JMCS
BANGMOD INTERNATIONAL JOURNAL OF MATHEMATICAL & COMPUTATIONAL SCIENCE
Volume 11 (2025)
Pages 23–50

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

A Two-Step Inertial CQ Method for Split Feasibility Problems with Applications

Abubakar Adamu[1,2,3], **Grace Nnennaya Ogwo**[4,5,*], **Chibueze Christain Okeke**[5], **Bertin Zinsou**[6]

[1] *Operational Research Center in Healthcare, Near East University, TRNC Mersin 10, Nicosia 99138, Turkey*
[2] *School of Mathematics, Chongqing Normal University, Chongqing 400047, China*
[3] *Charles Chidume Mathematics Institute, African University of Science and Technology, Abuja 900107, Nigeria*
E-mails: abubakar.adamu@neu.edu.tr, aadamu@aust.edu.ng
[4] *School of Mathematical Sciences, Zhejiang Normal University, Jinhua 321004, Peoples Republic of China*
E-mails: graceogwo@zjnu.edu.cn
[5] *School of Mathematical Sciences, University of the Witwatersrand, Johannesburg, South Africa*
E-mails: chibueze.okeke@wits.ac.za
[6] *National Institute for Theoretical and Computational Sciences (NITheCS)*
E-mails: bertin.zinsou@wits.ac.za

**Corresponding author.*

Received: 5 November 2024 / Accepted: 4 March 2025

Abstract This paper introduces an algorithm for approximating solutions of split feasibility problems by employing a two-step inertial acceleration strategy along with a self-adaptive step size. This combination enhances the convergence rate and reduces computational complexity of the proposed algorithm. The nonasymptotic $O(1/t)$ convergence rate and global convergence of the proposed method are established within the context of Euclidean spaces. The algorithm is extended to handle multiple set split feasibility problems, and a sensitivity analysis is conducted to identify optimal inertial parameter choices. Additionally, the algorithm is applied to the LASSO problem. Comparative evaluations with various algorithms from existing literature showcase the superior performance of the proposed algorithm.

MSC: 47H05, 47H10, 49J20, 47J25

Keywords: Split feasibility problems; two-step inertial technique; CQ methods; global convergence

Published online: 7 March 2025

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

Please cite this article as: A. Adamu et al., A Two-Step Inertial CQ Method for Split Feasibility Problems with Applications, Bangmod Int. J. Math. & Comp. Sci., Vol. 11 (2025), 23–50. https://doi.org/10.58715/bangmodjmcs.2025.11.2

1. Introduction

The split feasibility problem (SFP) is defined as the problem of finding a point $\hat{x} \in \mathcal{C}$ such that

$$A\hat{x} \in \mathcal{Q}, \tag{1.1}$$

where $\mathcal{C} \subseteq \mathbb{R}^k$ and $\mathcal{Q} \subseteq \mathbb{R}^m$ are nonempty, closed and convex sets, and $A : \mathbb{R}^k \to \mathbb{R}^m$ is a bounded and linear operator. We denote the solution set of problem (1.1) by Ω. For the purpose of solving inverse problems arising from phase retrievals and medical image reconstruction, Censor and Elfving [10] first presented the SFP in the setting of a finite dimensional real Hilbert space. Numerous studies have demonstrated the applicability of SFP in many fields of study, including computer tomography, image restoration, and data reduction (see [3, 11–13, 29, 30] and other references therein).

Various iterative methods for solving the SFP have been introduced and investigated by a number of researchers (see [9, 31, 38] and other references therein). The CQ algorithm, developed by Bryne [8], is a famous method for approximating solutions of the SFP (1.1). Iteratively, this algorithm generates the sequence $\{x_t\} \subset \mathbb{R}^k$:

$$x_{t+1} = P_{\mathcal{C}}(x_t - \lambda A^T(I - P_{\mathcal{Q}})Ax_t), \ \forall\ t \geq 1, \tag{1.2}$$

where $\lambda \in \left(0, \frac{2}{L}\right)$, with L being the largest eigenvalue of the matrix A^TA, $P_{\mathcal{C}}$, $P_{\mathcal{Q}}$ are the orthogonal projections onto $\mathcal{C}$ and $\mathcal{Q}$, respectively and I is the identity map. The author proved that the sequence generated by (1.2) converges to a solution of the SFP (1.1) or, more generally, to a minimizer of $\|P_{\mathcal{Q}}Ax^* - Ax^*\|$ over x^* in C. The drawback of this approach is that getting the step size λ requires the calculation of the spectral radius of the matrix A^TA or the norm estimate of the linear operator A, both of which can be challenging to compute in practice. To overcome this, Byrne [8] presented a method for estimating matrix norms (see [8], Proposition 4.1). However, the hypothesis of [8], Proposition 4.1 appears to be mechanical and may not be friendly to implement in practice. In order to overcome this drawback, using the idea of Yang [37], López et al. [20] introduced an adaptive step size which has no connection with matrix norms. They defined their stepsize λ_t as follows:

$$\lambda_t = \frac{\sigma^t F(x_t)}{\|\nabla F(x_t)\|^2}, \ t \geq 1,$$

where $\sigma_t \in (0, 4)$, $F(x_t) = \frac{1}{2}\|(I - P_{\mathcal{Q}})Ax_t\|^2$ and $\nabla F(x_t) = A^T(I - P_{\mathcal{Q}}Ax_t)$, for all $t \geq 1$. Due to the importance of the SFP, several modifications, extensions and generalization of the problem and iterative methods for solving the SFP have been proposed by several researchers (see [14, 15, 18, 19] and other references therein).

It is well-known that iterative methods for approximating solutions of SFP have slow convergence properties. In recent years, several authors have dedicated a reasonable research effort towards enhancing the convergence properties of existing iterative algorithms. One of the famous strategy for doing so is the inertial extrapolation technique which dates back to the early work of Polyak [25] in the context smooth convex minimization problems. Simply put, the inertial acceleration strategy is a procedure which involves a nonconvex combination of two previous terms to get the next iterate. For more on this technique and its applications to iterative methods for solving the SFP (1.1), interested readers may see, for example, [1, 2, 4, 6, 7, 28, 35], and the references therein.

Dang et al. [17] incorporate the inertial technique in the classical CQ algorithm and proposed a one-step inertial relaxed CQ algorithm for finding solutions of (1.1) in the setting of a real Hilbert space. Their algorithm is the following:

Algorithm 1. One-step inertial CQ method

Initialization: Let $x_0, x_1 \in \mathbb{H}$ be chosen arbitrarily.
Step 1: For $t \geq 0$, given the iterates x_{t-1} and x_t
Step 2: Compute

$$x_{t+1} = P_{\mathcal{C}_t}\big(G_t\big(x_t + \theta_t(x_t - x_{t-1})\big)\big),$$

where $G_t = (I - \lambda F_t),\ F_t = A^T(I - P_{\mathcal{Q}_t})A,\ \lambda \in (0, \frac{2}{L})$, L denotes the spectral radius of $A^T A$, $0 \leq \theta_t \leq \bar{\theta}_t$ with

$$\bar{\theta}_t := \min\left\{\theta, \big(\max\{\{t^2\|x_t - x_{t-1}\|^2,\ t^2\|x_t - x_{t-1}\|\}\big)^{-1}\right\},\ \theta \in [0, 1).$$

and C_t and Q_t are nonempty closed and convex half-spaces.
Set $t \leftarrow t + 1$, and go to **Step 2**.

The authors proved that the sequence $\{x_t\}$ generated by Algorithm 1 converges weakly to a point in Ω.

Remark 1.1. Observe that the step size λ depends on the knowledge of the spectral of $A^T A$ which requires the knowledge of the operator norm. It is well-known that computing norm of operators is not an easy task practice. It is also known that the sequences generated by inertial algorithms do not obey the Férjer monotonicity (that is, they do not satisfy $\|x_{n+1} - x\| \leq \|x_n - x\|,\ \forall x \in \Omega$) which is crucial in proving boundedness and convergence and in addition to this, some researchers have also discovered certain cases where the one-step inertial procedure fails to provide acceleration see, e.g., [27].

To address some of the points raised in the foregoing remark, some authors discovered that considering the alternating inertial technique solves the problem of Férjer monotonicity of the sequence. In fact, with regards to the SFP (1.1), Shehu et al [32] used the idea of López et al. [20] to dispense with the dependency of step size on the operator norm and used the alternating inertial technique to recover Férjer monotonicity. They introduced the following algorithm:

Algorithm 2. Alternated Inertial CQ Method

Initialization: Choose $x_0, x_1 \in \mathbb{R}^k$ and set $t = 1$.
Step 1: Compute

$$w_t = \begin{cases} x_t, & t = \text{even} \\ x_t + \theta_t(x_t - x_{t-1}), & t = \text{odd}. \end{cases}$$

Step 2: Compute

$$x_{t+1} = P_{\mathcal{C}}\big(w_t - \lambda_t \nabla F(w_t)\big),$$

Step 3: Set $t \leftarrow t + 1$, and go to **Step 2**,

---

where $F(w_t) = \frac{1}{2}\|(I - P_{\mathcal{Q}})Aw_t)\|,\ \nabla F(w_t) = A^T(I - P_{\mathcal{Q}})Aw_t$,

$$\lambda_t = \begin{cases} \frac{\rho_t F(w_t)}{\|\nabla F(w_t)\|^2}, & \|\nabla F(w_t)\| \neq 0 \\ 0, & \text{otherwise}, \end{cases}$$

$\{\rho_t\} \subset (0,4)$ is nondecreasing, $\{\theta_t\} \subset (0,\infty)$, $\mathcal{C}$ and $\mathcal{Q}$ are nonempty closed and convex sets.

The authors proved that the sequence $\{x_t\}$ generated by Algorithm 2 converges to a point in Ω.

Remark 1.2. Observe that Algorithm 2 is the alternated inertial version of Algorithm 1 in the setting of Euclidean spaces. Furthermore, the dependency of the stepsize on the knowledge of the operator norm in Algorithm 1 has been dispensed with in Algorithm 2.

Still on Remark 1.1, with regards to the failure of the one step inertial acceleration technique, in [26], Polyak discussed that the multi-step inertial methods can boost the speed of optimization methods even though no convergence results of such multi-step inertial methods was given by Polyak [26]. Recent research on multi-step inertial algorithms have been explored, revealing improved numerical efficiency in the results. Notably, consider the findings presented in, for instance, [5, 16, 23, 24]. The idea of the two-step inertia is to compute the inertial term as follows: given three points x_t, x_{t-1}, x_{t-2} the inertial term y_t is computed by

$$y_t = x_t + \theta(x_t - x_{t-1}) + \delta(x_{t-1} - x_{t-2}), \quad \text{where} \quad \theta > 0 \quad \text{and} \quad \delta < 0.$$

Recently, various scholars who have examined the multi-step inertial approaches have shown that numerically, this method has advantage over the one-step inertial algorithms. For some recent results interested readers may see [16, 22, 24, 39] and the references therein.

