An inertial projective splitting method for the sum of two maximal monotone operators

Majela Pentón Machado *

Abstract

We propose a projective splitting type method to solve the problem of finding a zero of the sum of two maximal monotone operators. Our method considers inertial and relaxation steps, and also allows inexact solutions of the proximal subproblems within a relative-error criterion. We study the asymptotic convergence of the method, as well as its iteration-complexity. We also discuss how the inexact computations of the proximal subproblems can be carried out when the operators are Lipschitz continuous. In addition, we provide numerical experiments comparing the computational performance of our method with previous (inertial and non-inertial) projective splitting methods.

Keywords. splitting algorithms; inertial algorithms; relative error; maximal monotone operators; complexity.

AMS Classification: 47H05, 90C30, 90C25.

1 Introduction

We are concerned with the *monotone inclusion* problem (MIP) of finding $z \in \mathcal{X}$ such that

$$0 \in G^* A(Gz) + B(z), \tag{1}$$

where $\mathcal{X}, \mathcal{Y}$ are real Hilbert spaces, $A : \mathcal{Y} \rightrightarrows \mathcal{Y}$ and $B : \mathcal{X} \rightrightarrows \mathcal{X}$ are *maximal monotone* point-to-set operators, and $G : \mathcal{X} \to \mathcal{Y}$ is a bounded linear operator. Problem (1) provides a framework for studying a broad class of problems [21]. In particular, an important instance of (1) is the optimization problem

$$\min_{z \in \mathcal{X}} f(Gz) + g(z) \tag{2}$$

where $f : \mathcal{Y} \to (-\infty, \infty]$ and $g : \mathcal{X} \to (-\infty, +\infty]$ are *proper*, convex and lower-semicontinuous functions. It is well-known that under appropriate assumptions, (2) is equivalent to (1) when $A = \partial f$ and $B = \partial g$ are the *subdifferentials* of the functions f and g.

Splitting methods (or *decomposition methods*) are a powerful tool for solving the structured monotone inclusion problem (1) (and (2)). The *Peaceman-Rachford* and *Douglas-Rachford* methods, first introduced in [30, 16] for the case of linear mappings and later generalized in [24] to address MIPs, are examples of this kind of algorithm. Other well-known examples are the *forward-backward* methods [24, 31, 36], which generalize gradient projection methods, and double-backward methods [8].

A relatively new class of splitting algorithms is the *projective splitting* method. This class has its origins in the works [18, 19] and since then, a great effort has been devoted to the design and study of the projective splitting methods due to their flexibility in the selection of the stepsizes, compositions with linear operators, and the possibility of block-iterative and asynchronous implementations, see [1, 20, 15, 26, 22, 32, 23] and references therein. Projective splitting methods can be seen as separator-projection methods. Indeed, they

*Instituto de Matemática e Estatística, Universidade Federal da Bahia, Campus de Ondina, Av. Adhemar de Barros, 40170-110 Salvador, BA, Brasil, majela.penton@ufba.br. ORCID: 0000-0003-1975-8017

work by (inexactly) solving *proximal subproblems*, each of which involves a single operator, to construct a separating hyperplane between the current iterate and an *extended solution set* of (1). Then, the next iterate is taken as a relaxed projection of the current one onto this separating hyperplane.

In this work, we propose an *inertial* projective splitting method to solve (1). Inertial algorithms for monotone inclusion problems defined by one operator were first analyzed in [2], where the authors studied the inertial *proximal point* (PP) method. Later, the inertial PP method was used to develop inertial versions of inexact PP algorithms, Douglas-Rachford and ADMM methods, among others, see [3, 5, 6, 10, 14, 25] and references therein.

Our main goal is to study a projective splitting method that combines inertial steps, relaxed projections and inexact calculations of the proximal subproblems within relative-error criteria. To the best of our knowledge, the only work that studies inertial projective methods is [4], where a relative-error inertial-relaxed inexact projective splitting algorithm is developed for solving structured monotone inclusion problems involving the sum of finitely many maximal monotone operators. For the case $n = 2$, the method of [4] is different from ours, as it only considers handling the operators in a parallel manner. In contrast, we allow treating the operators sequentially, which, as shown in our preliminary numerical experiments, could be advantageous in practice since the second proximal subproblem to be solved uses the more recent information generated at the iteration. The inertial and relaxation parameters of the proposed projective splitting method follow the mutual constraint of [3, 4], improving the usual $1/3$ upper bound for the sequence of inertial parameters. To approximately solve the proximal subproblems, we consider the relative-error criterion proposed in [34] that allows the use of the *ε-enlargements* of the operators. We observe that this criterion is slightly more flexible than the one used in [4], which does not consider elements of the enlargements of the maximal monotone operators.

For our inertial projective splitting method, we first establish the weak convergence of the generated sequences to a solution of (1). To do this, we show that it can be recast as the inertial-relaxed separator-projection method proposed in [4], and we use the convergence properties of this latter method. Also, we study the iteration-complexity of the proposed inertial projective splitting method. By considering a notion of approximate solution for problem (1) in terms of the ε-enlargements of the operators A and B, we obtain $\mathcal{O}(1/\sqrt{k})$ *pointwise* and $\mathcal{O}(1/k)$ *ergodic* convergence rates (iteration-complexity) for our method. Up to our knowledge, this is the first time that iteration complexities of inertial projective splitting-like algorithms are analyzed.

One of the main questions regarding inexact projective splitting methods is how to inexactly calculate at each iteration the proximal subproblems associated with each operator. Typically, some computational scheme is used to solve the corresponding problem iteratively until the imposed error criterion is satisfied. However, for the case where the operators are Lipschitz continuous and an estimation of the Lipschitz constants is available, the subproblem calculations can be replaced by two appropriate evaluations of the operators. These evaluations, also called *forward-steps*, were proposed in [23] for handling any Lipschitz continuous operator in the non-inertial (parallel) projective splitting method. We extend this procedure to our inertial (non-parallel) projective splitting algorithm for solving (1). Also, if the Lipschitz constant is unknown, we discuss a backtracking procedure that can be used, which returns an inexact solution of the proximal subproblems in the sense we consider here.

The remainder of this paper is organized as follows. Section 2 reviews the definitions and some basic properties of maximal monotone operators and their ε-enlargements. This section also introduces the inertial-relaxed separator projection algorithm of [4] and discusses its convergence properties. Section 3 presents our inertial inexact projective splitting method for solving (1) and establishes its convergence and iteration-complexity. Section 4 discusses how to compute inexact solutions of the proximal subproblems in our projective splitting method for the case where the operators are Lipschitz continuous. Finally, Section 5 applies our method to two common test problems and exhibits its computational performance. The main goal of these numerical experiments is to compare the inertial and the non-inertial projective splitting methods.

2 Mathematical preliminaries

2.1 Notation and basic results

Let $\mathcal{E}$ denote a real Hilbert space with inner product and induced norm denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$, respectively. The convergence in the weak topology of $\mathcal{E}$ is denoted by $\rightharpoonup$. The orthogonal projection of a point $x \in \mathcal{E}$ onto a closed convex set $S \subset \mathcal{E}$ is denoted by $P_S(x) := \operatorname{argmin}\{\|x-y\| \mid y \in S\}$, and the distance from x to S is denoted by $d(x,S) := \|x - P_S(x)\|$. We indicate by $\mathbb{R}^+$ the set of non-negative real numbers.

A *point-to-set* operator $T : \mathcal{E} \rightrightarrows \mathcal{E}$ is a function of $\mathcal{E}$ into the family $\wp(\mathcal{E}) = 2^{\mathcal{E}}$ of subsets of $\mathcal{E}$. An operator $T : \mathcal{E} \rightrightarrows \mathcal{E}$ is *monotone* if

$$\langle x' - x, v' - v \rangle \geq 0, \quad \forall\, v \in T(x), v' \in T(x').$$

On the other hand, T is *maximal monotone* if it is monotone and its graph $\operatorname{Gr}(T) := \{(x,v) \in \mathcal{E} \times \mathcal{E} \mid v \in T(x)\}$ is not properly contained in the graph of any other monotone operator. The inverse of T is the point-to-set operator $T^{-1} : \mathcal{E} \rightrightarrows \mathcal{E}$, defined at any $v \in \mathcal{E}$ by $x \in T^{-1}(v)$ if and only if $v \in T(x)$. If T is maximal monotone and $\lambda > 0$, the *resolvent mapping* (or *proximal mapping*) associated with T, $(\lambda T + I)^{-1} : \mathcal{E} \to \mathcal{E}$ where I is the identity mapping, is everywhere defined and single-valued [28]. It follows directly from the definition that $z' = (\lambda T + I)^{-1}(z)$ if and only if z' is the solution of the *proximal subproblem*

$$0 \in \lambda T(z') + (z' - z). \tag{3}$$

If $T : \mathcal{E} \rightrightarrows \mathcal{E}$ is maximal monotone and $\varepsilon \geq 0$, the *ε-enlargement* [11, 29] of T is the operator $T^{\varepsilon} : \mathcal{E} \rightrightarrows \mathcal{E}$ defined by

$$T^{\varepsilon}(x) := \{v \in \mathcal{E} \mid \langle x - \tilde{x}, v - \tilde{v} \rangle \geq -\varepsilon,\ \forall (\tilde{x}, \tilde{v}) \in \operatorname{Gr}(T)\}, \quad \forall\, x \in \mathcal{E}.$$

The following proposition presents some useful and well-known properties of monotone operators and their ε-enlargements (see, for example, [11, 12]).

Proposition 1. *Let $T, T' : \mathcal{E} \rightrightarrows \mathcal{E}$ be maximal monotone operators, then*

(i) $\left(T^{-1}\right)^{\varepsilon} = \left(T^{\varepsilon}\right)^{-1}$ *for all* $\varepsilon \geq 0$;

(ii) $T^{\varepsilon}(x) + (T')^{\varepsilon'}(x) \subset (T + T')^{\varepsilon + \varepsilon'}(x)$ *for every* $x \in \mathcal{E}$ *and* $\varepsilon, \varepsilon' \in \mathbb{R}^+$;

(iii) *if* $v \in T^{\varepsilon}(x)$ *and* $v' \in (T')^{\varepsilon'}(x')$, *then* $(v, v') \in (T \times T')^{\varepsilon + \varepsilon'}(x, x')$;

(iv) *if* $\{(p^k, v^k, \varepsilon_k)\}$ *is such that* $v^k \in T^{\varepsilon_k}(p^k)$ *for each* k, *and* $p^k \rightharpoonup p$, $\lim_{k \to +\infty} v^k = v$ *and* $\lim_{k \to +\infty} \varepsilon_k = \varepsilon$, *then* $v \in T^{\varepsilon}(p)$.

We now state a *weak transportation formula* for computing points in the graph of T^{ε}. A proof of the following result can be found in [12].

Theorem 1. *Suppose that $T : \mathcal{E} \rightrightarrows \mathcal{E}$ is maximal monotone. Let $x^l \in \mathcal{E}$, $v^l \in \mathcal{E}$, and ε_l, $\alpha_l \in \mathbb{R}^+$, $l = 1, \dots, k$, be such that*

$$v^l \in T^{\varepsilon_l}(x^l),\ l = 1, \dots, k, \qquad \sum_{l=1}^{k} \alpha_l = 1,$$

and define

$$\hat{x} := \sum_{l=1}^{k} \alpha_l x^l, \qquad \hat{v} := \sum_{l=1}^{k} \alpha_l v^l, \qquad \hat{\varepsilon} := \sum_{l=1}^{k} \alpha_l \left(\varepsilon_l + \langle x^l - \hat{x}, v^l - \hat{v} \rangle\right) = \sum_{l=1}^{k} \alpha_l \left(\varepsilon_l + \langle x^l - \hat{x}, v^l \rangle\right).$$

Then, $\hat{\varepsilon} \geq 0$ and $\hat{v} \in T^{\hat{\varepsilon}}(\hat{x})$.

The following technical results will be useful in our work.

Lemma 1. *Consider the symmetric matrix*

$$M = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$$

and assume that $\operatorname{trace}(M) > 0$. *If* τ *is the smallest eigenvalue of* M, *then* $\tau \geq \dfrac{\det(M)}{\operatorname{trace}(M)}$.

Proof. Simple calculations show that $\operatorname{trace}(M)^2 - 4\det(M) = (a_{11} - a_{22})^2 + 4a_{12}^2 \geq 0$. Therefore, it is clear that

$$\tau = \frac{1}{2}\left(\operatorname{trace}(M) - \sqrt{\operatorname{trace}(M)^2 - 4\det(M)}\right). \tag{4}$$

Applying the inequality $\sqrt{a+b} \leq \sqrt{a} + \dfrac{b}{2\sqrt{a}}$ for $a > 0$, we obtain

$$\sqrt{\operatorname{trace}(M)^2 - 4\det(M)} \leq \operatorname{trace}(M) - \frac{2\det(M)}{\operatorname{trace}(M)}.$$

Thus, substituting the above equation into (4), we conclude the proof. □

Lemma 2. [3, Lemma 7] *Let* $\{h_k\}$, $\{s_k\}$, $\{\eta_k\}$ *and* $\{\delta_k\}$ *be sequences in* $[0, +\infty)$ *such that there exists* $\eta \in \mathbb{R}$ *with* $0 \leq \eta_k \leq \eta < 1$, $h_0 = h_{-1}$ *and*

$$h_{k+1} - h_k + s_{k+1} \leq \eta_k(h_k - h_{k-1}) + \delta_k, \qquad \forall k \geq 0.$$

Then, the following hold:

(i) *For all* $k \geq 1$,

$$h_k + \sum_{j=1}^{k} s_j \leq h_0 + \frac{1}{1-\eta}\sum_{j=0}^{k-1}\delta_j. \tag{5}$$

(ii) *If* $\sum_{k=0}^{\infty}\delta_k \leq +\infty$, *then* $\lim_{k\to\infty} h_k$ *exists.*

For all $p, q \in \mathcal{E}$ and $t \in \mathbb{R}$, it holds that

$$\|tp + (1-t)q\|^2 = t\|p\|^2 + (1-t)\|q\|^2 - t(1-t)\|p-q\|^2. \tag{6}$$

2.2 An inertial-relaxed separator-projection method

In this section, we review the separator projection framework (Algorithm 1 below) presented in [4] and its convergence properties. Algorithm 1 is a general-separator projection method for finding a point in some closed and convex subset of a Hilbert space, with the feature of considering inertial steps. Hence, by reformulating problem (1) as the convex feasibility problem defined by a certain closed and convex extended solution set, this framework can be used to study inertial algorithms for solving (1). In particular, Algorithm 1 will be used in Section 3 to analyze the inertial-relaxed projective splitting method proposed in this work to solve (1).

Algorithm 1. Let $\mathcal{E}$ be a Hilbert space and $S \subset \mathcal{E}$ be some closed and convex subset. Start with an arbitrary $p^0 = p^{-1} \in \mathcal{E}$, $\eta \in [0,1)$ and $0 < \underline{\beta} < \overline{\beta} < 2$. For $k = 0, 1, \ldots$

1. Choose $\eta_k \in [0, \eta]$ and set
$$\overline{p}^k = p^k + \eta_k(p^k - p^{k-1}). \tag{7}$$
2. Find an affine function φ_k such that $\nabla\varphi_k \neq 0$ and $\varphi_k(p) \leq 0$ for all $p \in S$.
3. Choose $\beta_k \in [\underline{\beta}, \overline{\beta}]$ and set
$$p^{k+1} = \overline{p}^k - \beta_k \frac{\max\{0, \varphi_k(\overline{p}^k)\}}{\|\nabla\varphi_k\|^2}\nabla\varphi_k.$$

Remark 1.

(i) The inertial effect of Algorithm 1 is produced by the extrapolation step in (7), where $\eta_k \geq 0$ controls the magnitude of the extrapolation in the direction of the vector $p^k - p^{k-1}$. If $\eta_k \equiv 0$, Algorithm 1 reduces to a generic linear separator-projection method for finding a point in $S \subset \mathcal{E}$ [7].

(ii) The update in step 3 is the β_k-relaxed projection onto the halfspace $\{p \in \mathcal{E} : \varphi_k(p) \leq 0\}$. That is, if $\tilde{p}^{k+1}$ is the orthogonal projection of $\overline{p}^k$ onto this halfspace:

$$\tilde{p}^{k+1} = \overline{p}^k - \frac{\max\{0, \varphi_k(\overline{p}^k)\}}{\|\nabla\varphi_k\|^2}\nabla\varphi_k, \tag{8}$$

then the update is $p^{k+1} = (1-\beta_k)\overline{p}^k + \beta_k\tilde{p}^{k+1}$.

