of both the training and testing accuracies of ℓ_1-SCAAiRA and ℓ_1-ℓ_2-SCAAiRA are extremely smaller than those of ℓ_1-CZWC and ℓ_1-ℓ_2-CZWC for the two activation functions. These demonstrate that ℓ_1-SCAAiRA and ℓ_1-ℓ_2-SCAAiRA achieve better stability and generalization performance in the experiments.

(ii) It is also observed that due to the presence of the ℓ_2 penalty in ℓ_1-ℓ_2-SCAAiRA, it achieves higher training and testing accuracies in most results than its corresponding ℓ_1-SCAAiRA, which demonstrate its ability to achieve better generalization performance.

5. Conclusion

This work introduces a computationally efficient algorithm called strong convergence accelerated alternated inertial relaxed algorithm (SCAAiRA) with three-term conjugate gradient-like direction in finite-dimensional real Hilbert spaces. The performance results of the proposed algorithm in solving classification problems based on the extreme learning machine (ELM) for the Wisconsin breast cancer dataset are analyzed and compared with the relaxed algorithm in [11]. In all the experiments, the numerical results show that the proposed algorithm is robust, computationally efficient and achieves better generalization performance and stability than the algorithms in [11]. It also illustrates that the proposed algorithm achieves better accuracy based on the ℓ_1-ℓ_2 hybrid regularization model with ℓ_2 penalty.

Acknowledgments

The authors acknowledge the provided financial support by Center of Excellence in Theoretical and Computational Science (TaCS-CoE), KMUTT. Additionally, the first author acknowledged the support of Petchra Pra Jom Klao Doctoral Research Scholarship from King Mongkut's University of Technology Thonburi with Contract No. 48/2565 and Center of Excellence in Theoretical and Computational Science (TaCS-CoE), KMUTT.

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval and Consent to participate
Not Applicable.

Consent for publication
The authors agreed to publish this article in the Bangmod International Journal of Mathematical & Computational Science.

Availability of data and materials
The dataset analyzed in this study is available in https://archive.ics.uci.edu/.

Code availability
Available upon reasonable request.

Authors' contributions
All authors contributed equally in this article.

Orcid

Abdulwahab Ahmad https://orcid.org/0009-0000-0460-0204
Poom Kumam https://orcid.org/0000-0002-5463-4581

References

[1] A. Ahmad, P. Kumam, M.H. Harbau, Convergence theorems for common solutions of nonlinear problems and applications, Carpathian Journal of Mathematics 40(2) (2024) 207241. https://doi.org/10.37193/CJM.2024.02.01.

[2] A. Ahmad, P. Kumam, M.M. Yahaya, K. Sitthithakerngkiet, W. Cholamjiak, Two-step alternated inertial algorithm for split feasibility problems with some applications, Jounal of Nonlinear and Convex Analysis 25(12) (2024) 3165–3191.

[3] F. Alvarez, H. Attouch, An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping, Set-Valued Analysis 9 (2001) 3–11. https://doi.org/10.1023/A:1011253113155.

[4] H.H. Bauschke, P.L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, Springer, 2017. https://doi.org/10.1007/978-3-319-48311-5.

[5] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM Journal on Imaging Sciences 2(1) (2009) 183–202. https://doi.org/10.1137/080716542.

[6] C. Byrne, Iterative oblique projection onto convex sets and the split feasibility problem, Inverse problems 18(2) (2002) 441. https://doi.org/10.1088/0266-5611/18/2/310.

[7] F. Cao, Y. Tan, M. Cai, Sparse algorithms of random weight networks and applications, Expert Systems with Applications 41(5) (2014) 2457–2462. https://doi.org/10.1016/j.eswa.2013.09.045.

[8] L.C. Ceng, Q.H. Ansari, J.C. Yao, An extragradient method for solving split feasibility and fixed point problems, Computers & Mathematics with Applications 64(4) (2012) 633–642. https://doi.org/10.1016/j.camwa.2011.12.074.

[9] Y. Censor, T. Bortfeld, B. Martin, A. Trofimov, A unified approach for inversion problems in intensity-modulated radiation therapy, Physics in Medicine & Biology 51(10) (2006) 2353. https://doi.org/10.1088/0031-9155/51/10/001.

[10] Y. Censor, T. Elfving, A multiprojection algorithm using Bregman projections in a product space, Numerical Algorithms 8 (1994) 221–239. https://doi.org/10.1007/BF02142692.