Inspired by the mentioned works above and the growing interest on multi-step inertial algorithms, our contributions are the following:

- we introduce a new CQ method with a two-step inertial extrapolation and self-adaptive stepsize for finding the solution of the SFP (1.1). The global convergence result and nonasymptotic $O(1/t)$ convergence rate of the sequence generated by our proposed method are presented.
- Our approach includes two-step inertia (which hastens convergence) and self-adaptive step size (which lessens computational complexity). Consequently, our approach gets over the restrictions of the one-step inertia approaches examined in [4, 17, 21] and also the limitation of estimating the linear operator or the spectral radius of a matrix used in [8].
- We solve a multiple set split feasibility problem (which will be discussed in section 5) using our proposed method.
- We give numerical comparison using different problems arising from applications and compared the performances of our proposed algorithm with several algorithms established in the literature.

We organize the rest of the paper as follows: In Section 2, we present some basic definitions, concepts, lemmas and results that are needed in the subsequent sections. The proposed method is presented in Section 3. We study the global convergence analysis and present nonasymptotic $O(1/t)$ convergence rate of the proposed method in Section 4. In Section 5, using our proposed method, we solve the multiple set split feasibility problem. In Section 5.2, we present some numerical results of the proposed method to illustrate the applicability of our method. Finally, we conclude in Section 6.

2. Preliminaries

To obtain our global convergence and nonasymptotic $O(1/t)$ convergence rate, we present some basic results, lemmas and definitions in this section.

Definition 2.1. A mapping $S : \mathbb{R}^k \to \mathbb{R}^k$ is called

(i) *nonexpansive* if

$$\|Sx - Sy\| \leq \|x - y\|, \ \forall \ x, y \in \mathbb{R}^k,$$

(ii) *firmly nonexpansive* if

$$\|Sx - Sy\|^2 \leq \|x - y\|^2 - \|(I - S)x - (I - S)y\|^2, \ \forall \ x, y \in \mathbb{R}^k.$$

Equivalently, the firmly nonexpansive mapping is given by

$$\|Sx - Sy\|^2 \leq \langle x - y, Sx - Sy\rangle, \ \forall \ x, y \in \mathbb{R}^k.$$

Recall that for a nonempty, closed and convex subset $\mathcal{C}$ of $\mathbb{R}^k$, the metric projection denoted by $P_{\mathcal{C}}$, is a map defined on $\mathbb{R}^k$ onto $\mathcal{C}$ which assigns to each $x \in \mathbb{R}^k$, the unique point in $\mathcal{C}$, denoted by $P_{\mathcal{C}}x$ such that

$$||x - P_{\mathcal{C}}x|| \leq ||x - y||, \forall \ y \in \mathcal{C}.$$

Lemma 2.2. *Let $\mathcal{C}$ be a closed and convex subset of a real Hilbert space $\mathbb{R}^k$ and $x, y \in \mathbb{R}^k$. Then*

(i) $\|P_{\mathcal{C}}x - P_{\mathcal{C}}y\|^2 \leq \langle P_{\mathcal{C}}x - P_{\mathcal{C}}y, \ x - y\rangle$;
(ii) $\|P_{\mathcal{C}}x - y\|^2 \leq \|x - y\|^2 - \|x - P_{\mathcal{C}}x\|^2$.

Lemma 2.3. *Let $\mathcal{C}$ be a nonempty closed convex subset of a real Hilbert space $\mathbb{R}^k$. For any $x \in \mathbb{R}^k$ and $z \in \mathcal{C}$, we have*

$$z = P_{\mathcal{C}}x \iff \langle x - z, z - y\rangle \geq 0, \quad \forall \ y \in \mathcal{C}.$$

Lemma 2.4. *The following assertions hold:*

(i) $2\langle x, y\rangle = \|x\|^2 + \|y\|^2 - \|x - y\|^2 = \|x + y\|^2 - \|x\|^2 - \|y\|^2, \ \forall x, y \in \mathbb{R}^k$;
(ii) $\|\alpha x+\beta y\|^2 = \alpha(\alpha+\beta)\|x\|^2+\beta(\alpha+\beta)\|y\|^2-\alpha\beta(1-\alpha)\|x-y\|^2, \ \forall x, y \in \mathbb{R}^k, \ \alpha, \beta \in \mathbb{R}$;
(iii) $\|x + y\|^2 \leq \|x\|^2 + 2\langle y, x + y\rangle, \ \forall \ x, y \in \mathbb{R}^k$.

Lemma 2.5. *Let $x, y, z \in \mathbb{R}^k$ and $\alpha, \beta \in \mathbb{R}$. Then*

$$\begin{aligned}\|(1 + \alpha)x - (\alpha - \beta)y - \beta z\|^2 = (1 + \alpha)\|x\|^2 - (\alpha - \beta)\|y\|^2 \\ - \beta\|z\|^2 + (1 + \alpha)(\alpha - \beta)\|x - y\|^2 \\ + \beta(1 + \alpha)\|x - z\|^2 - \beta(\alpha - \beta)\|y - z\|^2.\end{aligned}$$

Definition 2.6. A function $F : \mathbb{R}^k \to \mathbb{R}$ is called convex, if for all $v \in [0, 1]$ and $x, y \in \mathbb{R}^k$,

$$F(vx + (1 - v)y) \leq vF(x) + (1 - v)F(y).$$

Remark 2.7. If F is convex on $\mathbb{R}^k$ and differentiable then

$$F(y) \geq F(x) + \langle y - x, \ \nabla F(x)\rangle, \ \forall \ y \in \mathbb{R}^k.$$

Definition 2.8. A convex function $F : \mathbb{R}^k \to \mathbb{R}$ is said to be subdifferentiable at a point $x \in \mathbb{R}^k$ if the set

$$\partial F(x) = \{u \in \mathbb{R}^k \mid F(y) \geq F(x) + \langle u, y - x\rangle, \ \forall y \in \mathbb{R}^k\} \tag{2.1}$$

is nonempty, where each element in $\partial F(x)$ is called a subgradient of F at x, $\partial F(x)$ is called the subdifferential of F at x and the inequality in (2.1) is called the subdifferential inequality of F at x.

Remark 2.9. If F is convex and differential, then its gradient and subgradient coincide.

Definition 2.10. A function $F : \mathbb{R}^k \to \mathbb{R}$ is said to be lower semicontinuous at x if

$$x_n \to x \text{ implies } F(x) \leq \liminf_{k\to\infty} F(x_n).$$

Note that F is lower semicontinuous on $\mathbb{R}^k$ if it is lower semicontinuous at every point $x \in \mathbb{R}^k$.

Lemma 2.11. *[9] Let $F(x) := \frac{1}{2}\|(I - P_{\mathcal{Q}})Ax\|^2, \ x \in \mathcal{C}$. Then*

(i) *F is convex and differentiable.*
(ii) *$\nabla F(x) = A^T(I - P_{\mathcal{Q}})Ax, \ x \in \mathbb{R}^k$.*
(iii) *F is lower semicontinuous on $\mathbb{R}^k$.*
(iv) *∇F is Lipschitz continuous with Lipschitz constant $\|A\|^2$.*

3. Proposed method

Assumption 3.1. *The following assumptions will be used in the convergence analysis.*

(a) *$\sigma \in (0, 2)$.*
(b) *θ and δ lie in the region*

$$\mathcal{G} := \left\{(\delta, \theta) : 0 \leq \theta < \frac{4-\sigma}{8-\sigma} < \frac{1}{2}, \ \frac{(8-\sigma)\theta - (4-\sigma)}{8-\sigma+8\theta} < \delta \leq 0\right\}. \tag{3.1}$$

(c) *$\theta \in \left[0, \frac{1}{2}\right)$ and $\delta \leq 0$ such that*

$$|\delta| < 1 - \left(\frac{4+\sigma}{4-\sigma}\right)\theta.$$

(d) *A is a bounded linear operator with adjoint or transpose A^T, C and Q are nonempty closed and convex subsets of $\mathbb{R}^k$ and $\mathbb{R}^m$, respectively.*

Remark 3.2. Conditions (a), (b) and (c) of Assumption 3.1 are easy to verify. Here is a prototype of constants that satisfy these conditions

$$\sigma = 1.5, \ \theta = 0.25, \quad \text{and} \quad \delta = -0.1,$$

we will give a sensitivity analysis of these parameters to suggest the optimal choice for these parameters in section 5.2.

Algorithm 3. Two-step inertial CQ method

Initialization: Let $x_0, x_1, x_2 \in \mathbb{R}^k$ be chosen arbitrarily. Set $t := 0$.

Step 1: Given the iterates $x_{t-2},\ x_{t-1}, x_t$ for each $t \geq 2$, choose δ and θ satisfying Assumption 3.1.

Step 2: Compute

$$w_t = x_t + \theta(x_t - x_{t-1}) + \delta(x_{t-1} - x_{t-2})$$

and

$$x_{t+1} = P_{\mathcal{C}}(w_t - \eta_t \nabla F(w_t)),$$

where

$$F(w_t) := \frac{1}{2}\left\|\left(I - P_{\mathcal{Q}}\right)Aw_t\right\|^2,\ \nabla F(w_t) := A^T(I - P_{\mathcal{Q}})Aw_t$$

and

$$\eta_t := \begin{cases} \frac{\sigma F(w_t)}{\|\nabla F(w_t)\|^2},\ \|\nabla F(w_t)\| \neq 0 \\ 0,\ \text{otherwise.} \end{cases} \tag{3.2}$$

Set $t \leftarrow t + 1$, and go to **Step 2**.

Remark 3.3. When $\delta = 0$, our method reduces to a one-step inertial method for solving SFP studied in Algorithm 1 with constant inertial parameter and self adaptive step size.

4. Convergence Analysis

Lemma 4.1. *Assume that the solution set Ω of (1.1) is nonempty. Then the sequence $\{x_t\}$ generated by Algorithm 3 satisfying Assumption 3.1 is bounded.*

Proof. Let $p \in \Omega$. Since $\nabla F(w_t) = A^T(I - P_{\mathcal{Q}})Aw_t$, we obtain from the firmly nonexpansivity of $I - P_{\mathcal{Q}}$ and the definition of $F(w_t)$ that

$$\begin{aligned} \langle \nabla F(w_t),\ w_t - p\rangle &= \langle A^T(I - P_{\mathcal{Q}})Aw_t,\ w_t - p\rangle \\ &= \langle (I - P_{\mathcal{Q}})Aw_t,\ Aw_t - Ap\rangle \\ &= \langle (I - P_{\mathcal{Q}})Aw_t - (I - P_{\mathcal{Q}})Ap,\ Aw_t - Ap\rangle \\ &\geq \|(I - P_{\mathcal{Q}})Aw_t\|^2 \\ &= 2F(w_t). \end{aligned} \tag{4.1}$$

Using Lemma 2.2 (ii) and definition of η_t, we have

$$\begin{aligned} \|x_{t+1} - w_t\| &= \|P_{\mathcal{C}}(w_t - \eta_t \nabla F(w_t)) - w_t\| \\ &\leq \|w_t - \eta_t \nabla F(w_t) - w_t\| \\ &= \eta_t \|\nabla F(w_t)\| \\ &= \sigma \frac{F(w_t)}{\|\nabla F(w_t)\|}. \end{aligned} \tag{4.2}$$