The following lemma will be useful in our convergence and complexity analyses. For its proof we refer the reader to [4, Lemma 2.1].

Lemma 3. [4, Lemma 2.1(a)] *Let $\{p^k\}$, $\{\overline{p}^k\}$, $\{\eta_k\}$, $\{\beta_k\}$ be the sequences generated by Algorithm 1. For any $p^* \in S$ and $k \geq -1$ define*

$$h_k := \|p^k - p^*\|^2. \tag{9}$$

Then, for all $k \geq 0$

$$h_{k+1} - h_k - \eta_k(h_k - h_{k-1}) \leq \delta_k - s_{k+1}, \tag{10}$$

where

$$s_{k+1} := \beta_k(2-\beta_k)\|\overline{p}^k - \tilde{p}^{k+1}\|^2 \quad \textit{and} \quad \delta_k := \eta_k(1+\eta_k)\|p^k - p^{k-1}\|^2. \tag{11}$$

Inequality (10) plays a role in the convergence analysis of inertial proximal algorithms similar to that played by Fejér-monotonicity in the analysis of standard proximal algorithms. Further, we observe that if $\eta_k \equiv 0$, in which case Algorithm 1 is a generic linear-separator projection method for finding a point in $S \subset \mathcal{E}$, then equation (10) reduces to

$$\|p^{k+1} - p^*\|^2 - \|p^k - p^*\|^2 \leq -\beta_k(2-\beta_k)\|p^k - \tilde{p}^{k+1}\|^2,$$

which indeed guarantees the Fejér-monotonicity of Algorithm 1.

Next, we present the main results on the asymptotic convergence of Algorithm 1 obtained in [4]. The key assumption is the summability condition (12) below, for which sufficient conditions (13) and (14) on the inertial and relaxation parameters η_k and β_k are given in Theorem 3.

Theorem 2. [4, Theorem 2.2] *Let $\{p^k\}$, $\{\overline{p}^k\}$, $\{\varphi_k\}$, $\{\eta_k\}$, $\{\beta_k\}$ be generated by Algorithm 1 and assume that*

$$\sum_{k=0}^{\infty} \eta_k \|p^k - p^{k-1}\|^2 < \infty. \tag{12}$$

Then, the following hold:

(i) *$\{p^k\}$ and $\{\overline{p}^k\}$ are bounded sequences;*

(ii) *if every cluster point of $\{p^k\}$ belongs to S, then $\{p^k\}$ converges weakly to an element of S;*

(iii) $\dfrac{\max\{0, \varphi_k(\overline{p}^k)\}}{\|\nabla\varphi_k\|} \to 0.$

Theorem 3. [4, Theorem 2.3] *Let $\{p^k\}$ and $\{\eta_k\}$ be generated by Algorithm 1. Suppose that $\eta \in [0,1)$, $\overline{\beta} \in (0,2)$, the sequence $\{\eta_k\}$ satisfies, for some $\overline{\eta} > 0$,*

$$0 \leq \eta_k \leq \eta_{k+1} \leq \eta < \overline{\eta} < 1 \tag{13}$$

and

$$\overline{\beta} = \overline{\beta}(\overline{\eta}) := \frac{2(\overline{\eta}-1)^2}{2(\overline{\eta}-1)^2 + 3\overline{\eta} - 1}. \tag{14}$$

Then, we have for any $p^* \in S$ *and all* $k \geq 0$

$$\sum_{j=0}^{k} \left\| p^{j+1} - p^j \right\|^2 \leq \frac{2-\eta}{q(\eta)(1-\eta)} \left\| p^0 - p^* \right\|^2, \tag{15}$$

where $q(\eta) := 2(\overline{\beta}^{-1} - 1)\eta^2 - (4\overline{\beta}^{-1} - 1)\eta + 2\overline{\beta} - 1 > 0$. *In particular, we have that*

$$\sum_{k=0}^{\infty} \left\| p^k - p^{k-1} \right\|^2 < \infty.$$

The inequality in (15) can be obtained as a consequence of the proof of Theorem 2.3 in [4]. Conditions (13) and (14) guarantee that the summability condition (12) is satisfied, thus Theorem 2 holds. Since Algorithm 1 is the basis for studying the properties of the inertial projective splitting method developed in this work, these conditions will also play an important role in its convergence analysis.

We observe that if we set $\overline{\eta} = 1/3$ in (13), then by (14) it follows that $\overline{\beta} = 1$. Conversely, the overrelaxation effects in Algorithm 1 can be achieved at the price of choosing the inertial parameter upper bound $\overline{\eta}$ strictly smaller than $1/3$. For more details on the relation between inertial and relaxation parameters we refer the reader to [6, 3, 4].

3 An inertial projective splitting algorithm

In this section, we present an inertial projective splitting method to solve the monotone inclusion problem (1). It is well known [18] that (1) can be reformulated in terms of the convex feasibility problem of finding a point in the *extended solution set*

$$S_e := \{(z,w) \in \mathcal{X} \times \mathcal{Y} \mid -G^*w \in B(z), w \in A(Gz)\}.$$

Indeed, $z \in \mathcal{X}$ is a solution of (1) if and only if there exists $w \in \mathcal{Y}$ such that $(z,w) \in S_e$ [18, Lemma 1]. And also, S_e is a closed and convex subset of $\mathcal{X} \times \mathcal{Y}$ [18, Lemma 1]. Therefore, to find a solution to (1), we can apply the framework presented in Algorithm 1. To do this, we need to construct at each iteration an affine function φ_k in $\mathcal{X} \times \mathcal{Y}$ such that $\varphi_k(z,w) \leq 0$ for all $(z,w) \in S_e$. We perform this construction by inexactly solving two proximal subproblems subject to a relative-error criterion, each involving only one of the operators A or B, ensuring the splitting nature of the method. The resulting scheme is Algorithm 2 below, which is a relative-error inertial-relaxed projective splitting algorithm for solving (1).

We now present the notion of approximate solution of a proximal subproblem used in our inertial projective splitting algorithm, which was introduced in [34]. For this purpose, we first observe that given a Hilbert space $\mathcal{H}$ a maximal monotone operator $T : \mathcal{H} \rightrightarrows \mathcal{H}$, $\lambda > 0$ and $z \in \mathcal{H}$, the proximal subproblem (3) is equivalent to the *proximal system*

$$\begin{cases} w \in T(z'), \\ \lambda w + z' - z = 0. \end{cases} \tag{16}$$

Definition 1. Given $\sigma \in [0,1)$, a triplet $(z', w, \varepsilon) \in \mathcal{H} \times \mathcal{H} \times \mathbb{R}^+$ is called a σ*-approximate solution* of (16) at (λ, z), if

$$w \in T^{\varepsilon}(z') \quad \text{and} \quad \left\| \lambda w + z' - z \right\|^2 + 2\lambda\varepsilon \leq \sigma^2 \left(\left\| \lambda w \right\|^2 + \left\| z' - z \right\|^2 \right). \tag{17}$$

We note that if (z', w) is the exact solution of (16), then, taking $\varepsilon = 0$, the triplet (z', w, ε) satisfies the approximation criterion (17) for all $\sigma \in [0,1)$. On the other hand, if $\sigma = 0$, only the exact solution of (16), with $\varepsilon = 0$, will satisfy (17). Additionally, we observe that the equality in (16) is relaxed by introducing a relative-error condition and the inclusion is relaxed by means of the ε-enlargement of T.

The following lemma presents some basic properties of the inexact solutions of the proximal system (16) in the sense of Definition 1 that will be useful in our work.

Lemma 4. [34, Lemma 2] *Let $z \in \mathcal{H}$, $\lambda > 0$ and $\sigma \in [0,1)$ be given. The triplet (z', w, ε) is a σ-approximate solution of the proximal system (16) at (z, λ) if and only if*

$$\varepsilon \geq 0, \qquad w \in T^{\varepsilon}(z'), \qquad \langle w, z - z' \rangle - \varepsilon \geq \frac{1-\sigma^2}{2\lambda}\left(\|z' - z\|^2 + \|\lambda w\|^2\right). \tag{18}$$

In addition, we have

$$\left(\frac{1-\sigma^2}{1+\bar{\sigma}}\right)\|z' - z\| \leq \|\lambda w\| \leq \left(\frac{1+\bar{\sigma}}{1-\sigma^2}\right)\|z' - z\|, \tag{19}$$

where $\bar{\sigma} := \sqrt{1-(1-\sigma^2)^2}$.

From now on, we consider the Hilbert space $\mathcal{E} = \mathcal{X} \times \mathcal{Y}$ with the inner product and associated norm

$$\langle (z,w),(z',w') \rangle = \langle z, z' \rangle + \langle w, w' \rangle, \qquad \|(z,w)\|^2 = \|z\|^2 + \|w\|^2.$$

Next, we present our relative-error inertial-relaxed projective splitting algorithm to solve (1).

Algorithm 2. Start with an arbitrary $(z^0, w^0) = (z^{-1}, w^{-1}) \in \mathcal{E}$, $\eta \in [0,1)$, $0 < \underline{\beta} < \overline{\beta} < 2$ and $\sigma \in [0,1)$. For $k = 0, 1, \ldots$

1. Choose $\eta_k \in [0, \eta]$ and set
$$\begin{aligned} \overline{z}^k &= z^k + \eta_k(z^k - z^{k-1}), \\ \overline{w}^k &= w^k + \eta_k(w^k - w^{k-1}). \end{aligned} \tag{20}$$

2. Choose $\lambda_k > 0$ and find $(x^k, b^k, \varepsilon_k^x)$ such that
$$\begin{aligned} &b^k \in B^{\varepsilon_k^x}(x^k), \qquad \lambda_k b^k + x^k = \overline{z}^k - \lambda_k G^* \overline{w}^k + r^{x,k}, \\ &\left\|r^{x,k}\right\|^2 + 2\lambda_k \varepsilon_k^x \leq \sigma^2\left(\left\|x^k - \overline{z}^k\right\|^2 + \left\|\lambda_k(b^k + G^*\overline{w}^k)\right\|^2\right). \end{aligned} \tag{21}$$

3. Choose $\mu_k > 0$, $\alpha_k \in \mathbb{R}$, and find $(y^k, a^k, \varepsilon_k^y)$ such that
$$\begin{aligned} &a^k \in A^{\varepsilon_k^y}(y^k), \qquad \mu_k a^k + y^k = G((1-\alpha_k)\overline{z}^k + \alpha_k x^k) + \mu_k \overline{w}^k + r^{y,k}, \\ &\left\|r^{y,k}\right\|^2 + 2\mu_k \varepsilon_k^y \leq \sigma^2\left(\left\|y^k - G\overline{z}^k\right\|^2 + \left\|\alpha_k G(\overline{z}^k - x^k) + \mu_k(a^k - \overline{w}^k)\right\|^2\right). \end{aligned} \tag{22}$$

4. If $\left\|y^k - Gx^k\right\| = \left\|b^k + G^* a^k\right\| = 0$ stop. Otherwise, set
$$\gamma_k = \frac{\left\langle \overline{z}^k - x^k, b^k + G^*\overline{w}^k \right\rangle + \left\langle G\overline{z}^k - y^k, a^k - \overline{w}^k \right\rangle - \varepsilon_k^x - \varepsilon_k^y}{\|b^k + G^* a^k\|^2 + \|y^k - Gx^k\|^2}. \tag{23}$$

5. Choose $\beta_k \in [\underline{\beta}, \overline{\beta}]$ and set
$$\begin{aligned} z^{k+1} &= \overline{z}^k - \beta_k \gamma_k (b^k + G^* a^k), \\ w^{k+1} &= \overline{w}^k - \beta_k \gamma_k (y^k - Gx^k). \end{aligned} \tag{24}$$

Remark 2.

(i) Similar to Algorithm 1, Algorithm 2 includes inertial and relaxation effects. The inertial step is performed in (20) and controlled by the parameter η_k, whereas the relaxation step takes place in (24) and is controlled by β_k.

(ii) At each iteration k, the triplet $(x^k, b^k, \varepsilon_k^x)$ calculated in step 2 of Algorithm 2 is a σ-approximate solution of (16) at $(\lambda_k, \overline{z}_k)$, where $T = B + G^*\overline{w}_k$. Similarly, $(y^k, a^k, \varepsilon_k^y)$ is a σ-approximate solution of (16) (with $T = A + \alpha_k/\mu_k G(\overline{z}^k - x^k) - \overline{w}^k)$ at $(\mu_k, G\overline{z}^k)$. Therefore, from the comments after Definition 1, there exists at least one triplet satisfying (21) (resp. (22)). Also, by Lemma 4, for each $k \geq 0$, the error criteria (21) and (22) imply, respectively,

$$\frac{1-\sigma^2}{1+\overline{\sigma}} \left\|\overline{z}^k - x^k\right\| \leq \lambda_k \left\|b^k + G^*\overline{w}^k\right\| \leq \frac{1+\overline{\sigma}}{1-\sigma^2} \left\|\overline{z}^k - x^k\right\| \tag{25}$$

and

$$\frac{1-\sigma^2}{1+\overline{\sigma}} \left\|G\overline{z}^k - y^k\right\| \leq \left\|\alpha_k G(\overline{z}^k - x^k) + \mu_k(a^k - \overline{w}^k)\right\| \leq \frac{1+\overline{\sigma}}{1-\sigma^2} \left\|G\overline{z}^k - y^k\right\|. \tag{26}$$

(iii) Algorithm 2 does not specify how to find the triplets $(x^k, b^k, \varepsilon_k^x)$ and $(y^k, a^k, \varepsilon_k^y)$ satisfying steps 2 and 3, respectively. The procedures used to carry out these computations will depend on the implementation of the method and the properties of the operators A and B. Usually, the error conditions (21) and (22) can be used as stopping criteria for some computational scheme used to iteratively solving the corresponding proximal subproblem. In the case where the operators are Lipschitz continuous, we discuss in Section 4 below how steps 2 and 3 could be implemented in order to produce such triplets.

We refer the reader to [27, Section 4], where the calculation of inexact solutions of proximal subproblems in the sense of Definition 1 is discussed in the context of optimization, i.e. when each operator is the subdifferential of a proper convex lower-semicontinuous function.

(iv) If $\eta_k \equiv 0$ (that is, there is no inertia), then Algorithm 2 is an inexact version of the projective splitting method proposed in [18]. We observe that for this case, Algorithm 2 is different from the inexact projective splitting method studied in [19] when $n = 2$. Indeed, the error criterion used in that work is a generalization for the case of $n \geq 2$ maximal monotone operators of the relative error tolerance of the hybrid proximal extragradient method [33] (without the use of ε-enlargements), while we used the more flexible error condition introduced in [34]. Also, for $n = 2$, the projective splitting method proposed in [32] is different from Algorithm 2 without inertia, even for the case $\alpha_k \equiv 0$, see [32, Section 4]. Furthermore, the convergence and complexity analyses of [32] are based on the approximate proximal point method of [34] whereas the convergence and complexity properties of Algorithm 2 are derived using Algorithm 1, a separator projection method.

(v) When $\alpha_k \equiv 0$, Algorithm 2 is a *parallel* inexact inertial projective method, since at each iteration it could process the operators independently. For this case, Algorithm 2 is closely related to the inertial projective splitting method proposed in [4] (for $n = 2$). The main difference between both methods is that Algorithm 2 allows the use of elements in the ε-enlargements of the operators for the inexact computation of the proximal subproblems, different from the algorithm studied in [4].

(vi) It will be proven (see Remark 3 below) that if Algorithm 2 stops in step 4, then $(x^k, a^k) \in S_e$ and therefore x^k is a solution of (1). Hence, from now on, we assume that Algorithm 2 never stops in step 4, generating an infinite sequence of iterates.

Next, we show that Algorithm 2 is an instance of Algorithm 1 in the sense that any sequence generated by the first method can be seen as a sequence generated by the second method for finding a point in S_e. For this purpose, we first construct the nonconstant affine functions φ_k such that $\varphi_k(z^*, w^*) \leq 0$ for all $(z^*, w^*) \in S_e$ required in step 2 of Algorithm 1.