[11] H. Che, Y. Zhuang, Y. Wang, H. Chen, A relaxed inertial and viscosity method for split feasibility problem and applications to image recovery, Journal of Global Optimization 87(2–4) (2023) 619–639. https://doi.org/10.1007/s10898-022-01246-9.

[12] P.L. Combettes, V.R. Wajs, Signal recovery by proximal forward-backward splitting. Multiscale modeling & simulation 4(4) (2005) 1168–1200. https://doi.org/10.1137/050626090.

[13] Q.L. Dong, X.H. Li, T.M. Rassias, Two projection algorithms for a class of split feasibility problems with jointly constrained Nash equilibrium models, Optimization 70(4) (2021) 871–897. https://doi.org/10.1080/02331934.2020.1753741.

[14] Q.L. Dong, L. Liu, Y. Yao, Self-adaptive projection and contraction methods with alternated inertial terms for solving the split feasibility problem, Journal of Nonlinear and Convex Analysis 23(3) (2022) 591–605.

[15] Q.L. Dong, H.b. Yuan, Accelerated Mann and CQ algorithms for finding a fixed point of a nonexpansive mapping, Fixed Point Theory and Applications 2015(1) (2015) 112. https://doi.org/10.1186/s13663-015-0374-6.

[16] A. Gibali, K. Küfer, and P. Süss. Successive linear programing approach for solving the nonlinear split feasibility problem, Journal of Nonlinear and Convex Analysis 15(2) (2014) 345–353.

[17] A. Gibali, D.T. Mai, et al. A new relaxed cq algorithm for solving split feasibility problems in hilbert spaces and its applications, Journal of Industrial & Management Optimization 15(2) (2019). https://doi.org/10.3934/jimo.2018080.

[18] K. Goebel, S. Reich, Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings, Series of Monographs and Textbooks in Pure and Applied Mathematics, Dekker, New York, 1984.

[19] M.H. Harbau, A. Ahmad, B. Ali, G.C. Ugwunnadi, Inertial residual algorithm for fixed points of finite family of strictly pseudocontractive mappings in Banach spaces, International Journal of Nonlinear Analysis and Applications 13(2) (2022) 22572269. https://doi.org/10.22075/IJNAA.2021.23301.2517.

[20] M.H. Harbau, G.C. Ugwunnadi, L.O. Jolaoso, A. Abdulwahab, Inertial accelerated algorithm for fixed point of asymptotically nonexpansive mapping in real uniformly convex Banach spaces, Axioms 10(3) (2021). https://doi.org/10.3390/axioms10030147.

[21] M. Harbaua, A. Ahmadb. Inertial hybrid s-iteration algorithm for fixed point of asymptotically nonexpansive mappings and equilibrium problems in a real Hilbert space, Journal of Nonlinear Sciences & Applications (JNSA), 15(2) (2022) 123–135. https://doi.org/10.22436/jnsa.015.02.04.

[22] S. He, C. Yang, Solving the variational inequality problem defined on intersection of finite level sets, Abstract and applied analysis 2013 (2013) 1–8. https://doi.org/10.1155/2013/942315.

[23] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks. In 2004 IEEE international joint conference on neural networks (IEEE Cat. No. 04CH37541), volume 2, pages 985–990, IEEE, 2004. https://doi.org/10.1109/IJCNN.2004.1380068.

[24] H. Iiduka, Acceleration method for convex optimization over the fixed point set of a nonexpansive mapping, Mathematical Programming 149(1–2) (2015) 131-165. https://doi.org/10.1007/s10107-013-0741-1.

[25] C. Izuchukwu, G. Ogwo, O. Mewomo, An inertial method for solving generalized split feasibility problems over the solution set of monotone variational inclusions, Optimization 71(3) (2022) 583–611. https://doi.org/10.1080/02331934.2020.1808648.

[26] A. Kiri, A. Abubakar, A family of conjugate gradient projection method for nonlinear monotone equations with applications to compressive sensing. Nonlinear Convex Analysis and Optimization 1(1) (2022) 47–65.

[27] T. Ling, X. Tong, L. Shi, Modified relaxed CQ methods for the split feasibility problems in Hilbert spaces with applications, Journal of Applied Mathematics and Computing 69(2) (2023) 3067–3094. https://doi.org/10.1007/s12190-023-01875-7.