From the definition of x_{t+1} in Step 2, (4.1) and (4.2), we have

$$\begin{aligned}
\|x_{t+1}-p\|^2 &= \|P_{\mathcal{C}}(w_t-\eta_t\nabla F(w_t))-p\|^2\\
&\le \|w_t-p-\eta_t\nabla F(w_t)\|^2\\
&= \|w_t-p\|^2+(\eta_t)^2\|\nabla F(w_t)\|^2-2\eta_t\langle\nabla F(w_t),\ w_t-p\rangle\\
&\le \|w_t-p\|^2+(\eta_t)^2\|\nabla F(w_t)\|^2-4\eta_t F(w_t)\\
&= \|w_t-p\|^2-4\eta_t F(w_t)+(\eta_t)^2\|\nabla F(w_t)\|^2\\
&= \|w_t-p\|^2-\sigma(4-\sigma)\frac{F^2(w_t)}{\|\nabla F(w_t)\|^2} &(4.3)\\
&\le \|w_t-p\|^2-\frac{4-\sigma}{\sigma}\|x_{t+1}-w_t\|^2. &(4.4)
\end{aligned}$$

Also, by the definition of w_t and Lemma 2.5, we have

$$\begin{aligned}
\|w_t-p\|^2 &= \|(1+\theta)(x_t-p)-(\theta-\delta)(x_{t-1}-p)-\delta(x_{t-2}-p)\|^2\\
&= (1+\theta)\|x_t-p\|^2-(\theta-\delta)\|x_{t-1}-p\|^2\\
&\quad -\delta\|x_{t-2}-p\|^2+(1+\theta)(\theta-\delta)\|x_t-x_{t-1}\|^2\\
&\quad +\delta(1+\theta)\|x_t-x_{t-2}\|^2-\delta(\theta-\delta)\|x_{t-1}-x_{t-2}\|^2. &(4.5)
\end{aligned}$$

From the definition of w_t and applying the Cauchy Schwartz inequality, we have

$$\begin{aligned}
\|x_{t+1}-w_t\|^2 &= \|x_{t+1}-(x_t+\theta(x_t-x_{t-1})+\delta(x_{t-1}-x_{t-2}))\|^2\\
&= \|x_{t+1}-x_t\|^2-2\theta\langle x_{t+1}-x_t,\ x_t-x_{t-1}\rangle\\
&\quad +2\delta\langle x_t-x_{t+1},\ x_{t-1}-x_{t-2}\rangle+\theta^2\|x_t-x_{t-1}\|^2\\
&\quad +2\delta\theta\langle x_t-x_{t-1},\ x_{t-1}-x_{t-2}\rangle+\delta^2\|x_{t-1}-x_{t-2}\|^2\\
&\ge \|x_{t+1}-x_t\|^2-2\theta\|x_{t+1}-x_t\|\|x_t-x_{t-1}\|\\
&\quad -2|\delta|\|x_t-x_{t+1}\|\|x_{t-1}-x_{t-2}\|+\theta^2\|x_t-x_{t-1}\|^2\\
&\quad -2|\delta|\theta\|x_{t-1}-x_t\|\|x_{t-1}-x_{t-2}\|+\delta^2\|x_{t-1}-x_{t-2}\|^2\\
&\ge \|x_{t+1}-x_t\|^2-\theta\Big[\|x_{t+1}-x_t\|^2+\|x_t-x_{t-1}\|^2\Big]\\
&\quad -|\delta|\Big[\|x_{t+1}-x_t\|^2+\|x_{t-1}-x_{t-2}\|^2\Big]\\
&\quad +\theta^2\|x_t-x_{t-1}\|^2-|\delta|\theta\Big[\|x_{t-1}-x_t\|^2+\|x_{t-1}-x_{t-2}\|^2\Big]\\
&\quad +\delta^2\|x_{t-1}-x_{t-2}\|^2\\
&= (1-|\delta|-\theta)\|x_{t+1}-x_t\|^2+(\theta^2-\theta-|\delta|\theta)\|x_t-x_{t-1}\|^2\\
&\quad +(\delta^2-|\delta|-|\delta|\theta)\|x_{t-1}-x_{t-2}\|^2. &(4.6)
\end{aligned}$$

Using the estimate in (4.3), equation (4.5), inequality (4.6) and Assumption 3.1, we have

$$\begin{aligned}
\|x_{t+1}-p\|^2 &\le \|w_t-p\|^2 - \frac{4-\sigma}{\sigma}\|x_{t+1}-w_t\|^2 \\
&\le (1+\theta)\|x_t-p\|^2 - (\theta-\delta)\|x_{t-1}-p\|^2 - \delta\|x_{t-2}-p\|^2 \\
&\quad + \delta(1+\theta)\|x_t-x_{t-2}\|^2 - \delta(\theta-\delta)\|x_{t-1}-x_{t-2}\|^2 \\
&\quad - \frac{(4-\sigma)}{\sigma}\Big[(1-|\delta|-\theta)\|x_{t+1}-x_t\|^2 + (\theta^2-\theta-|\delta|\theta)\|x_t-x_{t-1}\|^2 \\
&\quad + (\delta^2-|\delta|-|\delta|\theta)\|x_{t-1}-x_{t-2}\|^2\Big] \\
&= (1+\theta)\|x_t-p\|^2 - (\theta-\delta)\|x_{t-1}-p\|^2 - \delta\|x_{t-2}-p\|^2 \\
&\quad + \Big[(1+\theta)(\theta-\delta) - \frac{(4-\sigma)}{\sigma}(\theta^2-\theta-|\delta|\theta)\Big]\|x_t-x_{t-1}\|^2 \\
&\quad - \Big[\delta(\theta-\delta) + \frac{(4-\sigma)}{\sigma}(\delta^2-|\delta|-|\delta|\theta)\Big]\|x_{t-1}-x_{t-2}\|^2 \\
&= (1+\theta)\|x_t-p\|^2 - (\theta-\delta)\|x_{t-1}-p\|^2 - \delta\|x_{t-2}-p\|^2 \\
&\quad + \Big[\Big(1-\frac{4-\sigma}{\sigma}\Big)\theta^2 + \theta\Big(1+\frac{4-\sigma}{\sigma}\Big) - \delta - \delta\theta + \Big(\frac{4-\sigma}{\sigma}\Big)|\delta|\theta\Big]\|x_t-x_{t-1}\|^2 \\
&\quad + \Big[\Big(1-\frac{4-\sigma}{\sigma}\Big)\delta^2 - \delta\theta + \frac{(4-\sigma)}{\sigma}|\delta| + \frac{(4-\sigma)}{\sigma}|\delta|\theta\Big]\|x_{t-1}-x_{t-2}\|^2 \\
&\quad - \frac{(4-\sigma)}{\sigma}\Big(1-|\delta|-\theta\Big)\|x_{t+1}-x_t\|^2 \\
&\le (1+\theta)\|x_t-p\|^2 - (\theta-\delta)\|x_{t-1}-p\|^2 - \delta\|x_{t-2}-p\|^2 \\
&\quad + \Big[\Big(\frac{4}{\sigma}\Big)\theta - \delta - \delta\theta + \Big(\frac{4}{\sigma}-1\Big)|\delta|\theta\Big]\|x_t-x_{t-1}\|^2 \\
&\quad + \Big[\Big(\frac{4}{\sigma}-1\Big)|\delta|\theta - \delta\theta + \Big(\frac{4}{\sigma}-1\Big)|\delta|\Big]\|x_{t-1}-x_{t-2}\|^2 \\
&\quad - \Big(\frac{4}{\sigma}-1\Big)\Big(1-|\delta|-\theta\Big)\|x_{t+1}-x_t\|^2. \qquad (4.7)
\end{aligned}$$

From (4.7), we have that

$$\begin{aligned}
\Gamma_{t+1} \le \Gamma_t &+ \Big[\Big(\frac{8}{\sigma}-1\Big)\theta - \Big(\frac{4}{\sigma}-1\Big) + (1+\theta)\Big(\Big(\frac{4}{\sigma}-1\Big)|\delta| - \delta\Big)\Big]\|x_t-x_{t-1}\|^2 \\
&+ \Big[\Big(\frac{4}{\sigma}-1\Big)|\delta|\theta - \delta\theta + \Big(\frac{4}{\sigma}-1\Big)|\delta|\Big]\|x_{t-1}-x_{t-2}\|^2, \qquad (4.8)
\end{aligned}$$

where

$$\begin{aligned}
\Gamma_t &= \|x_t-p\|^2 - \theta\|x_{t-1}-p\|^2 - \delta\|x_{t-2}-p\|^2 \\
&\quad + \Big(\frac{4}{\sigma}-1\Big)\Big(1-|\delta|-\theta\Big)\|x_t-x_{t-1}\|^2, \ \forall\, t \ge 1.
\end{aligned}$$

Also, we observe that

$$\|x_{t-1}-p\|^2 \le 2\|x_t-x_{t-1}\|^2 + 2\|x_t-p\|^2. \qquad (4.9)$$

Now, we show that $\Gamma_t \ge 0, \ \forall t \ge 1$.

Using (4.9) and Assumption 3.1, we have

$$\begin{aligned}\Gamma_t &= \|x_t-p\|^2-\theta\|x_{t-1}-p\|^2-\delta\|x_{t-2}-p\|^2+\Big(\frac{4}{\sigma}-1\Big)\Big(1-|\delta|-\theta\Big)\|x_t-x_{t-1}\|^2\\ &\geq \|x_t-p\|^2-2\theta\|x_t-x_{t-1}\|^2-2\theta\|x_t-p\|^2-\delta\|x_{t-2}-p\|^2\\ &\quad+\Big(\frac{4}{\sigma}-1\Big)\Big(1-|\delta|-\theta\Big)\|x_t-x_{t-1}\|^2\\ &= (1-2\theta)\|x_t-p\|^2-\delta\|x_{t-2}-p\|^2+\Big(\frac{4}{\sigma}-1\Big)\Big(1-|\delta|-\Big(\frac{4+\sigma}{4-\sigma}\Big)\theta\Big)\|x_t-x_{t-1}\|^2\\ &\geq 0. \end{aligned}\tag{4.10}$$