Lemma 5. *Consider the sequences* $\{(x^k, b^k, \varepsilon_k^x)\}$ *and* $\{(y^k, a^k, \varepsilon_k^y)\}$ *generated by Algorithm 2, and for each* $k \geq 0$ *define the function* $\varphi_k : \mathcal{E} \to \mathbb{R}$ *by*

$$\varphi_k(z, w) := \langle z - x^k, b^k + G^* w \rangle + \langle Gz - y^k, a^k - w \rangle - \varepsilon_k^x - \varepsilon_k^y. \tag{27}$$

Then, the following hold:

(i) φ_k *is affine in* $\mathcal{E}$ *and* $\nabla \varphi_k = (b^k + G^* a^k, y^k - Gx^k)$*;*

(ii) $\varphi_k(p^*) \leq 0$ *for all* $p^* \in S_e$*.*

Proof. By adding and subtracting $\langle Gx^k, a^k - w \rangle$ on the right-hand side of (27), we can rewrite $\varphi_k(z, w)$ as

$$\varphi_k(z, w) = \langle z - x^k, b^k + G^* a^k \rangle + \langle Gx^k - y^k, a^k - w \rangle - \varepsilon_k^x - \varepsilon_k^y. \tag{28}$$

From above it is easy to see that φ_k is affine in $\mathcal{E}$ and $\nabla \varphi_k = (b^k + G^* a^k, y^k - Gx^k)$, proving item (i). Item (ii) is a direct consequence of (27) and the definitions of the ε-enlargement of a point-to-set operator and the set S_e. □

Proposition 2. *Let the sequences* $\{(z^k, w^k)\}$*,* $\{(\overline{z}^k, \overline{w}^k)\}$*,* $\{\eta_k\}$ *and* $\{\beta_k\}$ *be generated by Algorithm 2 and* $\{\varphi_k\}$ *given in (27). Define, for each* $k \geq -1$*,*

$$p^k = (z^k, w^k) \quad \text{and} \quad \overline{p}^k = (\overline{z}^k, \overline{w}^k). \tag{29}$$

Then, the following hold:

(i) *for all* $k \geq 0$*,* $\nabla \varphi_k \neq 0$ *and* $\varphi_k(p^*) \leq 0$ *for all* $p^* \in S_e$*;*

(ii) *for all* $k \geq 0$*,* $\overline{p}^k = p^k + \eta_k(p^k - p^{k-1})$ *and*

$$p^{k+1} = \overline{p}^k - \beta_k \frac{\max\{0, \varphi_k(\overline{p}^k)\}}{\|\nabla \varphi_k\|^2} \nabla \varphi_k. \tag{30}$$

As a consequence, Algorithm 2 is a special instance of Algorithm 1 for finding a point in S_e*.*

Proof. Item (i) is a direct consequence of the assumption that Algorithm 2 does not stop at step 4 (see Remark 2(vi)) and Lemma 5. The fact that $\overline{p}^k = p^k + \eta_k(p^k - p^{k-1})$ is due to the definitions of p^k and $\overline{p}^k$ in (29), and step 1 of Algorithm 2. Equation (30) follows combining (23) with (27), Lemma 5(i), (24) and (29).

Finally, the last statement of the proposition is a consequence of items (i) and (ii), and the definition of Algorithm 1. □

From now on, we consider the following assumptions on Algorithm 2:

(A1) the extended solution set S_e is nonempty;

(A2) there exist $\underline{\lambda}$ and $\overline{\lambda}$ such that $\overline{\lambda} \geq \underline{\lambda} > 0$ and $\lambda_k, \mu_k \in [\underline{\lambda}, \overline{\lambda}]$ for all $k \geq 0$;

(A3) $\nu := \inf_{k \geq 0} \left\{ \left(\frac{1 - \sigma^2}{1 + \overline{\sigma}} \right)^2 \cdot \frac{\mu_k}{\lambda_k} - \left(\frac{\alpha_k \|G\|}{2} \right)^2 \right\} > 0$, where $\overline{\sigma}$ is as in Lemma 4.

We observe that if $\sigma = 0$, $\mathcal{X} = \mathcal{Y}$ and $G = I$, condition (A3) is the same hypothesis considered in [18] for proving the convergence of the non-inertial projective splitting method. Also, if $\alpha_k \equiv 0$, in which case Algorithm 2 is reduced to a *parallel* inertial projective splitting method, then (A.3) trivially holds. In general, (A.3) may be ensured by any sufficient condition bounding the absolute values $|\alpha_k|$. For example, if there exists $\overline{\alpha} > 0$ such that $|\alpha_k| \leq \overline{\alpha}$ for all $k \geq 0$, then taking $\overline{\lambda}$ and $\underline{\lambda}$ in (A2) such that

$$\sqrt{\frac{\underline{\lambda}}{\overline{\lambda}}} > \frac{(1 + \overline{\sigma})\overline{\alpha} \|G\|}{2(1 - \sigma^2)},$$

we have that (A3) holds.

3.1 Convergence analysis

In this section, we study the asymptotic behavior of the sequences generated by Algorithm 2. Because of Proposition 2, Algorithm 2 is an instance of Algorithm 1. Therefore, by Theorem 2, under assumption (12), we will obtain the global convergence of Algorithm 2 by guaranteeing that every cluster point of the sequence $\{p^k\}$ defined in (29) belongs to S_e. Before establishing these convergence results, we need the following proposition, which presents several lower bounds for $\varphi_k(\overline{z}^k, \overline{w}^k)$.

Proposition 3. *Consider the sequences $\{(\overline{z}^k, \overline{w}^k)\}$, $\{(x^k, b^k, \varepsilon_k^x)\}$ and $\{(y^k, a^k, \varepsilon_k^y)\}$ generated by Algorithm 2 and $\{\varphi_k\}$ given in (27). Then, the following inequalities hold:*

$$\varphi_k(\overline{z}^k, \overline{w}^k) \geq \zeta \left(\left\| \overline{z}^k - x^k \right\|^2 + \left\| G\overline{z}^k - y^k \right\|^2 \right); \tag{31}$$

$$\varphi_k(\overline{z}^k, \overline{w}^k) \geq \frac{\zeta}{C} \left(\left\| b^k + G^*\overline{w}^k \right\|^2 + \left\| a^k - \overline{w}^k \right\|^2 \right); \tag{32}$$

$$\varphi_k(\overline{z}^k, \overline{w}^k) \geq \frac{\zeta \underline{\lambda}(1-\sigma^2)^2}{\sigma^2(1+\overline{\sigma})} \left(\varepsilon_k^x + \varepsilon_k^y \right); \tag{33}$$

$$\varphi_k(\overline{z}^k, \overline{w}^k) \geq \frac{\zeta}{2D(C+1)} \left\| \nabla \varphi_k \right\|^2; \tag{34}$$

where $\zeta := \dfrac{(1+\overline{\sigma})\underline{\lambda}\nu}{2(1-\sigma^2)\overline{\lambda}^2}$, $C := \dfrac{2}{\underline{\lambda}^2} \left[\left(\dfrac{1+\overline{\sigma}}{1-\sigma^2} \right)^2 + \dfrac{4\overline{\lambda}}{\underline{\lambda}} \right]$, *and* $D := \max\{1, \|G\|^2\}$.

Proof. First, we use the definition of φ_k in (27) and perform simple manipulations to obtain

$$\begin{aligned}
\varphi_k(\overline{z}^k, \overline{w}^k) =& \frac{1}{2\lambda_k} \left[2\lambda_k \left\langle \overline{z}^k - x^k, b^k + G^*\overline{w}^k \right\rangle - 2\lambda_k \varepsilon_k^x \right] + \frac{1}{2\mu_k} \left[2\mu_k \left\langle G\overline{z}^k - y^k, a^k - \overline{w}^k \right\rangle - 2\mu_k \varepsilon_k^y \right] \\
=& \frac{1}{2\lambda_k} \left[\left\| \overline{z}^k - x^k \right\|^2 + \left\| \lambda_k (b^k + G^*\overline{w}^k) \right\|^2 - \left\| r^{x,k} \right\|^2 - 2\lambda_k \varepsilon_k^x \right] \\
&+ \frac{1}{2\mu_k} \left[2 \left\langle G\overline{z}^k - y^k, \alpha_k G(\overline{z}^k - x^k) + \mu_k (a^k - \overline{w}^k) \right\rangle - 2\alpha_k \left\langle G\overline{z}^k - y^k, G(\overline{z}^k - x^k) \right\rangle - 2\mu_k \varepsilon_k^y \right] \\
=& \frac{1}{2\lambda_k} \left\| \overline{z}^k - x^k \right\|^2 + \frac{1}{2\lambda_k} \left\| \lambda_k (b^k + G^*\overline{w}^k) \right\|^2 - \frac{1}{2\lambda_k} \left[\left\| r^{x,k} \right\|^2 + 2\lambda_k \varepsilon_k^x \right] - \frac{\alpha_k}{\mu_k} \left\langle G\overline{z}^k - y^k, G(\overline{z}^k - x^k) \right\rangle \\
&+ \frac{1}{2\mu_k} \left\| G\overline{z}^k - y^k \right\|^2 + \frac{1}{2\mu_k} \left\| \alpha_k G(\overline{z}^k - x^k) + \mu_k (a^k - \overline{w}^k) \right\|^2 - \frac{1}{2\mu_k} \left[\left\| r^{y,k} \right\|^2 + 2\mu_k \varepsilon_k^y \right].
\end{aligned}$$

Now, by the error criteria (21) and (22), we have

$$\begin{aligned}
\varphi_k(\overline{z}^k, \overline{w}^k) \geq& \frac{1-\sigma^2}{2\lambda_k} \left\| \overline{z}^k - x^k \right\|^2 + \frac{1-\sigma^2}{2\lambda_k} \left\| \lambda_k (b^k + G^*\overline{w}^k) \right\|^2 - \frac{\alpha_k}{\mu_k} \left\langle G\overline{z}^k - y^k, G(\overline{z}^k - x^k) \right\rangle \\
&+ \frac{1-\sigma^2}{2\mu_k} \left\| G\overline{z}^k - y^k \right\|^2 + \frac{1-\sigma^2}{2\mu_k} \left\| \alpha_k G(\overline{z}^k - x^k) + \mu_k (a^k - \overline{w}^k) \right\|^2.
\end{aligned}$$

Combining the above equation with the first inequalities in (25) and (26), we obtain

$$\begin{aligned}
\varphi_k(\overline{z}^k, \overline{w}^k) \geq& \frac{1-\sigma^2}{2\lambda_k} \left\| \overline{z}^k - x^k \right\|^2 + \frac{(1-\sigma^2)^3}{2\lambda_k (1+\overline{\sigma})^2} \left\| \overline{z}^k - x^k \right\|^2 - \frac{\alpha_k}{\mu_k} \left\langle G\overline{z}^k - y^k, G(\overline{z}^k - x^k) \right\rangle \\
&+ \frac{1-\sigma^2}{2\mu_k} \left\| G\overline{z}^k - y^k \right\|^2 + \frac{(1-\sigma^2)^3}{2\mu_k (1+\overline{\sigma})^2} \left\| G\overline{z}^k - y^k \right\|^2 \\
=& \frac{1-\sigma^2}{\lambda_k (1+\overline{\sigma})} \left\| \overline{z}^k - x^k \right\|^2 - \frac{\alpha_k}{\mu_k} \left\langle G\overline{z}^k - y^k, G(\overline{z}^k - x^k) \right\rangle + \frac{1-\sigma^2}{\mu_k (1+\overline{\sigma})} \left\| G\overline{z}^k - y^k \right\|^2 \\
\geq& \frac{1-\sigma^2}{\lambda_k (1+\overline{\sigma})} \left\| \overline{z}^k - x^k \right\|^2 - \frac{|\alpha_k| \, \|G\|}{\mu_k} \left\| G\overline{z}^k - y^k \right\| \left\| \overline{z}^k - x^k \right\| + \frac{1-\sigma^2}{\mu_k (1+\overline{\sigma})} \left\| G\overline{z}^k - y^k \right\|^2,
\end{aligned} \tag{35}$$

where the equality above follows from simple calculations and the definition of $\overline{\sigma}$ in Lemma 4, and the last inequality is a consequence of the Cauchy-Schwartz inequality.

Next, for each $k \geq 0$, we define the symmetric matrix

$$M_k := \begin{pmatrix} \dfrac{1-\sigma^2}{\lambda_k(1+\overline{\sigma})} & -\dfrac{|\alpha_k|\,\|G\|}{2\mu_k} \\ -\dfrac{|\alpha_k|\,\|G\|}{2\mu_k} & \dfrac{1-\sigma^2}{\mu_k(1+\overline{\sigma})} \end{pmatrix}$$

and let τ_k be the smallest eigenvalue of M_k. By interpreting the last expression in (35) as a quadratic form applied to $\left(\left\|\overline{z}^k - x^k\right\|, \left\|G\overline{z}^k - y^k\right\|\right) \in \mathbb{R}^2$, we have

$$\varphi_k(\overline{z}^k, \overline{w}^k) \geq \begin{pmatrix} \left\|\overline{z}^k - x^k\right\| \\ \left\|G\overline{z}^k - y^k\right\| \end{pmatrix}^T M_k \begin{pmatrix} \left\|\overline{z}^k - x^k\right\| \\ \left\|G\overline{z}^k - y^k\right\| \end{pmatrix} \geq \tau_k \left(\left\|\overline{z}^k - x^k\right\|^2 + \left\|G\overline{z}^k - y^k\right\|^2\right). \tag{36}$$

Since $\mathrm{trace}(M_k) > 0$, applying Lemma 1 to the matrix M_k, we obtain

$$\tau_k \geq \frac{\det(M_k)}{\mathrm{trace}(M_k)}.$$

Further, assumptions (A2) and (A3) imply

$$\det(M_k) = \left(\frac{1-\sigma^2}{1+\overline{\sigma}}\right)^2 \frac{1}{\lambda_k\mu_k} - \frac{\alpha_k^2\,\|G\|^2}{4\mu_k^2} = \frac{1}{\mu_k^2}\left(\left(\frac{1-\sigma^2}{1+\overline{\sigma}}\right)^2 \frac{\mu_k}{\lambda_k} - \frac{\alpha_k^2\,\|G\|^2}{4}\right) \geq \frac{1}{\overline{\lambda}^2}\nu$$

and

$$\mathrm{trace}(M_k) = \frac{1-\sigma^2}{1+\overline{\sigma}}\left(\frac{1}{\lambda_k} + \frac{1}{\mu_k}\right) \leq \left(\frac{1-\sigma^2}{1+\overline{\sigma}}\right)\frac{2}{\underline{\lambda}}.$$

Thus, combining the three equations above and using the definition of ζ, we have $\tau_k \geq \zeta$ for all $k \geq 0$, which together with (36) implies (31).

Next, we use the second inequality in (25) to obtain

$$\left\|b^k + G^*\overline{w}^k\right\|^2 \leq \frac{1}{\underline{\lambda}^2}\left(\frac{1+\overline{\sigma}}{1-\sigma^2}\right)^2 \left\|\overline{z}^k - x^k\right\|^2.$$

By the triangle inequality for norms and the second inequality in (26), we have

$$\begin{aligned} \mu_k \left\|a^k - \overline{w}^k\right\| &\leq \left\|\mu_k(a^k - \overline{w}^k) + \alpha_k G(\overline{z}^k - x^k)\right\| + \left\|\alpha_k G(\overline{z}^k - x^k)\right\| \\ &\leq \left(\frac{1+\overline{\sigma}}{1-\sigma^2}\right)\left\|G\overline{z}^k - y^k\right\| + |\alpha_k|\,\|G\|\,\left\|\overline{z}^k - x^k\right\|. \end{aligned}$$

Hence, squaring the inequality above, combining with the previous equation and performing simple manipulations yield

$$\begin{aligned} \left\|b^k + G^*\overline{w}^k\right\|^2 + \left\|a^k - \overline{w}^k\right\|^2 &\leq \frac{1}{\underline{\lambda}^2}\left[\left(\frac{1+\overline{\sigma}}{1-\sigma^2}\right)^2 + 2\alpha_k^2\,\|G\|^2\right]\left\|\overline{z}^k - x^k\right\|^2 + \frac{2}{\underline{\lambda}^2}\left(\frac{1+\overline{\sigma}}{1-\sigma^2}\right)^2\left\|G\overline{z}^k - y^k\right\|^2 \\ &\leq \frac{2}{\underline{\lambda}^2}\left[\left(\frac{1+\overline{\sigma}}{1-\sigma^2}\right)^2 + \frac{4\overline{\lambda}}{\underline{\lambda}}\right]\left(\left\|\overline{z}^k - x^k\right\|^2 + \left\|G\overline{z}^k - y^k\right\|^2\right), \end{aligned}$$

where in the second inequality above we used that $\alpha_k^2\,\|G\|^2 \leq 4\overline{\lambda}/\underline{\lambda}$, which is a consequence of assumptions (A2) and (A3), and the facts that $\sigma \in [0,1)$ and $\overline{\sigma} > 0$. Equation above, together with the definition of C and (31), now implies (32).