[28] L. Liu, Q.L. Dong, S. Wang, M. T. Rassias, A viscosity iterative method with alternated inertial terms for solving the split feasibility problem. In High-Dimensional Optimization and Probability: With a View Towards Data Science, pages 359–372. Springer, 2022. https://doi.org/10.1007/978-3-031-00832-0_11.

[29] G. López, V. Martín-Márquez, F. Wang, H.K. Xu, Solving the split feasibility problem without prior knowledge of matrix norms, Inverse Problems 28(8) (2012) 085004. https://doi.org/10.1088/0266-5611/28/8/085004.

[30] D.A. Lorenz, T. Pock, An inertial forward-backward algorithm for monotone inclusions, Journal of Mathematical Imaging and Vision 51(2) (2015) 311–325. https://doi.org/10.1007/s10851-014-0523-2.

[31] Z. Mu, Y. Peng, A note on the inertial proximal point method. Statistics, Optimization & Information Computing 3(3) (2015) 241–248. https://doi.org/10.19139/soic.v3i3.124.

[32] S. Penfold, R. Zalas, M. Casiraghi, M. Brooke, Y. Censor, R. Schulte, Sparsity constrained split feasibility for dose-volume constraints in inverse planning of intensity-modulated photon or proton therapy, Physics in Medicine & Biology 62(9) (2017) 3599. https://doi.org/10.1088/1361-6560/aa602b.

[33] B.T. Polyak, Some methods of speeding up the convergence of iteration methods, Ussr computational mathematics and mathematical physics 4(5) (1964) 1–17. https://doi.org/10.1016/0041-5553(64)90137-5.

[34] D.R. Sahu, Y. J. Cho, Q.L. Dong, M. Kashyap, X. Li, Inertial relaxed CQ algorithms for solving a split feasibility problem in Hilbert spaces, Numerical Algorithms 87 (2021) 1075–1095. https://doi.org/10.1007/s11075-020-00999-2.

[35] D. Serre. Matrices, volume 216 of Graduate Texts in Mathematics. Springer-Verlag, New York, 2002. Theory and applications, Translated from the 2001 French original.

[36] Y. Shehu, Q.L. Dong, L.L. Liu, Global and linear convergence of alternated inertial methods for split feasibility problems, Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas 115(2) (2021) 1–26. https://doi.org/10.1007/s13398-020-00979-0.

[37] Y. Shehu, A. Gibali, New inertial relaxed method for solving split feasibilities, Optimization Letters 15(6) (2021) 2109–2126. https://doi.org/10.1007/s11590-020-01603-1.

[38] S. Suantai, P. Peeyada, A. Fulga, W. Cholamjiak, Heart disease detection using inertial Mann relaxed CQ algorithms for split feasibility problems, AIMS Mathematics 8(8) (2023) 18898–18918. https://doi.org/10.3934/math.2023962.

[39] A. Padcharoen, P. Kumam, P. Chaipunya, Y. Shehu. Convergence of Inertial Modified Krasnoselskii-Mann Iteration with Application to Image Recovery, Thai Journal

of Mathematics 18(1) (2020) 126–142.

[40] B. Tan, X. Qin, X. Wang, Alternated inertial algorithms for split feasibility problems, Numerical Algorithms 95(2) (2023) 773–812. https://doi.org/10.1007/s11075-023-01589-8.

[41] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B (Methodological) 58(1) (1996) 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[42] W. Wolberg, O. Mangasarian, N. Street, W. Street, Breast Cancer Wisconsin (Diagnostic), 1995.

[43] H.K. Xu, Iterative methods for the split feasibility problem in infinite-dimensional Hilbert spaces, Inverse problems 26(10) (2010) 105018. https://doi.org/10.1088/0266-5611/26/10/105018.

[44] Q. Yang, The relaxed CQ algorithm solving the split feasibility problem, Inverse problems 20(4) (2004) 1261. https://doi.org/10.1088/0266-5611/20/4/014.

[45] Q. Yang, On variable-step relaxed projection algorithm for variational inequalities, Journal of mathematical analysis and applications 302(1) (2005) 166–179. https://doi.org/10.1016/j.jmaa.2004.07.048.

[46] H. Ye, F. Cao, D. Wang, A hybrid regularization approach for random vector functional-link networks, Expert Systems with Applications 140 (2020) 112912. https://doi.org/10.1016/j.eswa.2019.112912.