From (4.8), we have

$$\begin{aligned}\Gamma_{t+1}-\Gamma_t &\leq -\Big[\Big(\frac{8}{\sigma}-1\Big)\theta-\Big(\frac{4}{\sigma}-1\Big)+(1+\theta)\Big(\Big(\frac{4}{\sigma}-1\Big)|\delta|-\delta\Big)\Big]\Big(\|x_{t-1}-x_{t-2}\|^2\\ &\quad-\|x_t-x_{t-1}\|^2\Big)-\Big[-\Big(\Big(\frac{8}{\sigma}-1\Big)\theta-\Big(\frac{4}{\sigma}-1\Big)+(1+\theta)\Big(\Big(\frac{4}{\sigma}-1\Big)|\delta|-\delta\Big)\Big)\\ &\quad-\Big(\Big(\frac{4}{\sigma}-1\Big)|\delta|\theta-\delta\theta+\Big(\frac{4}{\sigma}-1\Big)|\delta|\Big)\Big]\|x_{t-1}-x_{t-2}\|^2\\ &= -\Big[\Big(\frac{8}{\sigma}-1\Big)\theta-\Big(\frac{4}{\sigma}-1\Big)+(1+\theta)\Big(\Big(\frac{4}{\sigma}-1\Big)|\delta|-\delta\Big)\Big]\Big(\|x_{t-1}-x_{t-2}\|^2\\ &\quad-\|x_t-x_{t-1}\|^2\Big)-\Big[\Big(\frac{4}{\sigma}-1\Big)-\Big(\frac{8}{\sigma}-1\Big)\theta-2\Big(\frac{4}{\sigma}-1\Big)|\delta|\\ &\quad-2\Big(\frac{4}{\sigma}-1\Big)|\delta|\theta+\delta+2\theta\delta\Big]\|x_{t-1}-x_{t-2}\|^2\\ &= a_1\Big(\|x_{t-1}-x_{t-2}\|^2-\|x_t-x_{t-1}\|^2\Big)-a_2\|x_{t-1}-x_{t-2}\|^2, \end{aligned}\tag{4.11}$$

where

$$a_1=-\Big[\Big(\frac{8}{\sigma}-1\Big)\theta-\Big(\frac{4}{\sigma}-1\Big)+(1+\theta)\Big(\Big(\frac{4}{\sigma}-1\Big)|\delta|-\delta\Big)\Big]$$

and

$$a_2=\Big[\Big(\frac{4}{\sigma}-1\Big)-\Big(\frac{8}{\sigma}-1\Big)\theta-2\Big(\frac{4}{\sigma}-1\Big)|\delta|-2\Big(\frac{4}{\sigma}-1\Big)|\delta|\theta+\delta+2\theta\delta\Big].$$

Since $\delta\leq 0$, $|\delta|=-\delta$ and we have

$$\begin{aligned}a_1 &= -\Big[\Big(\frac{8}{\sigma}-1\Big)\theta-\Big(\frac{4}{\sigma}-1\Big)+(1+\theta)\Big(\Big(\frac{4}{\sigma}-1\Big)|\delta|-\delta\Big)\Big]>0\\ &\iff \frac{(8-\sigma)\theta-(4-\sigma)}{4(1+\theta)}<\delta.\end{aligned}$$

Also,

$$\begin{aligned}a_2 &= \Big[\Big(\frac{4}{\sigma}-1\Big)-\Big(\frac{8}{\sigma}-1\Big)\theta-2\Big(\frac{4}{\sigma}-1\Big)|\delta|-2\Big(\frac{4}{\sigma}-1\Big)|\delta|\theta+\delta+2\theta\delta\Big]>0\\ &\iff \frac{(8-\sigma)\theta-(4-\sigma)}{8-\sigma+8\theta}<\delta.\end{aligned}$$

For $0\leq\theta<\frac{4-\sigma}{8-\sigma}<\frac{1}{2}$, we have

$$(8-\sigma)\theta-(4-\sigma)<\frac{(8-\sigma)\theta-(4-\sigma)}{4(1+\theta)}<\frac{(8-\sigma)\theta-(4-\sigma)}{8-\sigma+8\theta}$$

which implies by Assumption 3.1 that $a_1 > 0$ and $a_2 > 0$ if

$$\frac{(8-\sigma)\theta-(4-\sigma)}{8-\sigma+8\theta} < \delta \leq 0.$$

From (4.11) we have

$$\Gamma_{t+1} + a_1\|x_t - x_{t-1}\|^2 \leq \Gamma_t + a_1\|x_{t-1} - x_{t-2}\|^2 - a_2\|x_{t-1} - x_{t-2}\|^2. \tag{4.12}$$

Also from (4.12), we have

$$\Delta_{t+1} \leq \Delta_t - a_2\|x_{t-1} - x_{t-2}\|^2, \tag{4.13}$$

where $\Delta_{t+1} = \Gamma_{t+1} + a_1\|x_t - x_{t-1}\|^2$ and $\Delta_t = \Gamma_t + a_1\|x_{t-1} - x_{t-2}\|^2$.
Now, since $a_2 > 0$, we have from (4.13) that

$$\Delta_{t+1} \leq \Delta_t$$

which implies that the sequence $\{\Delta_t\}$ is decreasing, thus the limit $\lim_{t\to\infty} \Delta_t$ exists. Consequently, from (4.12) we have that

$$\lim_{t\to\infty} a_2\|x_{t-1} - x_{t-2}\|^2 = 0$$

and

$$\lim_{t\to\infty} \|x_{t-1} - x_{t-2}\| = 0.$$

From the previous equation and the fact that $\lim_{t\to\infty} \Delta_t$ exists, we have that $\lim_{t\to\infty} \Gamma_t$ exists. Also, from the definition of w_t, we have

$$\begin{aligned}\|x_{t+1} - w_t\| &= \|x_{t+1} - x_t - \theta(x_t - x_{t-1}) - \delta(x_{t-1} - x_{t-2})\| \\ &\leq \|x_{t+1} - x_t\| + \theta\|x_t - x_{t-1}\| + |\delta|\|x_{t-1} - x_{t-2}\| \to 0, \text{ as } t \to \infty.\end{aligned} \tag{4.14}$$

Also,

$$\|w_t - x_t\| \leq \theta\|x_t - x_{t-1}\| + |\delta|\|x_{t-1} - x_{t-2}\| \to 0, \ t \to \infty.$$

Using the fact that $\lim_{t\to\infty} \Gamma_t$ exists and $\lim_{t\to\infty} \|x_{t+1} - x_t\| = 0$, we have

$$\lim_{t\to\infty} \left[\|x_t - p\|^2 - \theta\|x_{t-1} - p\|^2 - \delta\|x_{t-2} - p\|^2\right] \tag{4.15}$$

exists. We obtain from (4.10) that the sequence $\{x_t\}$ is bounded. From (4.3), (4.5) and (4.15), we have

$$\lim_{t\to\infty} \frac{F(x_t)}{\|\nabla F(x_t)\|} = 0. \tag{4.16}$$

Also, we note that

$$\begin{aligned}\|\nabla F(x_t)\| &= \|\nabla F(x_t) - \nabla F(p)\| \\ &\leq \|A\|^2\|x_t - p\|, \ \forall\, p \in \Omega.\end{aligned}$$

Hence, $\{\nabla F(x_t)\}$ is bounded. Therefore, from (4.16), we have

$$\lim_{t\to\infty} F(x_t) = 0. \tag{4.17}$$

∎

Theorem 4.2. *Let $\{x_t\}$ be a sequence generated by Algorithm 3 satisfying Assumption 3.1. Assume that the solution set Ω of (1.1) is nonempty. Then $\{x_t\}$ converges to a point in Ω.*

Proof. Since $\{x_t\}$ is bounded from Lemma 4.1, there exists a subsequence $\{x_{t_k}\}$ of $\{x_t\}$ such that $x_{t_k} \to u^* \in \mathbb{R}^k$. Also, since F is lower semicontinuous we have

$$0 \le F(u^*) \le \liminf_{k\to\infty} F(x_{t_k}) = \lim_{t\to\infty} F(x_t) = 0.$$

Therefore, $F(u^*) = 0$. Thus, $Au^* \in \mathcal{Q}$ which implies that $u^* \in \Omega$. Suppose that there exist $\{x_{t_k}\} \subset \{x_t\}$ and $\{x_{t_j}\} \subset \{x_t\}$ such that $x_{t_k} \to u^*$, $k \to \infty$ and $x_{t_j} \to u$, $j \to \infty$. We show that $u = u^*$.
Note that

$$2\langle x_t, u - u^*\rangle = \|x_t - u^*\|^2 - \|x_t - u\|^2 - \|u^*\|^2 + \|u\|^2, \tag{4.18}$$

$$2\langle -\theta x_{t-1}, u - u^*\rangle = -\theta\|x_{t-1} - u^*\|^2 + \theta\|x_{t-1} - u\|^2 + \theta\|u^*\|^2 - \theta\|u\|^2 \tag{4.19}$$

and

$$2\langle -\delta x_{t-2}, u - u^*\rangle = -\delta\|x_{t-2} - u^*\|^2 + \delta\|x_{t-2} - u\|^2 + \delta\|u^*\|^2 - \delta\|u\|^2. \tag{4.20}$$

Combining (4.18), (4.19) and (4.20), we have

$$\begin{aligned} 2\langle x_t - \theta x_{t-1} - \delta x_{t-2}, u - u^*\rangle &= \left(\|x_t - u^*\|^2 - \theta\|x_{t-1} - u^*\|^2 - \delta\|x_{t-2} - u^*\|^2\right) \\ &\quad - \left(\|x_t - u\|^2 - \theta\|x_{t-1} - u\|^2 - \delta\|x_{t-2} - u\|^2\right) \\ &\quad + (1 - \theta - \delta)\left(\|u\|^2 - \|u^*\|^2\right). \end{aligned}$$

From (4.15), we have

$$\lim \left[\|x_t - u\|^2 - \theta\|x_{t-1} - u\|^2 - \delta\|x_{t-2} - u\|^2\right]$$

exists. Also,

$$\lim \left[\|x_t - u^*\|^2 - \theta\|x_{t-1} - u^*\|^2 - \delta\|x_{t-2} - u^*\|^2\right]$$

exists which implies that

$$\lim_{t\to\infty} \langle x_t - \theta x_{t-1} - \delta x_{t-2}, u - u^*\rangle$$

exists. Now,

$$\begin{aligned} \langle u^* - \theta u^* - \delta u^*, u - u^*\rangle &= \lim_{k\to\infty} \langle x_{t_k} - \theta x_{t_{k-1}} - \delta x_{t_{k-2}},\ u - u^*\rangle \\ &= \lim_{t\to\infty} \langle x_t - \theta x_{t-1} - \delta x_{t-2},\ u - u^*\rangle \\ &= \lim_{j\to\infty} \langle x_{t_j} - \theta x_{t_{j-1}} - \delta x_{t_{j-2}}, u - u^*\rangle \\ &= \langle u - \theta u - \delta u, u - u^*\rangle \end{aligned}$$

which implies that

$$(1 - \theta - \delta)\|u - u^*\| = 0.$$

Now, since $\delta \le 0 < 1 - \theta$, see (3.1), we have that $u = u^*$. Hence, the sequence $\{x_t\}$ converges to a point in Ω. This completes the proof. ∎

Now, we are in the position to present a nonasymptotic $O(1/t)$ convergence rate of our proposed Algorithm 3.

Theorem 4.3. *Assume that $\Omega \neq \emptyset$ and $x_0 = x_{-1} = x_{-2}$. Let $\{x_t\}$ be a sequence generated by Algorithm 3 satisfying Assumption 3.1. Then, for any $p \in \Omega$ and $t > 0$, the following condition holds*

$$\min_{0 \le j \le t-2} \|x_{j+1} - w_j\|^2 \le 3\left(1 + \theta^2 + \delta^2\right) \frac{1}{a_2} \frac{(1-\theta-\delta)\|x_0 - p\|^2}{t-1}, \tag{4.21}$$

where $a_2 = \Big[\Big(\frac{4}{\sigma} - 1\Big) - \Big(\frac{8}{\sigma} - 1\Big)\theta - 2\Big(\frac{4}{\sigma} - 1\Big)|\delta| - 2\Big(\frac{4}{\sigma} - 1\Big)|\delta|\theta + \delta + 2\theta\delta\Big]$.