To prove (33), we observe that the error criterion in step 2 of Algorithm 2, together with the second inequality in (25) and the definition of $\overline{\sigma}$, yields

$$\varepsilon_k^x \leq \frac{\sigma^2}{2\lambda_k}\left(\left\|x^k - \overline{z}^2\right\|^2 + \frac{(1+\overline{\sigma})^2}{(1-\sigma^2)^2}\left\|x^k - \overline{z}^k\right\|^2\right) \leq \frac{\sigma^2(1+\overline{\sigma})}{\underline{\lambda}(1-\sigma^2)^2}\left\|x^k - \overline{z}^k\right\|^2.$$

Using the error criterion in step 3 of Algorithm 2 and the second inequality in (26), by a similar reasoning to the above, we obtain

$$\varepsilon_k^y \leq \frac{\sigma^2(1+\overline{\sigma})}{\underline{\lambda}(1-\sigma^2)^2}\left\|y^k - G\overline{z}^k\right\|^2.$$

Thus, adding the two equations above and using (31), we deduce (33).

Now, we proceed to prove (34). First, we observe that

$$\|\nabla\varphi_k\|^2 = \left\|b^k + G^* a^k\right\|^2 + \left\|y^k - Gx^k\right\|^2.$$

Then, using the triangle inequality for norms and the definition of D, we have

$$\begin{aligned}\left\|b^k + G^* a^k\right\|^2 &\leq \left(\left\|b^k + G^*\overline{w}^k\right\| + \left\|G^*(a^k - \overline{w}^k)\right\|\right)^2 \leq 2\left\|b^k + G^*\overline{w}^k\right\|^2 + 2\left\|G^*\right\|^2\left\|a^k - \overline{w}^k\right\|^2 \\ &\leq 2D\left(\left\|b^k + G^*\overline{w}^k\right\|^2 + \left\|a^k - \overline{w}^k\right\|^2\right)\end{aligned}$$

Similarly, from the triangle inequality it follows

$$\begin{aligned}\left\|y^k - Gx^k\right\|^2 &\leq 2\left(\left\|G\overline{z}^k - y^k\right\|^2 + \left\|G\overline{z}^k - Gx^k\right\|^2\right) \leq 2\left(\left\|G\overline{z}^k - y^k\right\|^2 + \|G\|^2\left\|\overline{z}^k - x^k\right\|^2\right) \\ &\leq 2D\left(\left\|G\overline{z}^k - y^k\right\|^2 + \left\|\overline{z}^k - x^k\right\|^2\right).\end{aligned}$$

Adding the two inequalities above, and using (31) and (32), we conclude (33). □

Remark 3. If Algorithm 2 stops at step 4, then $\nabla\varphi_k = 0$ and by (28), it follows that $\varphi_k(\overline{z}^k, \overline{w}^k) = -\varepsilon_k^x - \varepsilon_k^y$, which combined with (33) and the fact that $\varepsilon_k^x, \varepsilon_k^y \geq 0$ implies $\varepsilon_k^x = \varepsilon_k^y = 0$. Therefore, we have that $Gx^k = y^k$, $b^k = -G^* a^k$, $b^k \in B(x^k)$, $a^k \in A(y^k)$, and consequently $(x^k, a^k) \in S_e$. Hence, when Algorithm 2 stops at step 4, it has found a solution of (1).

We are now ready to establish the global convergence of Algorithm 2.

Theorem 4. *Let the sequences $\{(z^k, w^k)\}$, $\{(x^k, b^k)\}$, $\{(y^k, a^k)\}$ and $\{\eta_k\}$ be generated by Algorithm 2. Assume that*

$$\sum_{k=0}^{\infty} \eta_k \left\|(z^k, w^k) - (z^{k-1}, w^{k-1})\right\|^2 < \infty. \tag{37}$$

Then, $\{(z^k, w^k)\}$ converges weakly to some $(z^*, w^*) \in S_e$. Furthermore, we have $x^k \rightharpoonup z^*$, $y^k \rightharpoonup Gz^*$, $b^k \rightharpoonup -G^* w^*$ and $a^k \rightharpoonup w^*$.

Proof. By Proposition 2, Algorithm 2 is a special instance of Algorithm 1 for finding a point in S_e, with $\{p^k\}$, $\{\overline{p}^k\}$ as in (29) and $\{\varphi_k\}$ given in (27). Thus, the relation (37) implies condition (12), and we have that Theorem 2 holds.

Since we are assuming that Algorithm 2 never stops in step 4 (see Remark 2(vi)), (34) gives that $\varphi_k(\overline{z}^k, \overline{w}^k) > 0$ for all $k \geq 0$, and Theorem 2(iii) implies

$$\frac{\varphi_k(\overline{z}^k, \overline{w}^k)}{\|\nabla\varphi_k\|} \to 0. \tag{38}$$

Moreover, from (34) it follows

$$\frac{\varphi_k(\overline{z}^k, \overline{w}^k)}{\|\nabla\varphi_k\|} \geq \frac{\zeta}{2D(C+1)}\|\nabla\varphi_k\| \qquad \text{and} \qquad \frac{\varphi_k(\overline{z}^k, \overline{w}^k)^2}{\|\nabla\varphi_k\|^2} \geq \frac{\zeta}{2D(C+1)}\varphi_k(\overline{z}^k, \overline{w}^k). \tag{39}$$

Hence, by the first inequality above and (38), we have $\|\nabla\varphi_k\| \to 0$, which yields

$$b^k + G^*a^k \to 0, \qquad y^k - Gx^k \to 0. \tag{40}$$

The second inequality in (39), together with (38), implies that $\varphi_k(\overline{z}^k, \overline{w}^k) \to 0$, and combining with (31) and (32), we obtain

$$\overline{z}^k - x^k \to 0, \qquad G\overline{z}^k - y^k \to 0, \qquad b^k + G^*\overline{w}^k \to 0, \qquad a^k - \overline{w}^k \to 0. \tag{41}$$

Also, using the facts that $\varphi_k(\overline{z}^k, \overline{w}^k) \to 0$, $\varepsilon_k^x, \varepsilon_k^y \geq 0$, and equation (33), we have

$$\lim_{k\to\infty} \varepsilon_k^x = \lim_{k\to\infty} \varepsilon_k^y = 0.$$

Next, we prove that every cluster point of $\{p^k\}$ belongs to S_e. To do this, let $p^* = (z^*, w^*)$ be any cluster point of $\{p^k\}$ and observe that by (30) and (38), p^* is also a cluster point of $\{\overline{p}^k\}$. Let $\{\overline{p}^{k_j}\}$ be a subsequence such that $\overline{p}^{k_j} \rightharpoonup p^*$, then $\overline{z}^{k_j} \rightharpoonup z^*$ and $\overline{w}^{k_j} \rightharpoonup w^*$. Therefore, by the first and fourth limits in (41) one has $x^{k_j} \rightharpoonup z^*$ and $a^{k_j} \rightharpoonup w^*$.

Now we define the operators $F, \Phi, T : \mathcal{E} \rightrightarrows \mathcal{E}$ by

$$F(z,w) = (G^*w, -Gz), \qquad \Phi(z,w) = B(z) \times A^{-1}(w), \qquad T = F + \Phi$$

and observe that F, Φ, T are maximal monotone and $S_e = T^{-1}(0,0)$ (see [18, Lemma 2]). Also, note that the inclusion in (22) and Proposition 1(i) imply $y^k \in (A^{-1})^{\varepsilon_k^y}(a^k)$, which combined with the inclusion in (21) and Proposition 1(iii) gives $(b^k, y^k) \in \Phi^{\varepsilon_k^x + \varepsilon_k^y}(x^k, a^k)$. Using the definitions of F, Φ, T and Proposition 1(ii) we now obtain $(G^*a^k + b^k, y^k - Gx^k) \in T^{\varepsilon_k^x + \varepsilon_k^y}(x^k, a^k)$. Further, since $(G^*a^{k_j} + b^{k_j}, y^{k_j} - Gx^{k_j}) \to 0$, $(x^{k_j}, a^{k_j}) \rightharpoonup (z^*, w^*)$ and $\varepsilon_{k_j}^x + \varepsilon_{k_j}^y \to 0$, Proposition 1(iv) implies $(0,0) \in T(z^*, w^*)$. We thus deduce that $(z^*, w^*) \in S_e$, and by Theorem 2(ii) we conclude that $\{(z^k, w^k)\}$ converges weakly to a point (z^*, w^*) in S_e, from which also follows that $(\overline{z}^k, \overline{w}^k) \rightharpoonup (z^*, w^*)$. The remaining statements of the theorem are a consequence of this latter assertion and (41). □

The next theorem gives the sufficient conditions (42) and (43) on $\{\eta_k\}$, η, and $\overline{\beta}$ that ensure the summability condition (37) and, therefore, guarantee the convergence of Algorithm 2. We observe that these conditions are the same as (13) and (14) of Theorem 3.

Theorem 5. *Consider the sequences $\{(z^k, w^k)\}$, $\{(x^k, b^k)\}$, $\{(y^k, a^k)\}$ generated by Algorithm 2 and assume that $\eta \in [0,1)$, $\overline{\beta} \in (0,2)$ and $\{\eta_k\}$ satisfy, for some $\overline{\eta} > 0$, that*

$$0 \leq \eta_k \leq \eta_{k+1} \leq \eta < \overline{\eta} < 1 \tag{42}$$

and

$$\overline{\beta} = \overline{\beta}(\overline{\eta}) := \frac{2(\overline{\eta}-1)^2}{2(\overline{\eta}-1)^2 + 3\overline{\eta} - 1}. \tag{43}$$

Then, the relation in (37) holds. In particular, there is (z^*, w^*) in S_e such that $z^k, x^k \rightharpoonup z^*$, $y^k \rightharpoonup Gz^*$, $w^k, a^k \rightharpoonup w^*$ and $b^k \rightharpoonup -G^*w^*$.

Proof. In view of Proposition 2 and Theorem 3, assumptions (42) and (43) imply (37). Therefore, Theorem 4 holds and the statements of the theorem are satisfied. □

3.2 Complexity analysis

Our goal in this section is to study the iteration-complexity of Algorithm 2 for solving (1). For this purpose, we consider the following notion of approximate solution for (1): for given tolerances $\overline{\rho}, \overline{\varepsilon} > 0$, find $(x, b, \epsilon^x) \in \mathcal{X} \times \mathcal{X} \times \mathbb{R}^+$ and $(y, a, \epsilon^y) \in \mathcal{Y} \times \mathcal{Y} \times \mathbb{R}^+$ satisfying

$$b \in B^{\epsilon^x}(x), \qquad a \in A^{\epsilon^y}(y), \qquad \max\{\|G^*a + b\|, \|Gx - y\|\} \leq \overline{\rho}, \qquad \max\{\epsilon^x, \epsilon^y\} \leq \overline{\epsilon}. \tag{44}$$

We observe that if $\overline{\rho} = \overline{\varepsilon} = 0$, criterion (44) is reduced to $\varepsilon^x = \varepsilon^y = 0$, $G^*a + b = 0$, $Gx = y$, $b \in B(x)$ and $a \in A(y)$, in which case $(x, a) \in S_e$ and x is a solution of (1).

In what follows, we prove iteration-complexity results for Algorithm 2 to obtain approximate solutions of (1) according to (44). We first present *pointwise* complexity bounds, which estimate the quality of the best iterates among the first k generated by the algorithm.

Theorem 6. *Let the sequences $\{(x^k, b^k, \varepsilon_k^x)\}$ and $\{(y^k, a^k, \varepsilon_k^y)\}\}$ be generated by Algorithm 2. Assume the hypotheses of Theorem 5 and let $d_0 := d\big((z^0, w^0), S_e\big)$. Then, for every $k \in \mathbb{N}$, there exists an index $j_0 \in \{1, \ldots, k\}$ such that*

$$b^{j_0} \in B^{\varepsilon_{j_0}^x}(x^{j_0}), \qquad a^{j_0} \in A^{\varepsilon_{j_0}^y}(y^{j_0})$$

and

$$\begin{aligned} \left\|G^* a^{j_0} + b^{j_0}\right\| &\leq \frac{2D(C+1)d_0}{\zeta(1-\eta)\sqrt{k}} \sqrt{\frac{\eta(1+\eta)(2-\eta)}{\underline{\beta}(2-\overline{\beta})q(\eta)}}, \\ \left\|Gx^{j_0} - y^{j_0}\right\| &\leq \frac{2D(C+1)d_0}{\zeta(1-\eta)\sqrt{k}} \sqrt{\frac{\eta(1+\eta)(2-\eta)}{\underline{\beta}(2-\overline{\beta})q(\eta)}}, \\ \varepsilon_{j_0}^x + \varepsilon_{j_0}^y &\leq \frac{2D(C+1)\sigma^2(1+\overline{\sigma})d_0^2}{\zeta^2\underline{\lambda}(1-\sigma^2)^2\underline{\beta}(2-\overline{\beta})k} \cdot \frac{\eta(1+\eta)(2-\eta)}{(1-\eta)^2 q(\eta)}, \end{aligned} \tag{45}$$

where ζ, C and D are as in Proposition 3 and $q(\eta)$ is defined in Theorem 3.

Proof. The assertions that $b^{j_0} \in B^{\varepsilon_{j_0}^x}(x^{j_0})$ and $a^{j_0} \in A^{\varepsilon_{j_0}^y}(y^{j_0})$ are direct consequences of steps 2 and 3 of Algorithm 2.

Proposition 2 implies that the sequences $\{p^k\}$ and $\{\overline{p}^k\}$ in (29) satisfy Lemma 3 with $p^* := P_{S_e}(z^0, w^0)$. Therefore, by (10), the sequences $\{h_k\}$ given in (9), and $\{s_k\}$, $\{\delta_k\}$ defined in (11) satisfy Lemma 2. Consequently, (5), the definitions of these sequences, equation (8), and the assumption that $\varphi_k(\overline{z}^k, \overline{w}^k) > 0$ for all $k \geq 0$, give

$$\left\|p^k - p^*\right\|^2 + \sum_{j=0}^{k} \beta_j(2-\beta_j)\left(\frac{\varphi_j(\overline{z}^j, \overline{w}^j)}{\|\nabla\varphi_j\|}\right)^2 \leq \frac{1}{1-\eta}\sum_{j=0}^{k-1}\eta_j(1+\eta_j)\left\|p^j - p^{j-1}\right\|^2. \tag{46}$$

We observe that assumptions (42) and (43), together with Proposition 2, imply that Theorem 3 holds. Therefore, combining inequality (15) with (46) and using that $\left\|p^0 - p^*\right\| = d_0$, we obtain

$$\left\|p^k - p^*\right\|^2 + \sum_{j=0}^{k} \beta_j(2-\beta_j)\left(\frac{\varphi_j(\overline{z}^j, \overline{w}^j)}{\|\nabla\varphi_j\|}\right)^2 \leq \frac{\eta(1+\eta)(2-\eta)}{(1-\eta)^2 q(\eta)} d_0^2. \tag{47}$$

From (47) and the choice $\beta_j \in [\underline{\beta}, \overline{\beta}]$ for all $j \geq 0$, it follows that for every $k \geq 1$ there exists an index $j_0 \in \{1, \ldots, k\}$ such that

$$\underline{\beta}(2-\overline{\beta})k\left(\frac{\varphi_{j_0}(\overline{z}^{j_0}, \overline{w}^{j_0})}{\|\nabla\varphi_{j_0}\|}\right)^2 \leq \frac{\eta(1+\eta)(2-\eta)}{(1-\eta)^2 q(\eta)} d_0^2. \tag{48}$$

Now, equation (48), together with the first inequality in (39), gives

$$\frac{\zeta^2}{4D^2(C+1)^2}\left(\left\|G^*a^{j_0} + b^{j_0}\right\|^2 + \left\|Gx^{j_0} - y^{j_0}\right\|^2\right) \leq \frac{d_0^2}{\underline{\beta}(2-\overline{\beta})k} \cdot \frac{\eta(1+\eta)(2-\eta)}{(1-\eta)^2 q(\eta)},$$

from which the first two inequalities in (45) follow.

Finally, we combine (48) with the second inequality in (39) and (33) to obtain

$$\frac{\zeta^2\underline{\lambda}(1-\sigma^2)^2}{2D(C+1)\sigma^2(1+\overline{\sigma})}(\varepsilon_{j_0}^x + \varepsilon_{j_0}^y) \leq \frac{d_0^2}{\underline{\beta}(2-\overline{\beta})k} \cdot \frac{\eta(1+\eta)(2-\eta)}{(1-\eta)^2 q(\eta)},$$

which implies the last inequality in (45). □

We observe that Theorem 6 provides $\mathcal{O}(1/\sqrt{k})$ *pointwise* convergence rate and guarantees that, for given $\overline{\rho}, \overline{\varepsilon} > 0$, Algorithm 2 finds triplets (x, b, ε^x) and (y, a, ε^y) satisfying (44) in at most

$$\mathcal{O}\left(\max\left\{\frac{d_0^2}{\zeta^2\,\overline{\rho}^2}, \frac{d_0^2}{\zeta^2\,\overline{\varepsilon}}\right\}\right)$$

iterations.

Next, we derive alternative estimates for Algorithm 2, which we refer to as the *ergodic* iteration-complexity bounds. To this end, we will consider the following *ergodic sequences* associated with the sequences generated by Algorithm 2:

$$\begin{aligned}
&\Gamma_k = \sum_{l=0}^{k} \beta_l\gamma_l, \quad \hat{x}^k = \frac{1}{\Gamma_k}\sum_{l=0}^{k}\beta_l\gamma_l x^l, \quad \hat{b}^k = \frac{1}{\Gamma_k}\sum_{l=0}^{k}\beta_l\gamma_l b^l, \quad \hat{y}^k = \frac{1}{\Gamma_k}\sum_{l=0}^{k}\beta_l\gamma_l y^l, \quad \hat{a}^k = \frac{1}{\Gamma_k}\sum_{l=0}^{k}\beta_l\gamma_l a^l \\
&\hat{\varepsilon}_k^x = \frac{1}{\Gamma_k}\sum_{l=0}^{k}\beta_l\gamma_l\left(\varepsilon_l^x + \langle x^l - \hat{x}^k, b^l - \hat{b}^k\rangle\right), \quad \hat{\varepsilon}_k^y = \frac{1}{\Gamma_k}\sum_{l=0}^{k}\beta_l\gamma_l\left(\varepsilon_l^y + \langle y^l - \hat{y}^k, a^l - \hat{a}^k\rangle\right).
\end{aligned} \tag{49}$$

The theorem below estimates when the ergodic iterates above will meet the criterion (44).

Theorem 7. *Assume the hypotheses of Theorem 6. Additionally, suppose that $\eta_k \equiv \eta$ and consider the ergodic iterates defined in (49). Then, for any $k \in \mathbb{N}$, we have*

$$\hat{b}^k \in B^{\hat{\varepsilon}_k^x}(\hat{x}^k), \qquad \hat{a}^k \in A^{\hat{\varepsilon}_k^y}(\hat{y}^k)$$

and

$$\begin{aligned}
\left\|G^*\hat{a}^k + \hat{b}^k\right\| &\le \frac{4(1+\eta)D(C+1)d_0}{(k+1)(1-\eta)\underline{\beta}\zeta}\sqrt{\frac{\eta(1+\eta)(2-\eta)}{q(\eta)}}, \\
\left\|G\hat{x}^k - \hat{y}^k\right\| &\le \frac{4(1+\eta)D(C+1)d_0}{(k+1)(1-\eta)\underline{\beta}\zeta}\sqrt{\frac{\eta(1+\eta)(2-\eta)}{q(\eta)}}, \\
\hat{\varepsilon}_k^x + \hat{\varepsilon}_k^y &\le \frac{4D(C+1)d_0^2}{(k+1)\underline{\beta}\zeta}\left(\xi\frac{1+\eta}{1-\eta} + \frac{1}{4}\right)\frac{\eta(1+\eta)(2-\eta)}{(1-\eta)q(\eta)},
\end{aligned} \tag{50}$$

where ζ, C and D are the constants defined in Proposition 3, and $\xi := 3 + \dfrac{2CD(C+1)}{\zeta^2\underline{\beta}(2-\overline{\beta})}$.

Proof. The inclusions in the statement of the theorem follow from the definitions of the ergodic sequences in (49) and Theorem 1.

By steps 5 and 1 of Algorithm 2 and the assumption that $\eta_l \equiv \eta$, we have, for each $l \ge 0$,

$$\beta_l\gamma_l(G^*a^l + b^l) = \overline{z}^l - z^{l+1} = z^l + \eta(z^l - z^{l-1}) - z^{l+1} = z^l - z^{l+1} + \eta(z^l - z^{l-1}).$$

Adding equality above from $l = 0$ to k, we obtain

$$\sum_{l=0}^{k}\beta_l\gamma_l(G^*a^l + b^l) = (z^0 - z^{k+1}) + \eta(z^k - z^0).$$

Therefore, the above equation, the linearity of G^*, and the definitions of $\hat{a}^k$ and $\hat{b}^k$ in (49) imply

$$\Gamma_k(G^*\hat{a}^k + \hat{b}^k) = (z^0 - z^{k+1}) + \eta(z^k - z^0). \tag{51}$$

In the same way, we can prove

$$\Gamma_k(\hat{y}^k - G\hat{x}^k) = (w^0 - w^{k+1}) + \eta(w^k - w^0). \tag{52}$$

Next, we use (47) and the definition of p^k in (29) to obtain

$$\left\|z^j - z^*\right\|^2 + \left\|w^j - w^*\right\|^2 \le \frac{\eta(1+\eta)(2-\eta)}{(1-\eta)^2 q(\eta)} d_0^2, \qquad \forall j \ge 0. \tag{53}$$

Equation (53), together with the triangle inequality and simple calculations, yields

$$\left\|z^j - z^l\right\| \le \left\|z^j - z^*\right\| + \left\|z^l - z^*\right\| \le 2\frac{d_0}{1-\eta}\sqrt{\frac{\eta(1+\eta)(2-\eta)}{q(\eta)}}, \qquad \forall j, l \ge 0.$$

Hence, the inequality above with $j = 0$ and $l = k, k+1$, and equality (51) give

$$\Gamma_k \left\|G^*\hat{a}^k + \hat{b}^k\right\| \le 2\frac{d_0}{1-\eta}\sqrt{\frac{\eta(1+\eta)(2-\eta)}{q(\eta)}} + 2\eta\frac{d_0}{1-\eta}\sqrt{\frac{\eta(1+\eta)(2-\eta)}{q(\eta)}} = \frac{2(1+\eta)d_0}{1-\eta}\sqrt{\frac{\eta(1+\eta)(2-\eta)}{q(\eta)}}. \tag{54}$$

Similarly, we use (53), the triangle inequality, and (52) to deduce

$$\Gamma_k \left\|\hat{y}^k - G\hat{x}^k\right\| \le \frac{2(1+\eta)d_0}{1-\eta}\sqrt{\frac{\eta(1+\eta)(2-\eta)}{q(\eta)}}. \tag{55}$$

From step 4 of Algorithm 2 and Lemma 5, it is easy to see that

$$\gamma_l = \frac{\varphi_l(\overline{z}^l, \overline{w}^l)}{\|\nabla\varphi_l\|^2}, \qquad \forall\, l \ge 0. \tag{56}$$

Therefore, the definition of Γ_k, (56), the first inequality in (39), and the fact $\beta_l \ge \underline{\beta}$ imply

$$\Gamma_k = \sum_{l=0}^{k} \beta_l \gamma_l = \sum_{l=0}^{k} \beta_l \frac{\varphi_l(\overline{z}^l, \overline{w}^l)}{\|\nabla\varphi_l\|^2} \ge \sum_{l=0}^{k} \underline{\beta}\frac{\zeta}{2D(C+1)} = (k+1)\frac{\underline{\beta}\zeta}{2D(C+1)}. \tag{57}$$

Now, combining (57) with (54) and (55) we deduce the first two inequalities in (50).

To prove the last inequality in (50), we first note that the definition of φ_l in (27) and simple manipulations yield

$$\begin{aligned}\varphi_l(\hat{x}^k, \hat{a}^k) &= \left\langle \hat{x}^k - x^l, b^l + G^*\hat{a}^k \right\rangle + \left\langle G\hat{x}^k - y^l, a^l - \hat{a}^k \right\rangle - \varepsilon_l^x - \varepsilon_l^y \\ &= \left\langle \hat{x}^k - x^l, b^l - \hat{b}^k \right\rangle + \left\langle \hat{x}^k - x^l, \hat{b}^k + G^*\hat{a}^k \right\rangle + \left\langle G\hat{x}^k - \hat{y}^k, a^l - \hat{a}^k \right\rangle + \left\langle \hat{y}^k - y^l, a^l - \hat{a}^k \right\rangle - \varepsilon_l^x - \varepsilon_l^y.\end{aligned}$$

Multiplying the equation above by $\beta_l\gamma_l/\Gamma_k$, adding from $l = 0$ to k, and using the definitions of $\hat{x}^k$ and $\hat{a}^k$ in (49), we obtain

$$\frac{1}{\Gamma_k}\sum_{l=0}^{k} \beta_l\gamma_l\varphi_l(\hat{x}^k, \hat{a}^k) = \frac{1}{\Gamma_k}\sum_{l=0}^{k} \beta_l\gamma_l \left(\left\langle \hat{x}^k - x^l, b^l - \hat{b}^k \right\rangle + \left\langle \hat{y}^k - y^l, a^l - \hat{a}^k \right\rangle - \varepsilon_l^x - \varepsilon_l^y\right).$$

The equation above, together with the definitions of $\hat{\varepsilon}_k^x$ and $\hat{\varepsilon}_k^y$ in (49), now implies

$$\hat{\varepsilon}_k^x + \hat{\varepsilon}_k^y = -\frac{1}{\Gamma_k}\sum_{l=0}^{k} \beta_l\gamma_l\varphi_l(\hat{x}^k, \hat{a}^k). \tag{58}$$

Next, we use the update rule (24) and perform straightforward calculations to obtain

$$\begin{aligned}\left\|(\hat{x}^k, \hat{a}^k) - (z^{l+1}, w^{l+1})\right\|^2 &= \left\|(\hat{x}^k, \hat{a}^k) - (\overline{z}^l, \overline{w}^l)\right\|^2 + \beta_l^2\gamma_l^2 \|\nabla\varphi_l\|^2 + 2\beta_l\gamma_l \left\langle (\hat{x}^k, \hat{a}^k) - (\overline{z}^l, \overline{w}^l), \nabla\varphi_l \right\rangle \\ &= \left\|(\hat{x}^k, \hat{a}^k) - (\overline{z}^l, \overline{w}^l)\right\|^2 + \beta_l^2\gamma_l^2 \|\nabla\varphi_l\|^2 + 2\beta_l\gamma_l \left\langle (\hat{x}^k, \hat{a}^k) - (x^l, a^l), \nabla\varphi_l \right\rangle \\ &\quad + 2\beta_l\gamma_l \left\langle (x^l, a^l) - (\overline{z}^l, \overline{w}^l), \nabla\varphi_l \right\rangle.\end{aligned}$$

Adding and subtracting $2\beta_l\gamma_l(\varepsilon_l^x+\varepsilon_l^y)$ on the right-hand side of the last equality above and using the identity (28), we get

$$\begin{aligned}\left\|(\hat{x}^k,\hat{a}^k)-(z^{l+1},w^{l+1})\right\|^2 &= \left\|(\hat{x}^k,\hat{a}^k)-(\overline{z}^l,\overline{w}^l)\right\|^2+\beta_l^2\gamma_l^2\left\|\nabla\varphi_l\right\|^2+2\beta_l\gamma_l\varphi_l(\hat{x}^k,\hat{a}^k)-2\beta_l\gamma_l\varphi_l(\overline{z}^l,\overline{w}^l)\\ &= \left\|(\hat{x}^k,\hat{a}^k)-(\overline{z}^l,\overline{w}^l)\right\|^2-\beta_l(2-\beta_l)\gamma_l^2\left\|\nabla\varphi_l\right\|^2+2\beta_l\gamma_l\varphi_l(\hat{x}^k,\hat{a}^k),\end{aligned}$$

where the second equality follows from the relation in (56). Rearranging the equation above and using the fact that $\beta_l(2-\beta_l)\gamma_l^2\left\|\nabla\varphi_l\right\|^2\geq 0$, we deduce

$$-2\beta_l\gamma_l\varphi_l(\hat{x}^k,\hat{a}^k)\leq\left\|(\hat{x}^k,\hat{a}^k)-(\overline{z}^l,\overline{w}^l)\right\|^2-\left\|(\hat{x}^k,\hat{a}^k)-(z^{l+1},w^{l+1})\right\|^2.$$

Also, we use step 1 of Algorithm 2 and consider the relation (6) with $p=(z^{l-1},w^{l-1})-(\hat{x}^k,\hat{a}^k)$, $q=(z^l,w^l)-(\hat{x}^k,\hat{a}^k)$ and $t=-\eta$ to obtain

$$\left\|(\overline{z}^l,\overline{w}^l)-(\hat{x}^k,\hat{a}^k)\right\|^2=(1+\eta)\left\|(z^l,w^l)-(\hat{x}^k,\hat{a}^k)\right\|^2-\eta\left\|(z^{l-1},w^{l-1})-(\hat{x}^k,\hat{a}^k)\right\|^2+\delta_l,$$

where δ_l is as in (11) with p^l given in (29). Thus, from the two equations above, it follows that

$$\begin{aligned}-2\beta_l\gamma_l\varphi_l(\hat{x}^k,\hat{a}^k)\leq&\left\|(z^l,w^l)-(\hat{x}^k,\hat{a}^k)\right\|^2-\left\|(z^{l+1},w^{l+1})-(\hat{x}^k,\hat{a}^k)\right\|^2\\ &+\eta\left(\left\|(z^l,w^l)-(\hat{x}^k,\hat{a}^k)\right\|^2-\left\|(z^{l-1},w^{l-1})-(\hat{x}^k,\hat{a}^k)\right\|^2\right)+\delta_l.\end{aligned}$$

Further, by adding the equation above from $l=0$ to k and combining it with (58), we obtain

$$2\Gamma_k(\hat{\varepsilon}_k^x+\hat{\varepsilon}_k^y)\leq\left\|(z^0,w^0)-(\hat{x}^k,\hat{a}^k)\right\|^2+\eta\left\|(z^k,w^k)-(\hat{x}^k,\hat{a}^k)\right\|^2+\sum_{l=0}^k\delta_l. \tag{59}$$

We now proceed to find upper bounds for the first and second terms on the right-hand side of the equation (59). To this end, we define

$$(\hat{\overline{z}}^k,\hat{\overline{w}}^k):=\frac{1}{\Gamma_k}\sum_{l=0}^k\beta_l\gamma_l(\overline{z}^l,\overline{w}^l)$$

and use the definition of $(\hat{x}^k,\hat{a}^k)$ and the convexity of $\|\cdot\|^2$ to have, for all $j\geq 0$,

$$\begin{aligned}\left\|(z^j,w^j)-(\hat{x}^k,\hat{a}^k)\right\|^2&\leq 2\left\|(z^j,w^j)-(\hat{\overline{z}}^k,\hat{\overline{w}}^k)\right\|^2+2\left\|(\hat{\overline{z}}^k,\hat{\overline{w}}^k)-(\hat{x}^k,\hat{a}^k)\right\|^2\\ &\leq 2\frac{1}{\Gamma_k}\sum_{l=0}^k\beta_l\gamma_l\left\|(z^j,w^j)-(\overline{z}^l,\overline{w}^l)\right\|^2+2\frac{1}{\Gamma_k}\sum_{l=0}^k\beta_l\gamma_l\left\|(\overline{z}^l,\overline{w}^l)-(x^l,a^l)\right\|^2.\end{aligned} \tag{60}$$

Moreover, it is easy to see that

$$\left\|(z^j,w^j)-(\overline{z}^l,\overline{w}^l)\right\|^2\leq\ 2\left\|(z^j,w^j)-(z^*,w^*)\right\|^2+2\left\|(\overline{z}^l,\overline{w}^l)-(z^*,w^*)\right\|^2,\qquad\forall\, j,l\geq 0$$

and from (47) it follows

$$\left\|(z^j,w^j)-(z^*,w^*)\right\|^2\leq\frac{\eta(1+\eta)(2-\eta)}{(1-\eta)^2q(\eta)}d_0^2,\qquad\forall j\geq 0.$$

On the other hand, considering the identity (6) with $p=(z^{l-1},w^{l-1})-(z^*,w^*)$, $q=(z^l,w^l)-(z^*,w^*)$ and $t=-\eta$, (20), (47) and (15), we obtain

$$\begin{aligned}\left\|(\overline{z}^l,\overline{w}^l)-(z^*,w^*)\right\|^2&\leq(1+\eta)\left\|(z^l,w^l)-(z^*,w^*)\right\|^2+\eta(1+\eta)\left\|(z^l,w^l)-(z^{l-1},w^{l-1})\right\|^2\\ &\leq(1+\eta)\frac{\eta(1+\eta)(2-\eta)}{(1-\eta)^2q(\eta)}d_0^2+\eta(1+\eta)\frac{(2-\eta)}{(1-\eta)q(\eta)}d_0^2\\ &=2\frac{\eta(1+\eta)(2-\eta)}{(1-\eta)^2q(\eta)}d_0^2.\end{aligned}$$