Proof. Let $p \in \Omega$. Then, it follows from (4.13) that

$$\begin{aligned}
\Delta_0 = \Gamma_0 &= \|x_0 - p\|^2 - \theta\|x_{-1} - p\|^2 - \delta\|x_{-2} - p\|^2 + \frac{4-\sigma}{\sigma}(1 - |\delta| - \theta)\|x_0 - x_{-1}\|^2 \\
&= \|x_0 - p\|^2 - \theta\|x_{-1} - p\|^2 - \delta\|x_{-2} - p\|^2 \\
&= (1-\theta-\delta)\|x_0 - p\|^2.
\end{aligned} \tag{4.22}$$

Also, from (4.13) we have

$$a_2 \sum_{j=0}^{t} \|x_{j-1} - x_{j-2}\|^2 \le \Delta_0 - \Delta_{t+1}$$

which implies that

$$\begin{aligned}
\sum_{j=0}^{t} \|x_{j-1} - x_{j-2}\|^2 \le \frac{1}{a_2}\Delta_0 &= \frac{1}{a_2}\Gamma^0 \\
&= \frac{1}{a_2}(1-\theta-\delta)\|x_0 - p\|^2.
\end{aligned}$$

Hence,

$$\min_{0 \le j \le t} \|x_{j-1} - x_{j-2}\|^2 \le \frac{1}{a_2} \frac{(1-\theta-\delta)\|x_0 - p\|^2}{t+1}.$$

Consequently,

$$\min_{0 \le j \le t-1} \|x_j - x_{j-1}\|^2 \le \frac{1}{a_2} \frac{(1-\theta-\delta)\|x_0 - p\|^2}{t}.$$

and

$$\min_{0 \le j \le t-2} \|x_{j+1} - x_j\|^2 \le \frac{1}{a_2} \frac{(1-\theta-\delta)\|x_0 - p\|^2}{t-1}.$$

From (4.14) and using Young's inequality, we have

$$\begin{aligned}
\|x_{t+1} - w_t\|^2 &= \Big(\|x_{t+1} - x_t\| + \theta\|x_t - x_{t-1}\| + |\delta|\|x_{t-1} - x_{t-2}\|\Big)^2 \\
&\le 3\Big(\|x_{t+1} - x_t\|^2 + \theta^2\|x_t - x_{t-1}\|^2 + \delta^2|x_{t-1} - x_{t-2}\|^2\Big).
\end{aligned}$$

Hence,

$$\min_{0\le j\le t-2}\|x_{j+1}-w_j\|^2 \le 3\Big(\min_{0\le j\le t-2}\|x_{j+1}-x_j\|^2+\theta^2\min_{0\le j\le t-2}\|x_j-x_{j-1}\|^2 + \delta^2\min_{0\le j\le t-2}\|x_{j-1}-x_{j-2}\|^2\Big)$$
$$\le 3(1+\theta^2+\delta^2)\frac{1}{a_2}\frac{(1-\theta-\delta)\|x_0-p\|^2}{t-1}$$

which completes the proof. ∎

5. Applications and Numerical Results

5.1. Application to multiple set split feasibility problem

In this section, we present an extension of our proposed method to solve a multiple set split feasibility problem.
Consider the following multi set split feasibility problem:

$$\text{Find } \hat{x}\in\cap_{i=1}^{r}\mathcal{C}_i \text{ such that } A\hat{x}\in\cap_{j=1}^{s}\mathcal{Q}_j, \tag{5.1}$$

where $\mathcal{C}_i,\ i=1,2,\cdots,r,\ \mathcal{Q}_j,\ j=1,2,\cdots,s$ are families of convex closed sets in $\mathbb{R}^k$ and $\mathbb{R}^m$ respectively, and $A:\mathbb{R}^k\to\mathbb{R}^m$ a bounded linear operator. Let $\Theta\neq\emptyset$ be the solution set of the multi set split feasibility problem.
Now, we recall the proximity function which is associated with the multi set split feasibility problem (5.1) (see [36]): Assume that $\varphi_j>0,\ j=1,2,\cdots,s$, then

$$F(x):=\frac{1}{2}\sum_{j=1}^{s}\varphi_j\|A\hat{x}-P_{\mathcal{Q}_j}(A\hat{x})\|^2,\ \forall\hat{x}\in\mathbb{R}^k. \tag{5.2}$$

In literature, it have been shown that F is differentiable with gradient

$$\nabla F(x):=\sum_{j=1}^{s}\varphi_j A^T(I-P_{\mathcal{Q}_j})A\hat{x},\ \forall\ \hat{x}\in\mathbb{R}^k. \tag{5.3}$$

From our proposed method and the idea of Wang et al. [33], we propose a two-step inertial projected method to solve (5.1). The proposed method for solving (5.1) is defined as follows:

Algorithm 4. Two-step inertial CQ method for solving multi set split feasibility problem

---

Initialization: Let $x_0,x_1,x_2\in\mathbb{R}^k$ be chosen arbitrarily. Set $t:=2$.
Step 1: Given the iterates $x_{t-2},\ x_{t-1},x_t$, for each $t\ge 2$, choose δ and θ satisfying Assumption (3.1) and define $I:=\{1,2,\cdots,s\}$.
Step 2: Compute

$$w_t=x_t+\theta(x_t-x_{t-1})+\delta(x_{t-1}-x_{t-2})$$

Step 2: Define the weights $\{w_{t,i}\in(0,\infty):\ i\in I\}$ such that $\sum_{i=1}^{s}w_{t,i}>\tau$ and $\inf_{i\in I^t}w_{t,i}>\tau$, where $I_t:=\{i\in I:\ w_{t,i}>0\}$.
Step 3: Compute

$$x_{t+1}=\sum_{i=1}^{s}w_{t,i}P_{\mathcal{C}_i}(w_t-\eta_t\nabla F(w_t)),$$

where

$$F(w_t) := \frac{1}{2}\sum_{j=1}^{s}\varphi_j\left\|\left(I - P_{\mathcal{Q}_j}\right)Aw_t\right\|^2, \ \nabla F(w_t) := \sum_{j=1}^{s}\varphi_j A^T(I - P_{\mathcal{Q}_j})Aw_t$$

and

$$\eta_t := \begin{cases} \frac{\sigma F(w_t)}{\|\nabla F(w_t)\|^2}, & \|\nabla F(w_t)\| \neq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{5.4}$$

Set $t \leftarrow t+1$, and go to **Step 2**.

Definition 5.1. Let $\{x_t\}$ be a sequence generated by Algorithm 4 and let v be a non-negative integer. Then, the sequence $\{x_t\}$ satisfies the v-intermittent set control if

$$I_t \cup \cdots \cup I_{t+v-1} = I, \ \forall\, t \geq 1.$$

Following the same line of argument in Theorem 3.1 of Wang et al. [33] and Theorem 4.1 we obtain the following result for solving (5.1).

Theorem 5.2. *Assume that the solution set Ω of (1.1) is nonempty. Suppose that there exists a positive integer v such that the sequence $\{x_t\}$ satisfies the v-intermittent set control. Then the sequence $\{x_t\}$ generated by Algorithm 4 satisfying Assumption 3.1 converges to a point in Ω.*

5.2. Numerical illustrations

In this section we will give numerical illustrations and compare the performance of our proposed Algorithm 3 with some recent algorithms in the literature. In the first example, we will perform a sensitivity analysis on the parameters given in Assumption 3.1.

Example 5.3. Let $A : \mathbb{R}^5 \to \mathbb{R}^5$ be a symmetric and positive definite matrix defined by

$$Ax = \begin{pmatrix} 11 & -8 & 0.5 & 0.5 & 2 \\ -8 & 10 & -1.5 & -1.5 & 4 \\ 0.5 & -1.5 & 13 & -1 & -0.5 \\ 0.5 & -1.5 & -1 & 10 & 1 \\ 2 & 4 & -0.5 & 1 & 13 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} \quad \text{then} \quad A^T x = Ax.$$

$$\text{Let } \ \mathcal{C} = \{x \in \mathbb{R}^5 : \|x\| \leq \frac{3}{2}\} \text{ and } \mathcal{Q} = \{x \in \mathbb{R}^5 : \|x\| \leq 2\}.$$

Set $\eta_0 = 1.5$, $x_0 = x_1 = x_2 = (1,1,1,1,1)^T$. The simulation is terminated when $\|x_{t+1} - x_t\| < 10^{-8}$ or $n = 1001$. The results obtained are presented in Table 1 and Figures 1, 2 and 3.

Table 1. Convergence Performance of Algorithm 3 with Respect to Parameters

σ	θ	δ	**Number of Iterations**	**CPU Time (secs)**
1.5	0.25	-0.1	33	0.0049
1.5	0.25	-0.05	31	0.0036
1.5	0.25	-0.01	34	0.0065
1.5	0.25	0	36	0.0087
1.9	0	-0.05	33	0.0056
1.9	0.1	-0.05	32	0.0052
1.9	0.2	-0.05	30	0.0048
1.9	0.3	-0.05	28	0.0039
1.5	0.3	-0.05	31	0.0055
1	0.3	-0.05	39	0.0085
0.5	0.3	-0.05	39	0.0165
0.1	0.3	-0.05	244	0.0403

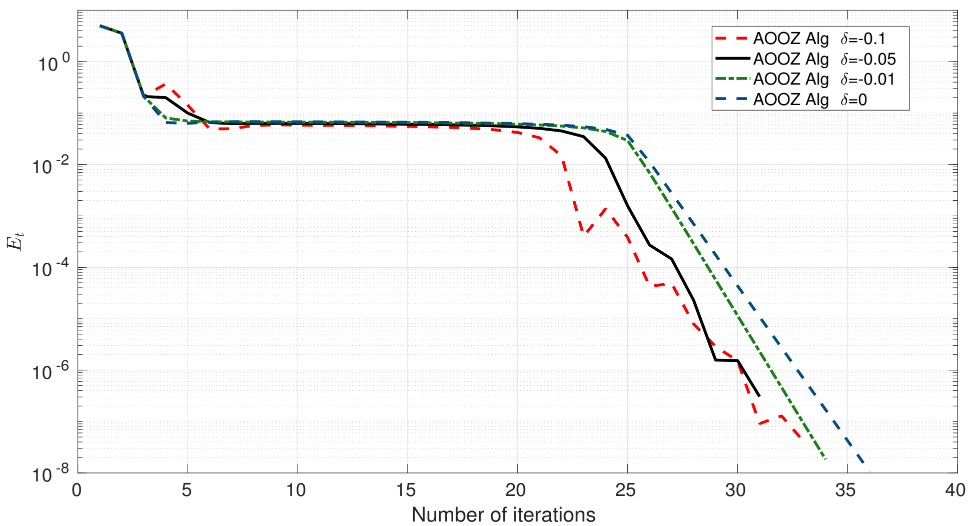

Figure 1. Behaviour of Algorithm 3 for $\theta = 0.25$, $\sigma = 1.5$ and varied values of δ