Combining the three equations above gives

$$\left\|(z^j,w^j)-(\overline{z}^l,\overline{w}^l)\right\|^2 \leq 6\frac{\eta(1+\eta)(2-\eta)}{(1-\eta)^2q(\eta)}d_0^2, \qquad \forall j\geq 0.$$

We substitute the inequality above into (60) to have

$$\left\|(z^j,w^j)-(\hat{x}^k,\hat{a}^k)\right\|^2 \leq 12\frac{\eta(1+\eta)(2-\eta)}{(1-\eta)^2q(\eta)}d_0^2+2\frac{1}{\Gamma_k}\sum_{l=0}^{k}\beta_l\gamma_l\left\|(\overline{z}^l,\overline{w}^l)-(x^l,a^l)\right\|^2.$$

Since $C>1$, (31), together with (32), implies

$$\varphi_l(\overline{z}^l,\overline{w}^l)\geq \frac{\zeta}{2C}\left(\left\|\overline{z}^l-x^l\right\|^2+\left\|a^k-\overline{w}^k\right\|^2\right)=\frac{\zeta}{2C}\left\|(\overline{z}^l,\overline{w}^l)-(x^l,a^l)\right\|^2.$$

Now, by the two equations above and (56), we obtain

$$\begin{aligned}\left\|(z^j,w^j)-(\hat{x}^k,\hat{a}^k)\right\|^2 &\leq 12\frac{\eta(1+\eta)(2-\eta)}{(1-\eta)^2q(\eta)}d_0^2+4\frac{C}{\zeta\Gamma_k}\sum_{l=0}^{k}\beta_l\gamma_l\varphi_l(\overline{z}^l,\overline{w}^l)\\ &=12\frac{\eta(1+\eta)(2-\eta)}{(1-\eta)^2q(\eta)}d_0^2+4\frac{C}{\zeta\Gamma_k}\sum_{l=0}^{k}\beta_l\frac{\varphi_l(\overline{z}^l,\overline{w}^l)^2}{\left\|\nabla\varphi_l\right\|^2}.\end{aligned}$$

From $\beta_l\leq\overline{\beta}$, (47), (57) and the definition of ξ, it follows

$$\begin{aligned}\left\|(z^j,w^j)-(\hat{x}^k,\hat{a}^k)\right\|^2 &\leq 12\frac{\eta(1+\eta)(2-\eta)}{(1-\eta)^2q(\eta)}d_0^2+8\frac{CD(C+1)}{\zeta^2\underline{\beta}}\cdot\frac{\eta(1+\eta)(2-\eta)}{(2-\overline{\beta})(1-\eta)^2q(\eta)}d_0^2\\ &=\frac{4\xi\eta(1+\eta)(2-\eta)}{(1-\eta)^2q(\eta)}d_0^2.\end{aligned}$$

Equation above with $j=0,k$, (59), the definition of $\{\delta_l\}$ in (11), and relations (15) and (57) imply the last inequality in (50). □

We emphasize that the ergodic complexity bounds for Algorithm 2 are asymptotically better than the pointwise complexity ones. Indeed, the bounds derived in Theorem 6 are $\mathcal{O}(1/\sqrt{k})$, whereas in Theorem 7 the bounds are $\mathcal{O}(1/k)$. However, we were only able to prove the results of this latter theorem in the case where the inertial parameters η_k were constant.

4 Lipschitz continuous operators

On each iteration, Algorithm 2 has to solve inexactly two proximal subproblems within a relative error criterion. In general, the procedure used to perform this computation will depend on the implementation of the method and the properties of the operators. For the case where A and B are Lipschitz continuous, we discuss in this section how approximate solutions of the proximal subproblems satisfying (21) and (22) can be computed.

If only one of the operators is Lipschitz, we can consider the procedure presented below regarding only this operator. Thus, from now on we assume that

(L1) A is L_A-Lipschitz, B is L_B-Lipschitz, and we define $\mathcal{L}:=\max\{L_A,L_B\}$.

We first show that if the Lipschitz constants of the operators are known, performing two appropriate evaluations (*forward steps*) of each operator, we can obtain triplets that satisfy the error conditions (21) and (22) in steps 2 and 3 of Algorithm 2. We observe that forward steps were proposed in [23] for handling Lipschitz continuous operators in the non-inertial projective splitting method in the parallel case ($\alpha_k\equiv 0$), and in [27], it was proved that they can be viewed as an inexact proximal update. Here, for the case $n=2$, we extend this procedure to any value of α_k.

Lemma 6. *Assume that $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is an L-Lipschitz continuous maximal monotone operator, $0 < c < 1/L$ and $z, v \in \mathcal{H}$. Define $\chi = z - c(Tz - v)$ and $\varpi = T\chi$, then for all $\sigma \in (cL, 1)$ it holds*

$$\|c(\varpi - v) + \chi - z\|^2 \leq \sigma^2 \left(\|\chi - z\|^2 + \|c(\varpi - v)\|^2 \right). \tag{61}$$

Proof. Using the definitions of χ and ϖ we have

$$\|c(\varpi - v) + \chi - z\|^2 = \|c(T\chi - Tz)\|^2 \leq c^2 L^2 \|\chi - z\|^2,$$

where the inequality above follows from the Lipschitz continuity of T. From the assumption that $\sigma > cL$, we deduce

$$\|c(\varpi - v) + \chi - z\|^2 \leq \sigma^2 \|\chi - z\|^2,$$

from which (61) directly follows. □

Proposition 4. *For all $k \geq 0$, take $0 < \lambda_k < 1/L_B$, $0 < \mu_k < 1/L_A$, $\alpha_k \in \mathbb{R}$ and define*

$$\begin{aligned} x^k &= \overline{z}^k - \lambda_k(B\overline{z}^k + G^*\overline{w}^k), & b^k &= Bx^k, \\ y^k &= G((1-\alpha_k)\overline{z}^k + \alpha_k x^k) - \mu_k(A\overline{z}^k - \overline{w}^k), & a^k &= Ay^k. \end{aligned} \tag{62}$$

Then, $(x^k, b^k, 0)$ and $(y^k, a^k, 0)$ satisfy, respectively, the relations in (21) and (22) for all $\sigma \in (max\{\lambda_k L_B, \mu_k L_A\}, 1)$.

Proof. It is clear that the inclusions in (21) and (22) hold. By applying Lemma 6 with $\mathcal{H} = \mathcal{X}$, $T = B$, $L = L_B$, $c = \lambda_k$, $z = \overline{z}^k$ and $v = -G^*\overline{w}^k$, we obtain equation (61) with $\chi = x^k$ and $\varpi = b^k$ for all $\sigma \in (\lambda_k L_B, 1)$, and thus the inequality in (21) follows.

Similarly, using Lemma 6 with $\mathcal{H} = \mathcal{Y}$, $T = A$, $L = L_A$, $c = \mu_k$, $z = G\overline{z}^k$ and $v = \alpha_k/\mu_k G(x^k - \overline{z}^k) + \overline{w}^k$, we have (61) with $\chi = y^k$ and $\varpi = a^k$ for all $\sigma \in (\mu_k L_A, 1)$, which implies the inequality in (22). Taking σ large enough, we conclude the proposition. □

As a consequence of the previous proposition, the procedure in (62), which only requires two evaluations of each operator per iteration, can be used in steps 2 and 3 of Algorithm 2 to obtain the desired triplets. However, if the Lipschitz constants of the operators are unknown, Proposition 4 cannot be applied. In that case, we can implement the following backtracking linesearch to be used in place of (62).

Backtracking procedure 1. Input: T, $z, u, v \in \mathcal{H}$, $c_1 > 0$, $\Delta > 0$

Set $\phi = Tz$ and for $j = 1, 2, \ldots$ do

1. $\tilde{\chi}^j = z + u - c_j(\phi - v)$

2. $\tilde{\varpi}^j = T\tilde{\chi}^j$

3. if $\Delta \left\|z - \tilde{\chi}^j\right\|^2 \leq \left\langle z - \tilde{\chi}^j, \tilde{\varpi}^j - v \right\rangle - \frac{1}{c_j} \left\langle z - \tilde{\chi}^j, u \right\rangle$ then

4. return $J \leftarrow j$, $c \leftarrow c_j$, $\chi \leftarrow \tilde{\chi}^j$, $\varpi \leftarrow \tilde{\varpi}^j$

5. else $c_{j+1} = c_j/2$

The backtracking procedure (BP) 1 is closely related to the backtracking linesearch proposed in [23]. However, different from [23], the BP 1 allows us to deal with Lipschitz operators not only in the parallel case. The following lemma shows that the loop in the BP 1 has finite termination and that the number of iterations is bounded. Its proof is similar to [23, Theorem 2], and for the sake of brevity we only present the necessary modifications.

Lemma 7. *Assume that $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is an L-Lipschitz continuous maximal monotone operator. Then, for the loop in the BP 1 we have*

$$J \leq \max\{2 + \log_2((\Delta + L)c_1), 1\}$$

and

$$c \geq \min\left\{\frac{1}{2(L+\Delta)}, c_1\right\}. \tag{63}$$

Proof. By simple manipulations we have

$$\begin{aligned}\left\langle z - \tilde{\chi}^j, \tilde{\varpi}^j - v\right\rangle - \frac{1}{c_j}\left\langle z - \tilde{\chi}^j, u\right\rangle &= \left\langle z - \tilde{\chi}^j, \tilde{\varpi}^j - \phi\right\rangle + \left\langle z - \tilde{\chi}^j, \phi - v\right\rangle - \frac{1}{c_j}\left\langle z - \tilde{\chi}^j, u\right\rangle \\ &= \left\langle z - \tilde{\chi}^j, \tilde{\varpi}^j - \phi\right\rangle + \left\langle z - \tilde{\chi}^j, \phi - v - \frac{1}{c_j}u\right\rangle \\ &= \left\langle z - \tilde{\chi}^j, T\tilde{\chi}^j - Tz\right\rangle + \left\langle z - \tilde{\chi}^j, \frac{1}{c_j}(z - \tilde{\chi}^j)\right\rangle \\ &\geq \left(\frac{1}{c_j} - L\right)\left\|z - \tilde{\chi}^j\right\|^2,\end{aligned}$$

where the third equality above follows from the definitions of ϕ, $\tilde{\chi}^j$ and $\tilde{\varpi}^j$ in the BP 1, and the inequality is due to the Cauchy-Schwartz inequality and the assumption that T is L-Lipschitz. The rest of the proof runs as [23, Theorem 2]. □

In order to compute admissible triplets satisfying steps 2 and 3 of Algorithm 2, we can use the BP 1 as follows:

2′. Choose $\lambda_{(1,k)} > 0$ and set (λ_k, x^k, b^k) = **Backtracking procedure** 1 $(B, \overline{z}^k, 0, -G^*\overline{w}^k, \lambda_{(1,k)}, \Delta)$.

3′. Choose $\mu_{(1,k)} > 0$, $\alpha_k \in \mathbb{R}$ and set (μ_k, y^k, a^k) = **Backtracking procedure** 1 $(A, G\overline{z}^k, \alpha_k G(x^k - \overline{z}^k), \overline{w}^k, \mu_{(1,k)}, \Delta)$.

Additionally, we consider the following assumption:

(L2) there exist $\underline{\mu}$ and $\overline{\lambda}$ such that $\overline{\lambda} \geq \underline{\mu} > 0$ and $\lambda_{(1,k)}, \mu_{(1,k)} \in [\underline{\mu}, \overline{\lambda}]$ for all $k \geq 0$.

We observe that since for all $k \geq 0$, λ_k is given by the BP 1 with initial trial stepsize $\lambda_{(1,k)}$, equation (63) yields

$$\lambda_k \geq \min\left\{\frac{1}{2(L_B+\Delta)}, \lambda_{(1,k)}\right\} \geq \min\left\{\frac{1}{2(\mathcal{L}+\Delta)}, \underline{\mu}\right\} =: \underline{\lambda}, \tag{64}$$

where the second inequality above follows from the definition of $\mathcal{L}$ and assumption (L2). Similarly, one can prove that $\mu_k \geq \underline{\lambda}$ for all $k \geq 0$. Further, by the BP 1 it follows that $\lambda_k \leq \lambda_{(1,k)}$ and $\mu_k \leq \mu_{(1,k)}$ for all $k \geq 0$. Thus, it holds

$$\lambda_k, \mu_k \in [\underline{\lambda}, \overline{\lambda}], \qquad \forall k \geq 0.$$

Consequently, assumption (L2) implies condition (A2).

Proposition 5. *Suppose that assumptions* (L1) *and* (L2) *hold. If $\Delta > 0$, then, for each iteration $k \geq 0$, (λ_k, x^k, b^k) and (μ_k, y^k, a^k) calculated by steps* 2′ *and* 3′*, respectively, satisfy* (21) *and* (22) *in Algorithm* 2.

Proof. Let λ_k, x^k and b^k be computed via line 2′. Since they are the outputs of the BP 1 with $T = B$, $z = \overline{z}^k$, $u = 0$, $v = -G^*\overline{w}^k$ and $c_1 = \lambda_{(1,k)}$, from lines 1, 2 and 4 of this procedure it follows

$$x^k = \overline{z}^k - \lambda_k(B\overline{z}^k + G^*\overline{w}^k) \qquad \text{and} \qquad b^k = Bx^k. \tag{65}$$

By rearranging the terms in the first equality in (65) and using the triangle inequality for norms and the Lipschitz continuity of B, we have

$$\left\|\lambda_k(b^k + G^*\overline{w}^k)\right\| = \left\|\lambda_k(Bx^k - B\overline{z}^k) + \overline{z}^k - x^k\right\| \leq (\lambda_k L_B + 1)\left\|x^k - \overline{z}^k\right\| \leq (\lambda_k \mathcal{L} + 1)\left\|x^k - \overline{z}^k\right\|. \quad (66)$$

Therefore, the BP 1, (66) and simple manipulations yield

$$\begin{aligned}\left\langle \overline{z}^k - x^k, b^k + G^*\overline{w}^k\right\rangle \geq & \Delta\left\|\overline{z}^k - x^k\right\|^2 = \frac{\Delta}{2}\left\|\overline{z}^k - x^k\right\|^2 + \frac{\Delta}{2}\left\|\overline{z}^k - x^k\right\|^2 \\ \geq & \frac{\Delta}{2}\left\|\overline{z}^k - x^k\right\|^2 + \frac{\Delta}{2(\mathcal{L}\lambda_k + 1)^2}\left\|\lambda_k(b^k + G^*\overline{w}^k)\right\|^2 \\ = & \frac{1}{2\lambda_k}\left[\Delta\lambda_k\left\|\overline{z}^k - x^k\right\|^2 + \frac{\Delta\lambda_k}{(\mathcal{L}\lambda_k + 1)^2}\left\|\lambda_k(b_k + G^*\overline{w}^k)\right\|^2\right].\end{aligned}$$

Further, since $\mathcal{L}\lambda_k + 1 > 1$, the equation above implies

$$\left\langle \overline{z}^k - x^k, b^k + G^*\overline{w}^k\right\rangle \geq \frac{1}{2\lambda_k}\cdot\frac{\Delta\lambda_k}{(\mathcal{L}\lambda_k + 1)^2}\left[\left\|\overline{z}^k - x^k\right\|^2 + \left\|\lambda_k(b_k + G^*\overline{w}^k)\right\|^2\right]. \quad (67)$$

Now, we use assumption (L2) and (64) to obtain

$$\frac{\Delta\lambda_k}{(\mathcal{L}\lambda_k + 1)^2} \geq \frac{\Delta\underline{\lambda}}{(\mathcal{L}\overline{\lambda} + 1)^2}.$$

From the fact that $\Delta, \mathcal{L} > 0$ and the definition of $\underline{\lambda}$ in (64), it follows that $\Delta\underline{\lambda} \leq \Delta/2(\mathcal{L} + \Delta) < 1$ and clearly $1/(\mathcal{L}\overline{\lambda} + 1)^2 < 1$. Hence, defining $\hat{\sigma}^2 := 1 - \Delta\underline{\lambda}/(\mathcal{L}\overline{\lambda} + 1)^2$, we have that $\hat{\sigma} \in (0, 1)$, and combining with (67) we deduce

$$\left\langle \overline{z}^k - x^k, b^k + G^*\overline{w}^k\right\rangle \geq \frac{1 - \hat{\sigma}^2}{2\lambda_k}\left[\left\|\overline{z}^k - x^k\right\|^2 + \left\|\lambda_k(b_k + G^*\overline{w}^k)\right\|^2\right].$$

The equation above, together with the second equality in (65) and Lemma 4, implies that λ_k and the triplet $(x^k, b^k, 0)$ satisfy the conditions in (21).

We next consider the outputs produced by line $3'$. Since the proof is similar to the previous one, we will only indicate the necessary modifications. Because μ_k, y^k and a^k are the outputs of the BP 1 with $T = A$, $z = G\overline{z}^k$, $u = \alpha_k G(x^k - \overline{z}^k)$, $v = \overline{w}^k$ and $c_1 = \mu_{(1,k)}$, it holds

$$y^k = G\overline{z}^k + \alpha_k G(x^k - \overline{z}^k) - \mu_k(A(G\overline{z}^k) - \overline{w}^k) \qquad \text{and} \qquad a^k = Ay^k.$$

Therefore, using the equalities above, the triangle inequality, and the Lipschitz continuity of A, in place of (66), we have

$$\left\|\alpha_k G(\overline{z}^k - x^k) + \mu_k(a^k - \overline{w}^k)\right\| = \left\|G\overline{z}^k - y^k + \mu_k(Ay^k - A(G\overline{z}^k))\right\| \leq (\mathcal{L}\mu_k + 1)\left\|G\overline{z}^k - y^k\right\|.$$

Thus, by the BP 1 and equation above, in analogy to (67), we obtain

$$\left\langle G\overline{z}^k - y^k, \frac{\alpha_k}{\mu_k}G(\overline{z}^k - x^k) + a^k - \overline{w}^k\right\rangle \geq \frac{1}{2\mu_k}\cdot\frac{\Delta\mu_k}{(\mathcal{L}\mu_k + 1)^2}\left[\left\|G\overline{z}^k - y^k\right\|^2 + \left\|\alpha_k G(\overline{z}^k - x^k) + \mu_k(a^k - \overline{w}^k)\right\|^2\right],$$

and as before, taking $\hat{\sigma}^2 = 1 - \Delta\underline{\lambda}/(\mathcal{L}\overline{\lambda} + 1)^2 \in (0, 1)$, we have

$$\frac{\Delta\mu_k}{(\mathcal{L}\mu_k + 1)^2} \geq 1 - \hat{\sigma}^2.$$

Now, combining the two equations above, we conclude

$$\left\langle G\overline{z}^k - y^k, \frac{\alpha_k}{\mu_k}G(\overline{z}^k - x^k) + a^k - \overline{w}^k\right\rangle \geq \frac{1 - \hat{\sigma}^2}{2\mu_k}\left[\left\|G\overline{z}^k - y^k\right\|^2 + \left\|\alpha_k G(\overline{z}^k - x^k) + \mu_k(a^k - \overline{w}^k)\right\|^2\right],$$

which by Lemma 4 gives that μ_k and $(y^k, a^k, 0)$ satisfy (22). □

Remark 4.

(i) We note that under assumptions (A1), (L1), (L2) and (A3), if the hypotheses of Theorems 4 (or Theorem 5) are satisfied, then the convergence of Algorithm 2, with the implementations in steps $2'$ and $3'$, is guaranteed.

(ii) If our goal is to find triplets (x, b, ε^x) and (y, a, ε^y) such that (44) is satisfied for fixed tolerances $\overline{\rho}, \overline{\varepsilon} > 0$, then Theorem 7, Proposition 5 and Lemma 7 give that Algorithm 2, with the implementations in steps $2'$ and $3'$, will find such triplets in at most

$$\mathcal{O}\left(\max\left\{\frac{d_0}{\zeta\,\overline{\rho}}, \frac{d_0^2}{\zeta\,\overline{\varepsilon}}\right\}\right)$$

iterations, with each one making at most

$$\mathcal{O}\left(\max\left\{2 + \log_2((\Delta + \mathcal{L})\overline{\lambda}), 1\right\}\right)$$

iterations of the BP 1.

(iii) If B is an affine operator, it is not necessary to use the BP 1 to find a valid triplet. For this case, even without knowing the Lipschitz constant, the operator B could be processed with only two forward steps. See [23, Section 5.2] for more details.

5 Numerical experiments

In this section, we apply Algorithm 2 to two common test problems in convex optimization. The main objective of these preliminary numerical experiments is to compare the computational performance of the inertial projective splitting method with its non-inertial counterpart and to illustrate the effect that different values of α_k have on the performance of the methods.

We recall that the convex minimization problem

$$\min_{z \in \mathcal{X}} f(Gz) + g(z) \tag{68}$$

where $f : \mathcal{Y} \to (-\infty, \infty]$ and $g : \mathcal{X} \to (-\infty, +\infty]$ are proper, convex and lower-semicontinuous functions, is equivalent to (1) (under appropriate constraint qualifications) with $A = \partial f$ and $B = \partial g$. Thus, we can apply Algorithm 2 to solve (68).

In our numerical experiments, we consider the *Lasso* and the *TV deblurring* problems, which are instances of (68). In both test problems, the function g has Lipschitz continuous gradient, which clearly gives that B is Lipschitz continuous. Therefore, we take $\lambda_k \equiv \sigma/L_B$ and use the scheme provided in the first line of (62) to calculate a valid triplet in step 2 of Algorithm 2 at each iteration k. We test Algorithm 2 with $\alpha_k \equiv \alpha = 0, -1, 1$ and another 7 uniformly distributed random values over $(-1, 1)$. Further, we set $\mu_k \equiv \mu$, and μ was chosen such that assumption (A3) is satisfied.

For a fair comparison with the non-inertial projective splitting method, we consider both algorithms with the same choices of the parameters $\alpha_k, \lambda_k, \mu_k$ and β_k. Also, the strategies for computing steps 2 and 3 in the methods are the same. We implement both algorithms in Matlab and, analogously to [37], we use the following condition to stop the methods:

$$\frac{|F(z^k) - F^*|}{F^*} \leq 10^{-4}. \tag{69}$$

In the equation above, $F(\cdot)$ denotes the objective function of (68), and F^* is the optimal value estimated after running 10^4 iterations of Algorithm 2 (with $\alpha_k \equiv 0$ and $\eta_k \equiv 0$) and taking the minimum objective value. We have preferred the stopping rule given by (69) since empirically we observed that it was satisfied faster than the condition (44) used in our complexity analysis. In the experiments, we refer to Algorithm 2 as the IR-PS method and the non-inertial projective splitting method as the PS method.

5.1 Lasso problem

In this section, we consider the *Lasso* problem [35]

$$\min_{z\in\mathbb{R}^n} \frac{1}{2}\left\|Mz - v\right\|_2^2 + \tau\left\|z\right\|_1, \tag{70}$$

where $M \in \mathbb{R}^{m\times n}$, $v \in \mathbb{R}^m$ and $\tau > 0$ is a regularization parameter. The goal of this problem is to find a sparse solution to the linear system of equations $Mz = v$. We observe that (70) is an instance of (68) taking $g(z) := (1/2)\left\|Mz - v\right\|_2^2$, $f(z) := \tau\left\|z\right\|_1$ and $G = I$.

It is well-known that the proximal operator of $A(z) = \partial(\tau\left\|\cdot\right\|_1)(z)$ has a closed form. Thus, to solve (70) via Algorithm 2, at each iteration k, we set $y^k = (\mu_k A + I)^{-1}((1-\alpha_k)\overline{z}^k + \alpha_k x^k + \mu_k\overline{w}^k)$ and $a^k = (1/\mu_k)((1-\alpha_k)\overline{z}^k + \alpha_k x^k + \mu_k\overline{w}^k - y^k)$ (observe that in this case $r^{y,k} = 0$ and $\varepsilon_k^y = 0$).

In the numerical tests, we use four non-artificial data sets taken from the UCI Machine Learning Repository [17]:

- the drive face dataset (DriveFace) with $m = 606$ and $n = 6400$;
- communities and crime dataset (Communities) with $m = 2215$ and $n = 145$;
- Breast Cancer Wisconsin (Diagnostic) dataset (Wisconsin) with $m = 569$ and $n = 30$;
- California Department of Transportation dataset (PEMS) with $m = 267$ and $n = 138672$.

We normalize the data for all problems and take $\tau = 0.1\left\|M^T v\right\|_\infty$. We use $\sigma = 0.24$, and $\eta_k \equiv 0.5$ because we have empirically found that it works well for all problems. Using equation (43), we compute $\overline{\beta} = 0.3425$ with $\overline{\eta} = 0.57$, and set $\beta_k \equiv \overline{\beta}$.

Tables 1, 2, 3 and 4 present the results of the numerical experiments for the Lasso problem for each dataset. They report, for different values of α (referred to as Alpha in the tables), the iteration count (It) and total time in seconds (time(s)) required for the IR-PS and PS to satisfy (69). In each table, the smallest value in each row appears in bold. We observe that the IR-PS solves almost all the problems faster than the PS. As the last line in each table indicates, the IR-PS takes, in the mean, approximately 15% less time than the PS, and executes around 1000 fewer iterations. Also, observe that the best IR-PS method (in terms of iterations and running time) for all the datasets is always for a value of α different from zero ($\alpha = 0$ corresponds to the parallel IR-PS of [4]).

For the fastest value of α for each data set, we report in figures 1 and 2, respectively, the error and residual curves for both methods. The error is given by (69), whereas the residue is defined as

$$\max\{\left\|a^k + b^k\right\|, \left\|x^k - y^k\right\|\},$$

see (44).

5.2 Total variation deblurring problem

The Total Variation deblurring problem seeks to estimate an unknown original image $z \in \mathbb{R}^{N\times N}$ from an observed blurred image $v \in \mathbb{R}^{N\times N}$ by solving the minimization problem

$$\min_{z\in\mathbb{R}^{N\times N}} \frac{1}{2}\left\|Mz - v\right\|_F^2 + \tau\sum_{i,j=1}^{N}\left\|(\nabla z)_{i,j}\right\|_2, \tag{71}$$

where $\nabla : \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times N}\times\mathbb{R}^{N\times N}$ is the *discrete gradient* operator (see [13] for the precise definition), $\tau > 0$ is a regularization parameter, $\left\|\cdot\right\|_2$ is the Euclidean norm in $\mathbb{R}^2$, and $M : \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times N}$ is a linear operator that represents some blurring operator. Problem (71) is an instance of (68) with $g(z) = \frac{1}{2}\left\|Mz - v\right\|_F^2$, $f(p) = \tau\sum_{i,j=1}^{N}\left\|p_{i,j}\right\|_2$ and $G = \nabla$.

	IR-PS		PS	
Alpha	time(s)	It	time(s)	It
1	**44.33**	**5547**	55.06	6881
-1	**65.05**	**8140**	70.79	8840
0	**46.15**	**5803**	52.19	6447
-0.8147	**59.45**	**7450**	66.46	8317
-0.1270	**47.33**	**5909**	52.88	6598
-0.6324	**56.98**	**7110**	66.08	8231
0.2785	**42.70**	**5375**	49.42	6240
0.5469	**40.36**	**5045**	46.29	5775
0.9575	**44.65**	**5547**	56.70	7058
-0.3584	**49.67**	**6221**	58.40	7303
Geometric Mean	**49.67**	**6214.7**	57.43	7169

Table 1: Results for the Lasso problem with the DriveFace dataset.

	IR-PS		PS	
Alpha	time(s)	It	time(s)	It
1	**0.5428**	**2269**	0.6564	2838
-1	**1.2063**	**4985**	1.3776	6808
0	**0.6888**	**3625**	0.7603	3838
-0.8147	**1.0097**	**5080**	1.2829	6559
-0.1270	0.9349	**4021**	**0.9050**	4774
-0.6324	**0.8943**	**4888**	1.0568	5609
0.2785	**0.5831**	**3066**	0.7329	3821
0.5469	**0.4718**	**2558**	0.5958	3159
0.9575	**0.4183**	**2228**	0.5592	3028
-0.3584	**0.7614**	**3966**	0.9900	5169
Geometric Mean	**0.7511**	**3668.6**	0.8917	4560.3

Table 2: Results for the Lasso problem with the Communities dataset.

We recall that for $B = \partial g$, we use the strategy given in the first line of (62) for implementing step 2 of Algorithm 2. On the other hand, $A = \partial f$ has a computational simple proximal operator. Hence, we compute $y^k = (\mu_k A + I)^{-1}(\nabla((1-\alpha_k)\overline{z}^k + \alpha_k x^k) + \mu_k \overline{w}^k)$ and $a_k = (1/\mu_k)(\nabla((1-\alpha_k)\overline{z}^k + \alpha_k x^k) + \mu_k \overline{w}^k - y^k)$.

In the numerical simulations, we consider the 256×256 *Cameraman* image blurred by a 4×4 Gaussian blur with standard deviation 2, followed by an additive normal noise with zero mean and standard deviation 10^{-4}. The regularization parameter τ was set to 10^{-4}. We use $\sigma = 1/\left\|\nabla\right\|^{1.3}$, and $\eta_k \equiv \eta = 0.1$ since we empirically observed that this choice performs better for this problem. Using (43), we compute $\overline{\beta} = 1.519$ with $\overline{\eta} = 0.17 > \eta$ and set $\beta_k = \overline{\beta}$ for all k.

Table 5 displays the results of this experiment. The columns mean: Alpha, the value of α used by the IR-PS and the PS; time(s), the time (in seconds) needed to reach the desired accuracy (69); and It, the number of iterations performed by the methods. We again highlight the smallest value in each row, and we mark with an asterisk (*) the instances that reach the maximum number of allowed iterations (10000). Observe that for all α, the IR-PS method requires fewer iterations and it was faster than the PS method, taking approximately less than one third of the execution time and iterations, in the mean. Also, observe that the best method tested for this problem, in terms of iterations and running time, was the IR-PS method with $\alpha = -0.8147$, with a 30% reduction of the iterations and the execution time of the parallel IR-PS method ($\alpha = 0$).

In figure 3, we plotted the error and residual curves of the IR-PS and PS methods with $\alpha = 1$, which is

	IR-PS		PS	
Alpha	time(s)	It	time(s)	It
1	0.0547	1881	**0.0507**	**1815**
-1	**0.1177**	**3931**	0.1426	4987
0	**0.0815**	**2765**	0.1022	3516
-0.8147	**0.1093**	**3755**	0.1296	4464
-0.1270	**0.0759**	**2685**	0.1159	3800
-0.6324	**0.1810**	**3369**	0.2141	4324
0.2785	**0.0788**	**2652**	0.0952	2995
0.5469	**0.0642**	**2203**	0.0736	2405
0.9575	**0.0466**	**1534**	0.0607	1925
-0.3584	**0.0923**	**3046**	0.1161	3928
Geometric Mean	**0.0902**	**2782.1**	0.1101	3315.9

Table 3: Results for the Lasso problem with the Wisconsin dataset.

	IR-PS		PS	
Alpha	time(s)	It	time(s)	It
1	**472.57**	**6036**	616.97	7831
-1	**612.82**	**7925**	740.65	9572
0	**495.11**	**6364**	578.82	7432
-0.8147	**588.41**	**7533**	710.94	9095
-0.1270	**499.74**	**6312**	597.38	7583
-0.6324	**555.49**	**7048**	676.15	8581
0.2785	**479.59**	**6027**	556.78	7065
0.5469	**434.45**	**5517**	550.12	6977
0.9575	**480.67**	**6065**	636.19	8062
-0.3584	**523.79**	**6658**	621.95	7922
Geometric Mean	**514.26**	**6548.5**	628.60	8012

Table 4: Results for the Lasso problem with the PEMS dataset.

the fastest α for both methods. In this case, the residue is given by

$$\max\{\left\|\nabla^* a^k + b^k\right\|, \left\|\nabla x^k - y^k\right\|\}$$

see (44).

	IR-PS		PS	
Alpha	time(s)	It	time(s)	It
1	**42.99**	**2777**	64.25	3989
-1	**10.91**	**704**	186.33	10000*
0	**12.35**	**777**	95.24	5830
-0.8147	**7.21**	**510**	43.66	2966
-0.1270	**10.57**	**691**	57.25	3960
-0.6324	**11.14**	**785**	71.14	4926
0.2785	**21.57**	**1473**	47.81	3392
0.5469	**23.45**	**1670**	40.09	2862
0.9575	**36.64**	**2474**	65.19	4649
-0.3584	**9.49**	**679**	44.39	3126
Geometric Mean	**18.63**	**1254**	71.54	4570

Table 5: Results for the TV deblurring problem.