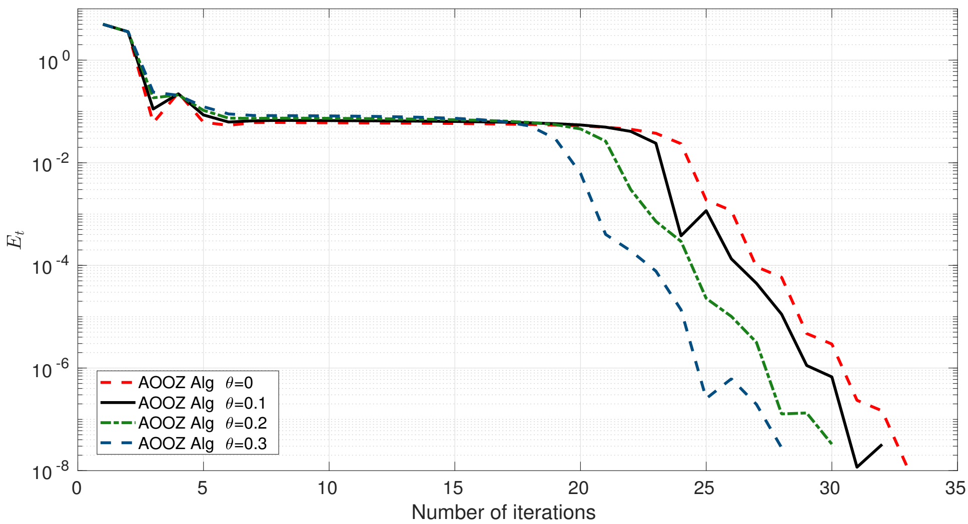

Figure 2. Behaviour of Algorithm 3 for $\delta = -0.05$, $\sigma = 1.9$ and varied values of θ

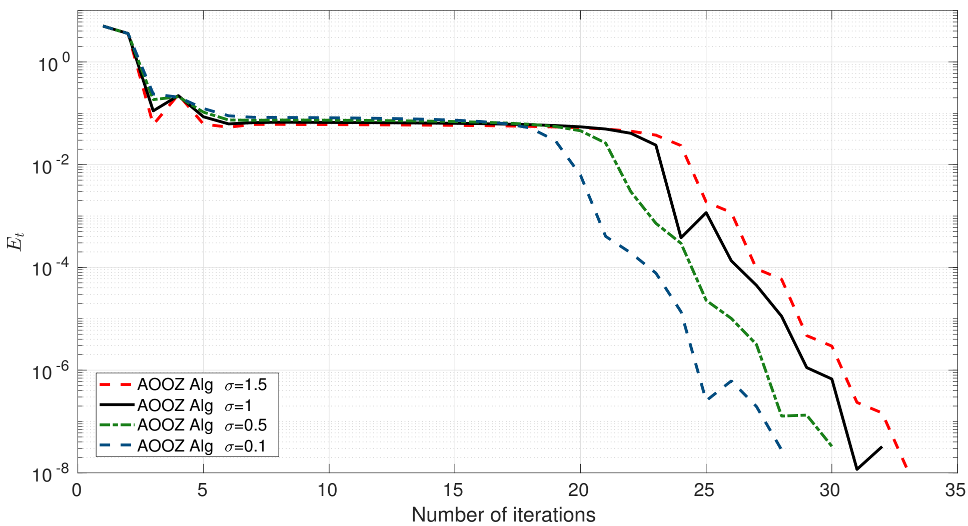

Figure 3. Behavior of Algorithm 3 for $\delta = -0.05$, $\theta = 0.3$ and varied values of σ

Discussion of Results. From Table 1 and Figures 1, 2 and 3, we observe that the choice of δ, θ and σ are very sensitive and affects the behavior of our proposed Algorithm 3. Based on the results in Table 1 and the simulations in Figures 1, 2 and 3, we see that the best set of parameters for Algorithm 3 are $\delta = -0.05$, $\theta = 0.3$ and $\sigma = 1.9$. Using the best set of parameters, we will compare the performance of our proposed Algorithm 3 (AOOZ Alg) with Algorithm 3.1 of Dang et al. [17] (DSX Alg), Algorithm 1 of Dong et al. [18] (DLQY Alg), Algorithm 1 of Shehu et al. [32] (SDL Alg) and Algorithm 4.1 of Wang and Xu [34] (WX Alg). Since the choice of parameters are sensitive to the behaviour algorithms, we will choose the same parameters used by the authors in their papers. However, Wang and Xu [34] did not give any numerical example in their paper. Hence, in WX Alg we choose $\alpha_n = \frac{1}{10^5(t+1)}$ and $\gamma = 10^{-4}$. We will consider two set of initial points to test the robustness of each algorithm. The simulation is terminated when $E_t = \|x_{t+1} - x_t\| < 10^{-8}$ or $t = 1001$. The results obtained for Case 1: $x_0 = x_1 = x_2 = (0.5, 1, 2, 0.25, -1)^T$ and Case 2: $x_0 = x_1 = x_2 = (-2, 1.5, -2, -0.25, 3)^T$ and are presented in Table 2, Figure 4 and Figure 5.

Table 2. Table for Two Cases of the Initial Points

Algorithm	Case	Number of Iterations	CPU Time (secs)
AOOZ Alg	Case 1	28	0.0036
	Case 2	31	0.0077
DSX Alg	Case 1	101	0.0206
	Case 2	92	0.0254
DLQY Alg	Case 1	99	0.0254
	Case 2	98	0.0174
SDL Alg	Case 1	415	0.0507
	Case 2	303	0.0571
WX Alg	Case 1	412	0.0829
	Case 2	410	0.0577

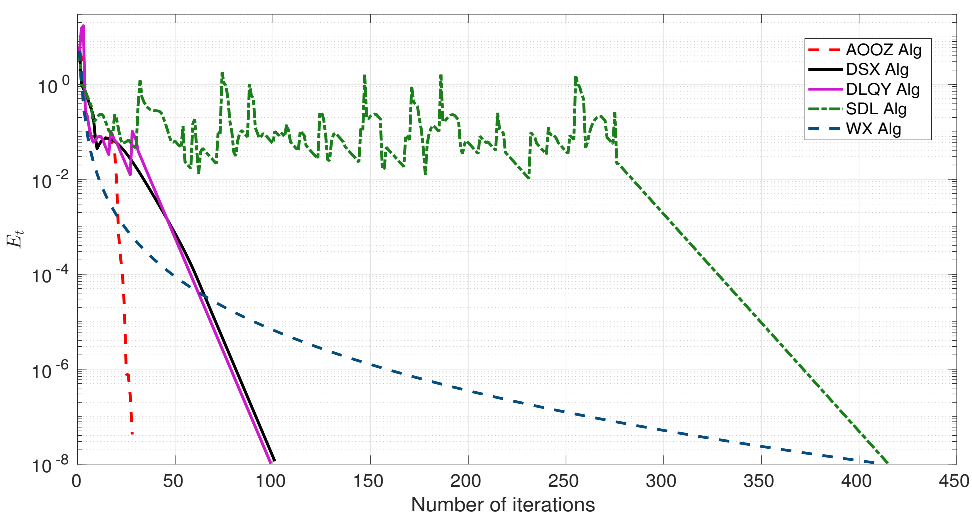

Figure 4. Graph the Iterates for the Cases 1

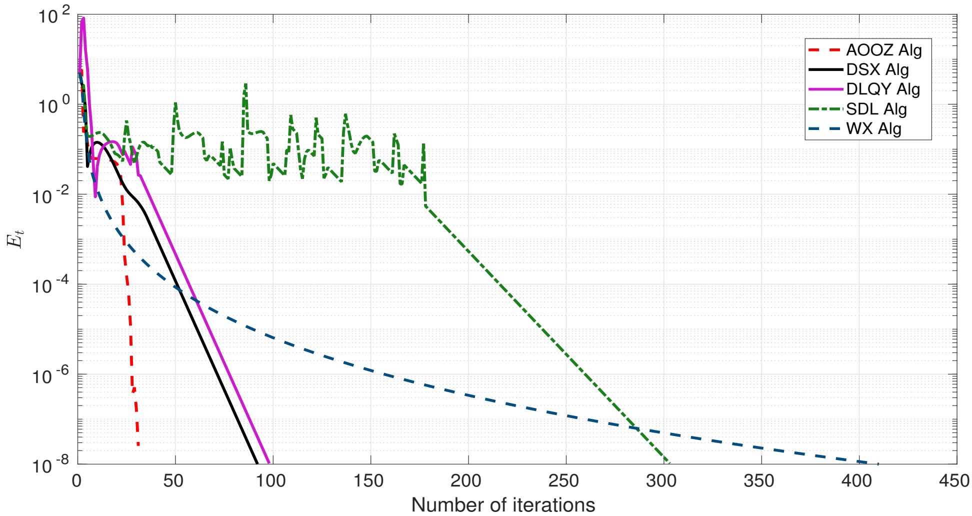

FIGURE 5. Graph the Iterates for the Cases 2

In the next example we test the robustness of each algorithm for higher dimension.

Example 5.4. Let $A : \mathbb{R}^k \to \mathbb{R}^k$ be defined as $Ax = Bx$ where B is a randomly generated matrix such that its entries $b_{ij} \in (0, 1)$.

$$\text{Let } \mathcal{C} = \left\{x \in \mathbb{R}^k : \|x\| \leq \frac{3}{2}\right\} \text{ and } \mathcal{Q} = \{x \in \mathbb{R}^k : \|x\| \leq 2\}.$$

We will consider two dimensions $k = 100$ and $k = 1000$ and study the behaviour of each algorithm with respect to these dimensions. The initial guess x_0 is generated randomly and we set $x_1 = x_2 = x_0$ for all the algorithms. Since x_0 is generated randomly, as we saw in Example 5.3, the choice of x_0 affects the required computational time and number of iterations to satisfy the stopping criteria. So, for fair comparison, for each algorithm, we run the simulation ten (10) times and we report the best performance. The simulation is terminated when $E_t = \|x_{t+1} - x_t\| < 10^{-8}$ or $t = 3001$. The results obtained for each dimension is presented in Table 3, Figure 6 and Figure 7.

TABLE 3. Table for Two Cases of the Initial Points

Algorithm	Dimension k	Number of Iterations	CPU Time (secs)
AOOZ Alg	100 1000	38 656	0.0096 2.3878
DSX Alg	100 1000	387 3000	0.0601 12.2461
DLQY Alg	100 1000	117 871	0.0432 3.0320
SDL Alg	100 1000	323 1397	0.0517 4.8420
WX Alg	100 1000	457 2537	0.0357 6.1402

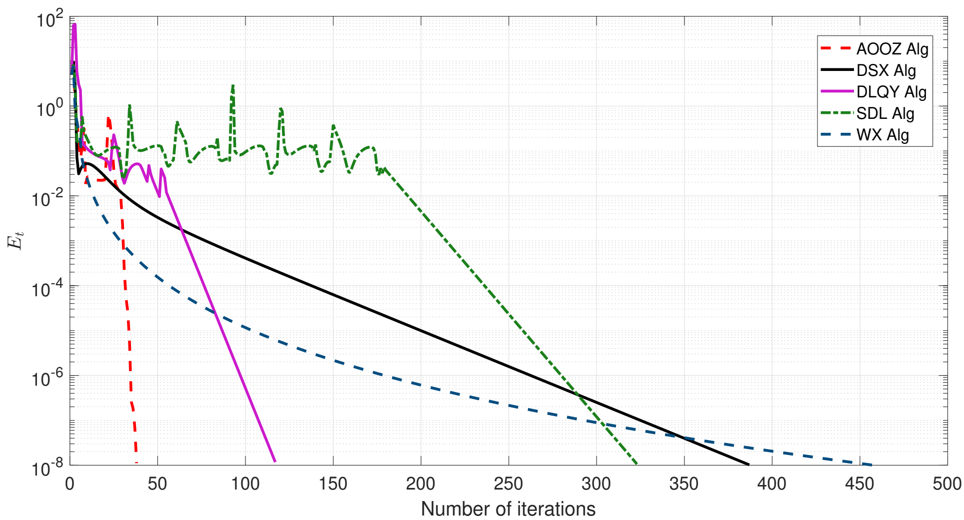

FIGURE 6. Graph the Iterates for $k = 100$

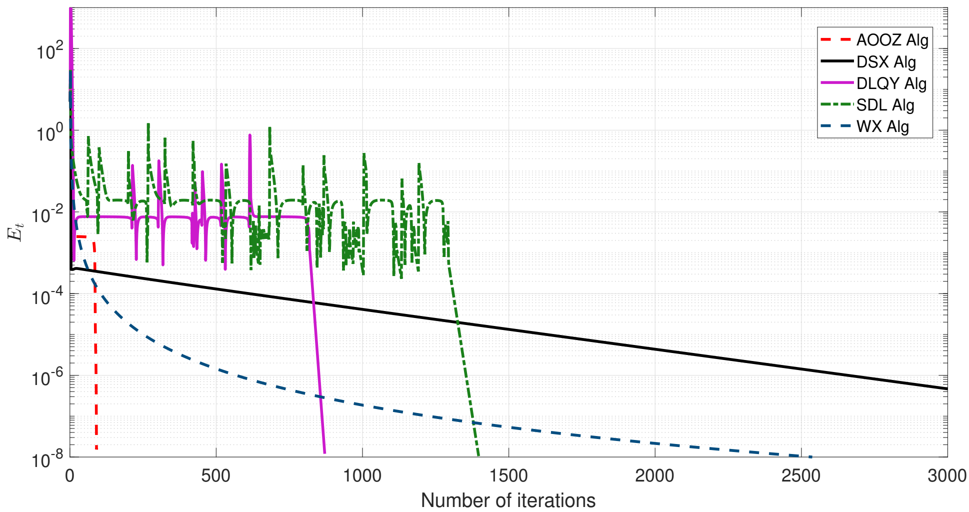

Figure 7. Graph the Iterates for 1000

Discussion of Results. We observe from Examples 5.3 and 5.4 our proposed Algorithm 3 outperforms the Algorithms Dang et al. [17] (DSX Alg), Dong et al. [18] (DLQY Alg), Shehu et al. [32] (SDL Alg) and Wang and Xu [34] (WX Alg) in terms of the computational time and number of iterations required to satisfy the stopping criterion. Surprisingly, for these examples, the algorithms of Dong et al. [18] and Shehu et al. [32] stopped after a few iteration when we choose the inertial term w_n to be alternating as given their respective papers. For the purpose of illustrations, it should be noted that the results presented for the algorithms of Dong et al. [18] and Shehu et al. [32] are for the fully inertial versions of their respective algorithms. This perhaps should serve as an explanation for the oscillatory behaviour of the iterates generated by the algorithms. Based on the iterates generated by the algorithm Dang et al. [17] (DSX Alg) when we increased the dimension to 1000, we can concluded that their algorithm is not suitable for higher dimension.

5.3. Application to LASSO Problem

Example 5.5. We consider the following LASSO problem

$$\min\left\{\frac{1}{2}\|Ax-b\|_2^2 : \; x\in\mathbb{R}^k, \; \|x\|_1 \le r\right\}, \tag{5.5}$$

where $A \in \mathbb{R}^{m\times k}$, $m < k$, $b \in \mathbb{R}^m$ and $r > 0$. We consider $k = 6144$ and $m = 1440$. A normal distribution with a standard deviation of zero and a unit variance serves as the basis for the matrix A. Additionally, the genuine spare signal x_* is formed by uniformly dispersing throughout the interval $[-1, 1]$ with spikes (nonzero entries) 90 and 180 while the rest are kept at zero. The sample data b is given as $b = Ax_*$.
The solution of the minimization problem (5.5) under certain conditions on the matrix A is similar to the ℓ_1-norm solution of the under determined linear system. For the problem

under consideration (1.1) , we define

$$\mathcal{C} = \Big\{x : \|x\|_1 \leq t\Big\} \text{ and } \mathcal{Q} = \{b\}.$$

We will use the subgradient projection since the projection onto the closed convex $\mathcal{C}$ does not have a closed form solution. Now, we define a convex function $d(x) := \|x\|_1 - t$ and define

$$\mathcal{C}_t = \Big\{x \in \mathbb{R}^k : d(w_t) + \langle \zeta_t,\ x - w_t\rangle \leq 0\},$$

where $\zeta \in \partial d(w_t)$. The orthogonal projection on $\mathcal{C}_t$ is given by

$$P_{\mathcal{C}_t}(\tilde{y}) = \begin{cases} y, & d(w_t) + \langle \zeta_t, \tilde{y} - w_t\rangle \leq 0, \\ y - \frac{d(w_t)+\langle \zeta_t, \tilde{y}-w_t\rangle \zeta_t}{\|\zeta_t\|^2}, & \text{otherwise.} \end{cases}$$

Therefore, at point x, the subdifferential ∂c is given by

$$\partial c(x) = \begin{cases} 1, & x > 0, \\ [-1, 1], & x = 0, \\ -1, & x < 0. \end{cases}$$

Now, we will compare the performance of our proposed Algorithm 3 (AOOZ Alg) with Algorithm 3.1 of Dang et al. [17] (DSX Alg), Algorithm 1 of Dong et al. [18] (DLQY Alg), Algorithm 1 of Shehu et al. [32] (SDL Alg) Algorithm 4.1 of Wang and Xu [34] (WX Alg). in the restoration process of the sparse signal. We use the same control parameters used in Examples 5.3 and 5.4. Furthermore, we evaluate the mean square error (MSE) defined by:

$$MSE = \frac{1}{k}\|x_* - x\|^2 \tag{5.6}$$

to make sure that the restored signal has a good length and observation compared to the original signal, where x_* is an approximated signal of x. The initial points x_0, x_1, x_2 are choosen to be zeros and we use tolerance of $\|x_t - x_*\| < 10^{-5}$ and maximum number of iterations $t = 1000$ are used as stopping criterion. We present the results of the numerical simulation in Figures 8 and 9.

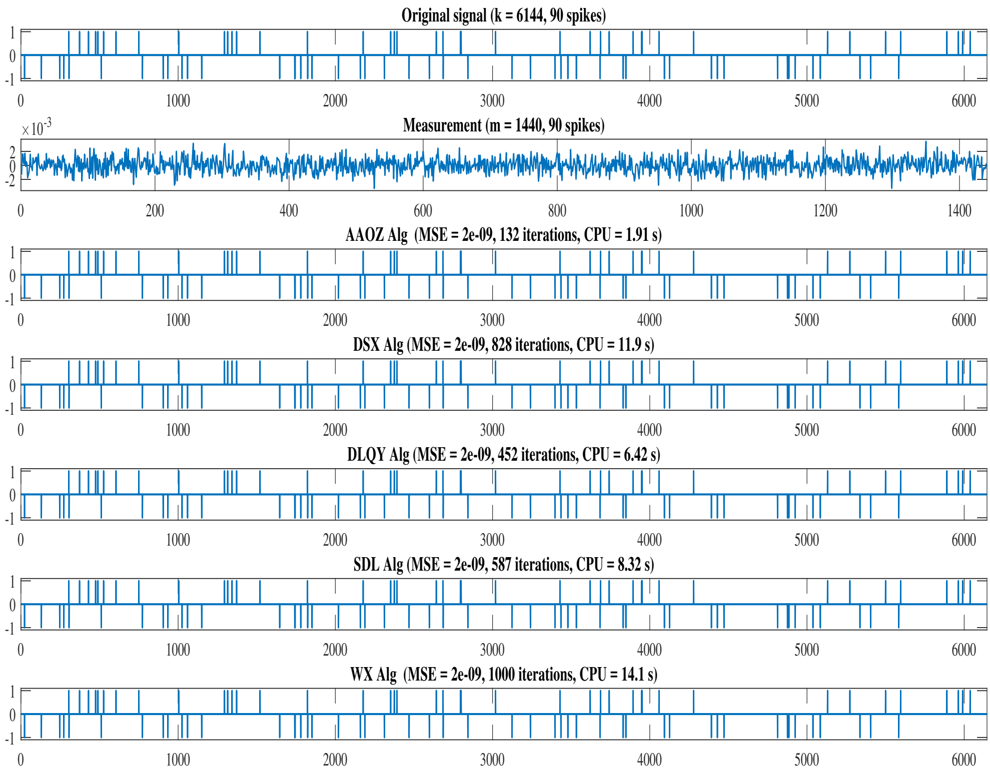

Figure 8. Restored Signal via One-Inertia and Two-Inertia for 90 spikes

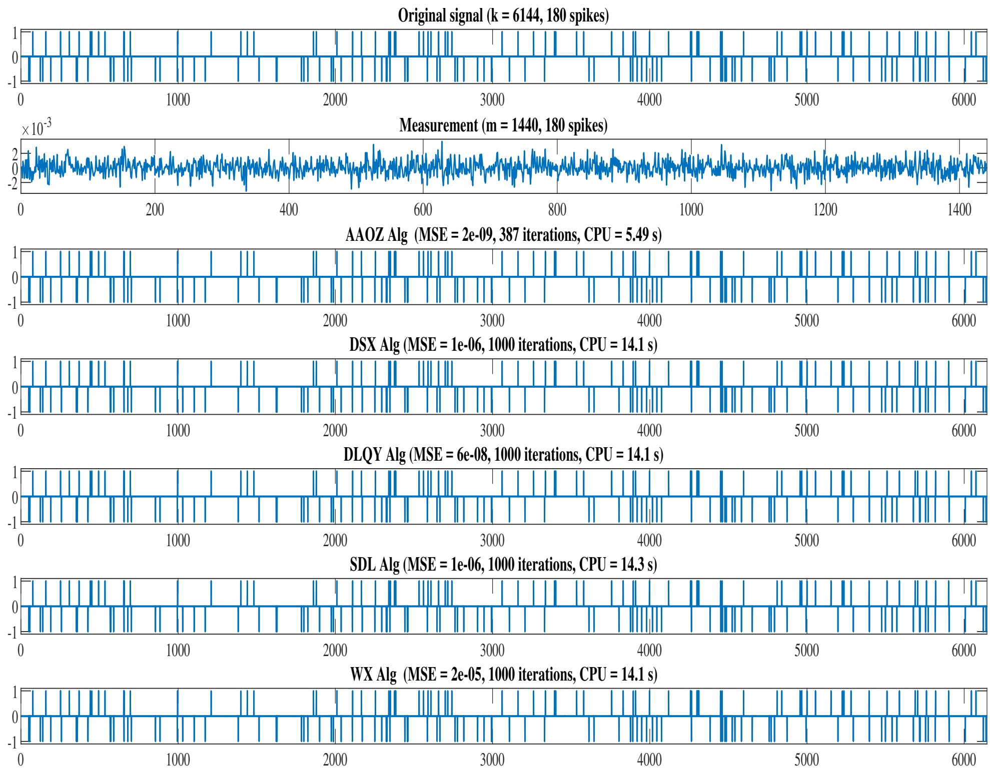

FIGURE 9. Restored Signal via One-Inertia and Two-Inertia for 180 spikes

Discussion of Results. We observe from Figures 8 and 9 the MSE values of all the algorithms show that restored signals have a good length and observation compared to the original signal. However, the superiority between the algorithms will be determined by the required number of iterations and CPU time to satisfy the stopping criteria since the MSE values is only a measure of the quality of the restored signals. Looking at these Figures again, one will see that our proposed algorithm (AOOZ Alg) restored the signals in much fewer iterations and CPU time with the best MSE value compared to other algorithms. Notably, when the number of spikes was increase to 180, all the methods failed to satisfy the stopping criteria before the specified maximum number of iteration (1000) was reached. But for 180 spikes, it took our proposed algorithm just 387 iteration and a CPU time of 5.49s to restore the signal with accuracy (MSE value) of 2E-09. Thus, based on these examples, we see that our proposed two-step inertial algorithm (Algorithm 3) provides computational advantage over the algorithms of Algorithm 3.1 of Dang et al. [17] (DSX Alg), Algorithm 1 of Dong et al. [18] (DLQY Alg), Algorithm 1 of Shehu et al. [32] (SDL Alg) Algorithm 4.1 of Wang and Xu [34] (WX Alg).