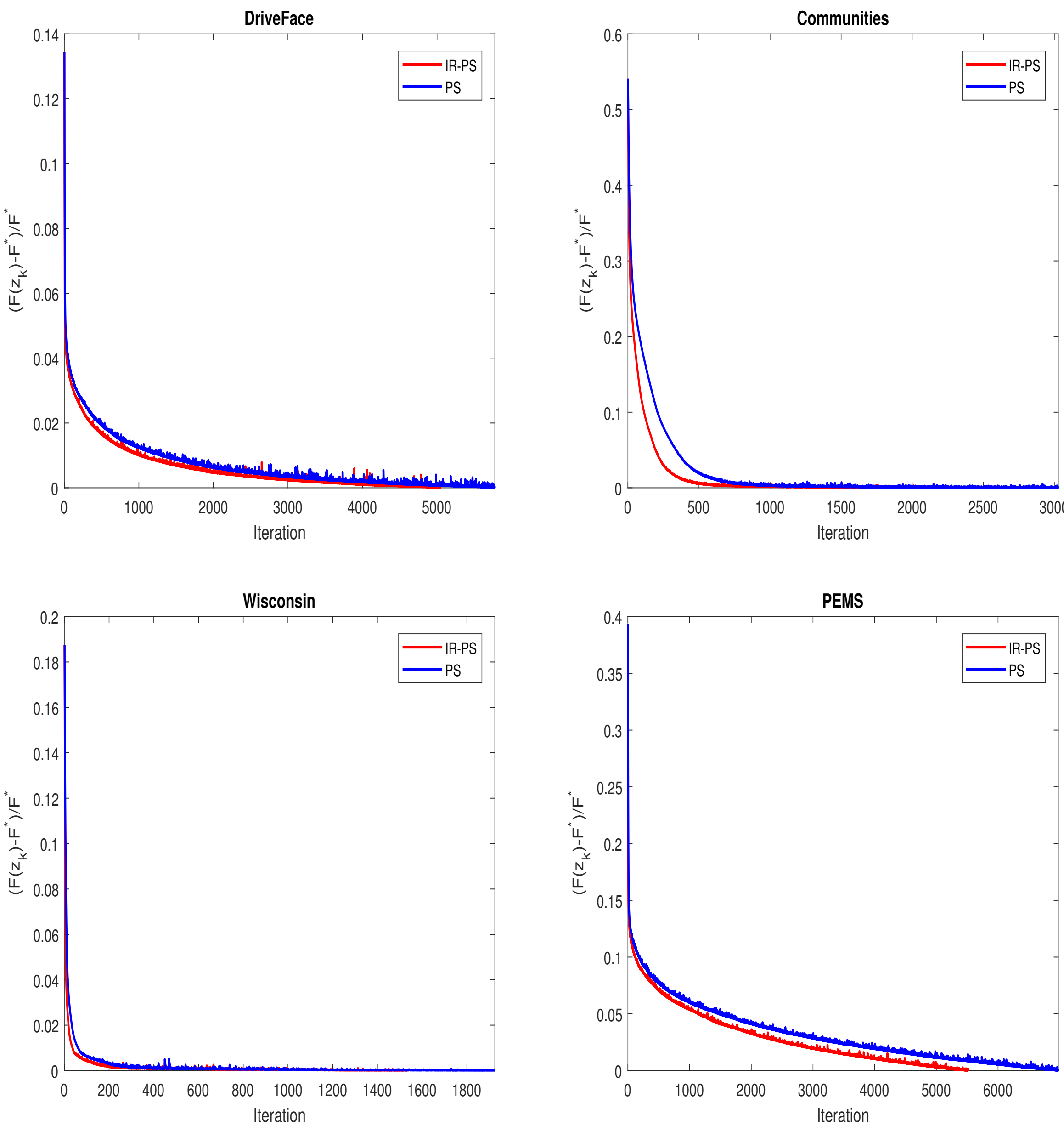

Figure 1: Error curves of the IR-PS and PS methods in the Lasso problems. (top left) $\alpha = 0.5469$, (top right) $\alpha = 0.9575$, (bottom left) $\alpha = 0.9575$, (bottom right) $\alpha = 0.5469$.

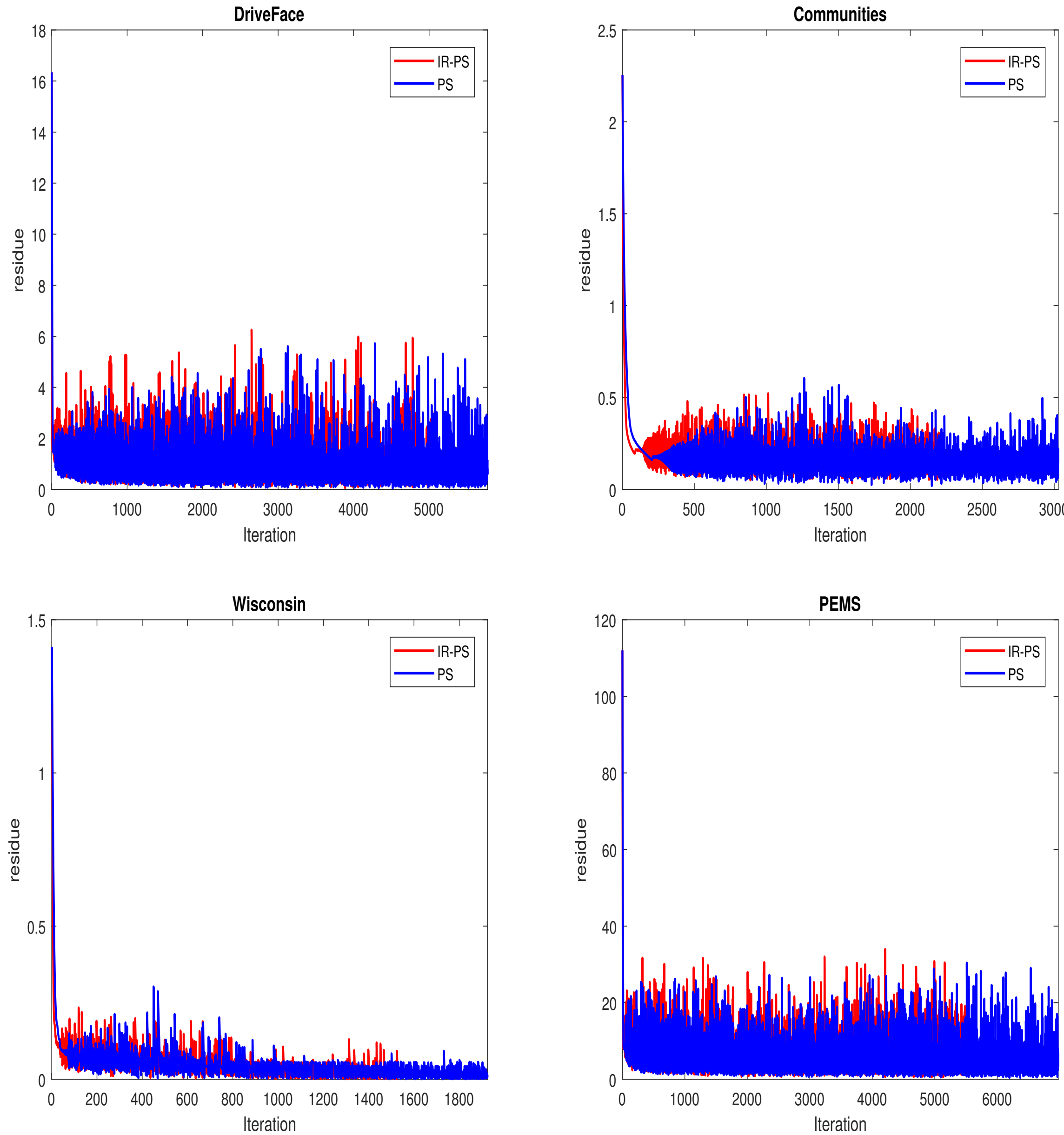

Figure 2: Residual curves of the IR-PS and PS methods in the Lasso problems. (top left) $\alpha = 0.5469$, (top right) $\alpha = 0.9575$, (bottom left) $\alpha = 0.9575$, (bottom right) $\alpha = 0.5469$.

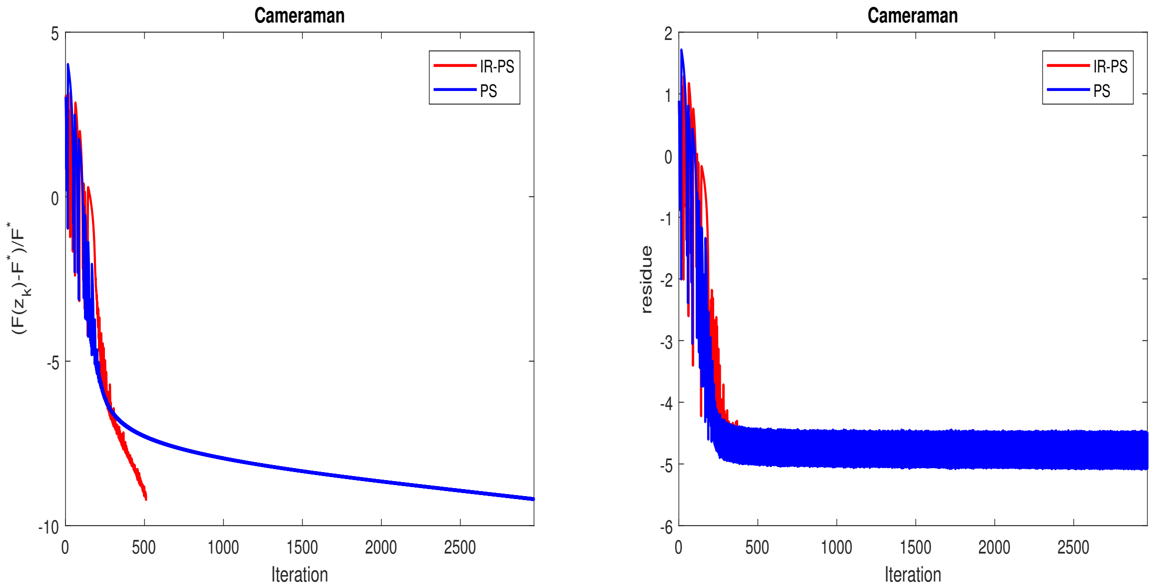

Figure 3: Error curve (left) and residual curve (right), in the logarithmic scale, of the IR-PS and PS methods with $\alpha = -0.8147$ for the TV-deblurring problem.

Declarations

Conflict of interest

The author has no competing financial interests or personal relationships that might be perceived to influence the results and/or discussion reported in this paper.

Authors' contributions

The author wrote the main manuscript text, performed the numerical implementations, prepared all figures and tables, and reviewed the manuscript.

References

[1] Alotaibi, A., Combettes, P.L., Shahzad, N.: Solving coupled composite monotone inclusions by successive fejér approximations of their kuhn–tucker set. SIAM Journal on Optimization. 24(4), 2076–2095 (2014)

[2] Alvarez, F., Attouch, H.: An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. Set Valued Var. Anal. 9(1–2), 3–11 (2001)

[3] Alves, M.M., Eckstein, J., Geremia, M., Melo J.G.: Relative-error inertial-relaxed inexact versions of Douglas-Rachford and ADMM splitting algorithms. Comput. Optim. Appl. 75, 389–422 (2020) https://doi.org/10.1007/s10589-019-00165-y

[4] Alves, M.M., Geremia, M., Marcavillaca, R.T.: A relative-error inertial-relaxed inexact projective splitting algorithm. Accepted for publication in Journal of Convex Analysis. (2023)

[5] Attouch, H., Cabot, A.: Convergence of a Relaxed Inertial Forward–Backward Algorithm for Structured Monotone Inclusions. Appl. Math. Optim. 80, 547–598 (2019) https://doi.org/10.1007/s00245-019-09584-z

[6] Attouch, H., Cabot, A.: Convergence of a relaxed inertial proximal algorithm for maximally monotone operators. Math. Program. 184, 243–287 (2020) https://doi.org/10.1007/s10107-019-01412-0

[7] Bauschke, H.H., Borwein, J.M.: On projection algorithms for solving convex feasibility problems. SIAM Rev. 38(3), 367–426 (1996) https://doi.org/10.1137/S0036144593251710

[8] Bauschke, H.H., Combettes, Reich, S.: The asymptotic behavior of the composition of two resolvents. Nonlinear Anal. 60, 283–301 (2005)

[9] Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. 2(1), 183–202 (2009)

[10] Bot, R.I., Csetnek, E.R., Hendrich, C.: Inertial Douglas-Rachford splitting for monotone inclusion problems. Appl. Math. Comput. 256, 472–487 (2015)

[11] Burachik, R.S., Iusem, A.N., Svaiter, B.F.: Enlargement of monotone operators with applications to variational inequalities. Set-Valued Anal. 5(2), 159–180 (1997)

[12] Burachik, R.S., Svaiter, B.F.: ε-enlargements of maximal monotone operators in Banach spaces. Set-Valued Analysis. 7(2), 117–132 (1999)

[13] Chambolle, A.: An algorithm for total variation minimization and applications. J.Math. Imaging Vis. 20, 89–97 (2004)

[14] Chen, C., Chan, R.H., Ma, S., Yang, J.: Inertial proximal ADMM for linearly constrained separable convex optimization. SIAM J. Imaging Sci. 8(4), 2239–2267 (2015)

[15] Combettes, P.L., Eckstein, J.: Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions. Mathematical Programming. 168(1-2), 645–672 (2018)

[16] Douglas Jr., J., Rachford Jr., H.H.: On the numerical solution of heat conduction problems in two and three space variables. Trans. Am. Math. Soc. 82, 421–439 (1956)

[17] Dua, D., Graff, C.: UCI machine learning repository. http://archive.ics.uci.edu/ml (2017)

[18] Eckstein, J., Svaiter, B.F.: A family of projective splitting methods for the sum of two maximal monotone operators. Mathematical Programming. 111(1-2), 173–199 (2008)

[19] Eckstein, J., Svaiter, B.F.: General projective splitting methods for sums of maximal monotone operators. SIAM Journal on Control and Optimization. 48(2), 787–811 (2009)

[20] Eckstein , J.: A simplified form of block-iterative operator splitting and an asynchronous algorithm resembling the multi-block alternating direction method of multipliers. J. Optim. Theory Appl. 173, 155–182 (2017)

[21] Glowinski, R., Osher, S.J., Yin, W.: Splitting Methods in Communication, Imaging, Science, and Engineering. Springer Publishing Company, Incorporated (2017)

[22] Johnstone, P.R., Eckstein, J.: Convergence Rates for Projective Splitting. SIAM J. Optim. 29(3), 1931–1957 (2019)

[23] Johnstone, P.R., Eckstein, J.: Projective Splitting with Forward Steps. Math. Program. 191, 631–670 (2022) https://doi.org/10.1007/s10107-020-01565-3

[24] Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. 16(6), 964–979 (1979)

[25] Lorenz, D.A., Pock, T.: An inertial forward-backward algorithm for monotone inclusions. J. Math. Imaging Vis. 51(2), 311–325 (2015)

[26] Machado, M.P.: On the Complexity of the Projective Splitting and Spingarn's Methods for the Sum of Two Maximal Monotone Operators. Journal of Optimization Theory and Applications. 178(1), 153–190 (2018)

[27] Machado, M.P., Sicre, M.R.: A projective splitting method for monotone inclusions: iteration-complexity and application to composite optimization. Journal of Optimization Theory and Applications. 198, 552–587 (2023) https://doi.org/10.1007/s10957-023-02214-3

[28] Minty, G.J.: Monotone (nonlinear) operators in Hilbert space. Duke Math. J. 29, 341–346 (1962)

[29] Monteiro, R.D.C., Svaiter, B.F.: On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. SIAM J. Optim. 20(6), 2755–2787 (2010)

[30] Peaceman, D.W., Rachford Jr., H.H.: The numerical solution of parabolic and elliptic differential equations. J. Soc. Ind. Appl. Math. 3, 28–41 (1955)

[31] Passty, G.B.: Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. J. Math. Anal. Appl. 72(2), 383–390 (1979)

[32] Sicre, M.R.: On the complexity of a hybrid proximal extragradient projective method for solving monotone inclusion problems. Comput. Optim. Appl. 76, 991–1019 (2020) https://doi.org/10.1007/s10589-020-00200-3

[33] Solodov, M.V., Svaiter, B.F.: A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. Set-Valued Anal. 7(4), 323–345 (1999)

[34] Solodov, M.V., Svaiter, B.F.: A unified framework for some inexact proximal point algorithms. Numer. Funct. Anal. Optim. 22(7-8), 1013–1035 (2001)

[35] Tibshirani, R.: Regression shrinkage and selection via the lasso. J. Roy. Statist. Soc. Ser. B. 58(1), 267–288 (1996)

[36] Tseng, P.: A modified forward-backward splitting method for maximal monotone mappings. SIAM J. Control Optim. 38(2), 431–446 (2000)

[37] Villa, S., Salzo, S., Baldassarre, L., Verri, A.: Accelerated and inexact forward-backward algorithms. J SIAM J. Optim. 23(3), 1607–1633 (2013)