6. Conclusion

In this work, we studied the split feasibility problem. We introduced an algorithm with two-step inertial extrapolation and self adaptive stepsize to solve the aforementioned problem. We presented a global convergence and non-asymptotic $O(1/t)$ convergence rate of the proposed method. Furthermore, we applied our proposed method solve a multiple set feasibility problem. Finally, we presented numerical results of our proposed method to illustrate the applicability of our method and presented an application of our result to LASSO problem.

7. Declaration

Acknowledgment The authors acknowledges with thanks the School of Mathematics at University of the Witwatersrand for making their facilities available for the research.

Funding
The first author is supported by the postdoctoral research grant from University of the Witwatersrand, South Africa.

Availability of data and material
Not applicable.

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

Authors' contributions
All authors worked equally on the results and approved the final manuscript.

References

[1] A.A. Adam, A. Adamu, A.H. Ibrahim, D.U. Ozsahin, Inertial Halpern-type methods for variational inequality with application to medical image recovery, Communications in Nonlinear Science and Numerical Simulation, (2024), Article ID 108315. https://doi.org/10.1016/j.cnsns.2024.108315.

[2] A. Adamu, D. Kitkuan, T. Seangwattana, An accelerated Halpern-type algorithm for solving variational inclusion problems with applications, Bangmond International Journal of Mathathematics and Computational Science, 8 (2022) 37–55. https://doi.org/10.58715/bangmodjmcs.2022.8.4.

[3] A. Adamu, H. Umar, S.E. Akinade, D.U. Ozsahin, Comparative analysis of FISTA and inertial Tseng algorithm for enhanced image restoration in prostate cancer imaging, Applied Mathematics in Science and Engineering, 32(1) (2024), Article ID 2388247. https://doi.org/10.1080/27690911.2024.2388247.

[4] H.A. Abass, K.O. Aremu, L.O. Jolaoso, O.T. Mewomo, An inertial forward-backward splitting method for approximating solutions of certain optimization problems, Journal of Nonlinear Functional Analysis, (2024), Article ID 6.

[5] H.A. Abass, A. Adamu, M.A. Phane, A two-step inertial algorithm for solving equilibrium problems on Hadamard manifolds, Carpathian Journal of Mathematics, 41(2) (2025) 253–272.

[6] F. Alvarez, H. Attouch, An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping, Set Valued Analysis, 9(1-2) (2001) 3–11. https://doi.org/10.1023/A:1011253113155.
[7] R.I. Boţ, E.R. Csetnek, C. Hendrich, Inertial Douglas-Rachford splitting for monotone inclusion, Applied Mathematics and Computation, 256 (2015) 472–487. https://doi.org/10.1016/j.amc.2015.01.017.
[8] C.L. Byrne, Iterative oblique projection onto convex sets and the split feasibility problem, Inverse Problems, 18 (2002) 441–453. https://doi.org/10.1088/0266-5611/18/2/310.
[9] C. Byrne, A unified treatment of some iterative algorithms in signal processing and image reconstruction, Inverse Problems, 20(1) (2004) 103–120. https://doi.org/10.1088/0266-5611/20/1/006.
[10] Y. Censor, T. Elfving, A multiprojection algorithm using Bregman projections in a product space, Numerical Algorithms, 8(2-4) (1994) 221–239. https://doi.org/10.1007/BF02142692.
[11] Y. Censor, A. Gibali, S. Reich, Algorithms for the split variational inequality problem, Numerical Algorithms, 59 (2012) 301–323. https://doi.org/10.1007/s11075-011-9490-5.
[12] Y. Censor, T. Bortfeld, B. Martin, A. Trofimov, A unified approach for inversion problems in intensity modulated radiation therapy, Physics in Medicine & Biology, 51 (2006) 2353–2365. https://doi.org/10.1088/0031-9155/51/10/001.
[13] C.E. Chidume, A. Adamu, Solving split equality fixed point problem for quasi-phi-nonexpansive mappings, Thai Journal of Mathematics, 19(4) 1699–1717.
[14] C.E. Chidume, A. Adamu, A new iterative algorithms for split feasibility and fixed point problems, Journal of Nonlinear Variational Analysis, 5(2) (2021) 201–210. https://doi.org/10.23952/jnva.5.2021.2.02.
[15] C.E. Chidume, A. Adamu, On split generalized mixed equality equilibrium and split equality fixed point problems. Applied Set-Valued Analysis and Optimization, 2(3) 273–283.
[16] C. Izuchukwu, Y. Shehu, Q.L. Dong, Two-step inertial forward-reflected-backward splitting based algorithm for nonconvex mixed variational inequalities, Journal of Computational and Applied Mathematics, 426 (2023), Article ID 115093. https://doi.org/10.1016/j.cam.2023.115093.
[17] Y. Dang, J. Sun, H. Xu, Inertial accelerated algorithms for solving a split feasibility problem, Journal of Industrial and Management Optimization, 13(3) (2017) 1383–1394. https://doi.org/10.3934/jimo.2016078.
[18] Q.L. Dong, L.L. Liu, X. Qin, J.C. Yao, An alternated inertial general splitting method with linearization for the split feasibility problem, Optimization, 71 (2023) 2585–2607. https://doi.org/10.1080/02331934.2022.2069567.
[19] A. Gibali, N.H. Ha, N.T. Thuong, T.H. Trang, N.T. Vinh, Polyaks gradient method for solving the split convex feasibility problem and its applications, Journal of Applied and Numerical Optimization, 1 (2019) 145–156. https://doi.org/10.23952/jano.1.2019.2.05.
[20] G. Lpez, V. Martn-Mrquez, F. Wang, H.K. Xu, Solving the split feasibility problem without prior knowledge of matrix norms, Inverse Prolems, 28(8) (2012), Article ID 085004. https://doi.org/10.1088/0266-5611/28/8/085004.

[21] 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.

[22] C. Guo, J. Zhao, Q.L. Dong, A stochastic two-step inertial Bregman proximal alternating linearized minimization algorithm for nonconvex and nonsmooth problems, Numerical Algorithms, (2023).

[23] A.E. Ofem, J.A. Abuchu, G.C. Ugwunnadi, H.A. Nabwey, A. Adamu, O.K. Narain, Double inertial steps extragradient-type methods for solving optimal control and image restoration problems, AIMS Mathematics, 9(5) 12870–12905. https://doi.org/10.3934/math.2024629.

[24] C.C. Okeke, A. Adamu, R. Promkam, P. Sunthrayuth, Two-step inertial method for solving split common null point problem with multiple output sets in Hilbert spaces, AIMS Mathematics, 8(9) 20201–20222. https://doi.org/10.3934/math.20231030.

[25] B.T. Polyak, Some methods of speeding up the convergence of iterative methods, USSR Computational Mathematics and Mathematical Physics, 4(5) (1964), 1–17. https://doi.org/10.1016/0041-5553(64)90137-5.

[26] B.T. Polyak, Introduction to optimization, New York, Optimization Software, Publications Division, 1987.

[27] C. Poon, J. Liang, Trajectory of Alternating Direction Method of Multipliers and Adaptive Acceleration, In Advances In Neural Information Processing Systems (2019).

[28] Y. Nesterov, A method for solving the convex programming problem with convergence rate $O(1/k^2)$, Doklady Akademii Nauk SSSR, 269 (1983) 543–547.

[29] A. Taiwo, L.O. Jolaoso, O.T. Mewomo, Viscosity approximation method for solving the multiple-set split equality common fixed-point problems for quasi-pseudocontractive mappings in Hilbert Spaces. Journal of Industrial and Management Optimization, (2020)

[30] A. Taiwo, L.O. Jolaoso, O.T. Mewomo, Parallel hybrid algorithm for solving pseudomonotone equilibrium and split common fixed point problems, Bulletin of the Malaysian Mathematical Sciences Society, 43 (2020) 1893–1918. https://doi.org/10.1007/s40840-019-00781-1.

[31] Y. Shehu, Iterative methods for split feasibility problems in certain Banach spaces, Journal of Nonlinear Convex Analysis, 16(12) (2015) 2351–2364.

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

[33] J. Wang, Y. Hu, C.K. W. Yu, X. Zhuang, A family of projection gradient methods for solving the multiple sets split feasibility problem, Journal of Optimization Theory and Applications, 183(2) (2019) 520–534. https://doi.org/10.1007/s10957-019-01563-2.

[34] F. Wang, H.K. Xu, Approximating curve and strong convergence of the CQ algorithm for the split feasibility problem. Journal of Inequalities and Applications, (2010) 1–13. https://doi.org/10.1155/2010/102085.

[35] Z. Wang, P. Sunthrayuth, R. Promkam, A. Adamu, Three novel inertial subgradient extragradient methods for quasi-monotone variational inequalities in Banach spaces, Computational and Applied Mathematics, 43 (8) (2019), Article ID 421.

https://doi.org/10.1007/s40314-024-02929-7.

[36] H.K. Xu, A variable Krasnosel'skii-Mann algorithm and multiple-set split feasibility problem, Inverse Problems, 22 (2006) 2021–2034. https://doi.org/10.1088/0266-5611/22/6/007.

[37] Q. Yang, On variable-step relaxed projection algorithm for variational inequalities, Journal of Mathematical Analysis and Applications, 302 (2004) 166–179. https://doi.org/10.1016/j.jmaa.2004.07.048.

[38] H. Yu, W. Zhan, F. Wang, The ball-relaxed CQ algorithms for the split feasibility problem, Optimization, 67(10) (2018) 1687–1699. https://doi.org/10.1080/02331934.2018.1485677.

[39] J. Zhao, Q.L. Dong, M.T. Rassias, F. Wang, Two-step inertial Bregman alternating minimization algorithm for nonconvex and nonsmooth problems, Journal of Global Optimization, 84(2022) 941–966. https://doi.org/10.1007/s10898-022-01176-6.