A relative-error inertial-relaxed inexact projective splitting algorithm

M. Marques Alves [*] Marina Geremia [†] Raul T. Marcavillaca [‡]

February 18, 2020

Abstract

For solving structured monotone inclusion problems involving the sum of finitely many maximal monotone operators, we propose and study a relative-error inertial-relaxed inexact projective splitting algorithm. The proposed algorithm benefits from a combination of inertial and relaxation effects, which are both controlled by parameters within a certain range. We propose sufficient conditions on these parameters and study the interplay between them in order to guarantee weak convergence of sequences generated by our algorithm. Additionally, the proposed algorithm also benefits from inexact subproblem solution within a relative-error criterion. Simple numerical experiments on LASSO problems indicate some improvement when compared with previous (noninertial and exact) versions of projective splitting.

2000 Mathematics Subject Classification: 90C25, 90C30, 47H05.

Key words: operator splitting, projective splitting, inertial algorithms, relative-error, monotone operators.

1 Introduction

Let $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_n$ be real Hilbert spaces and let $\langle \cdot, \cdot \rangle$ and $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$ denote the inner product and norm (respectively) in $\mathcal{H}_i$ $(i = 0, \dots, n)$. Assume that $\mathcal{H}_0 = \mathcal{H}_n$. Let $\boldsymbol{\mathcal{H}} := \mathcal{H}_0 \times \cdots \times \mathcal{H}_{n-1}$ be endowed with the inner product and norm defined, respectively, as follows (for some $\gamma > 0$):

$$\langle (z, w), (z', w') \rangle_\gamma = \gamma \langle z, z' \rangle + \sum_{i=1}^{n-1} \langle w_i, w_i' \rangle, \quad \|(z, w)\|_\gamma^2 = \gamma \|z\|^2 + \sum_{i=1}^{n-1} \|w_i\|^2, \tag{1}$$

where $z, z' \in \mathcal{H}_0$ and $w := (w_1, \dots, w_{n-1}), w' := (w_1', \dots, w_{n-1}') \in \mathcal{H}_1 \times \ldots \times \mathcal{H}_{n-1}$.

Consider the monotone inclusion problem of finding $z \in \mathcal{H}_0$ such that

$$0 \in \sum_{i=1}^{n} G_i^* T_i G_i(z) \tag{2}$$

where $n \geq 2$ and the following assumptions hold:

[*]Departamento de Matemática, Universidade Federal de Santa Catarina, Florianópolis, Brazil, 88040-900 (maicon.alves@ufsc.br). The work of this author was partially supported by CNPq grants no. 406250/2013-8, 306317/2014-1 and 405214/2016-2.

[†]Departamento de Matemática, Universidade Federal de Santa Catarina, Florianópolis, Brazil, 88040-900 (marinageremia@yahoo.com.br).

[‡]Departamento de Matemática, Universidade Federal de Santa Catarina, Florianópolis, Brazil, 88040-900 (rtm1111@hotmail.com). The work of this author was partially supported by CAPES.

(A1) For each $i = 1, \dots, n$, the operator $T_i : \mathcal{H}_i \rightrightarrows \mathcal{H}_i$ is (set-valued) maximal monotone and $G_i : \mathcal{H}_0 \to \mathcal{H}_i$ is a bounded linear operator.

(A2) The linear operator G_n is equal to the identity map in $\mathcal{H}_0 = \mathcal{H}_n$, i.e., $G_n : z \mapsto z$ for all $z \in \mathcal{H}_0$.

(A3) The solution set of (2) is nonempty, i.e., there exists at least one $z \in \mathcal{H}_0$ satisfying the inclusion in (2).

Problem (2) appears in different fields of applied mathematics and optimization, including machine learning, inverse problems and image processing [1, 13, 15], specially in connection with the convex optimization problem

$$\min_{z \in \mathcal{H}_0} \sum_{i=1}^{n} f_i(G_i z) \tag{3}$$

where, for $i = 1, \dots, n$, each $f_i : \mathcal{H}_i \to (-\infty, +\infty]$ is proper, closed and convex. Indeed, under mild assumptions on f_i and G_i, the minimization problem (3) is equivalent to the monotone inclusion problem (2) with $T_i = \partial f_i$ $(i = 1, \dots, n)$.

A very popular strategy to find approximate solutions of (2) is that of (monotone) operator splitting algorithms, which traces back to the development of some well-known numerical schemes like the Douglas-Rachford splitting algorithm, Spingarn's method of partial inverses, among others.

The family of *projective splitting algorithms* for solving (2), originated in [12] for the case that G_i is the identity $(i = 1, \dots, n)$, and later on developed in different directions, e.g., in [1, 13, 15], has deserved a lot of attention in modern operator splitting research, mainly due to its flexibility (when compared to other classes of operator splitting algorithms) regarding parameters and the activation of T_i and G_i separately during the iterative process.

The derivation of the class of projective splitting algorithms can be motivated as follows. First note that using Assumption (A2) above, we obtain that (2) can equivalently be written as

$$0 \in \sum_{i=1}^{n-1} G_i^* T_i G_i(z) + T_n(z) \tag{4}$$

which, in turn, is clearly equivalent to the (feasibility) problem of finding a point in the *extended solution set* of (2) (or (4)):

$$\mathcal{S} := \left\{ (z, w_1, \dots, w_{n-1}) \in \boldsymbol{\mathcal{H}} \;\middle|\; w_i \in T_i(G_i z),\, i = 1, \dots, n-1,\, -\sum_{i=1}^{n-1} G_i^* w_i \in T_n(z) \right\}. \tag{5}$$

Since $\mathcal{S}$ is nonempty (see Assumption (A3)), closed and convex (see, e.g., [1, 12]) in $\boldsymbol{\mathcal{H}}$, it follows that problem (2) reduces to the task of finding a point in $\mathcal{S}$ (fact that motivates the abstract framework developed in Section 2 below).

Note now that, if we pick $y_i^k \in T_i(x_i^k)$ $(i = 1, \dots, n)$, then from the monotonicity of T_i and the inclusions in (5), it follows that

$$\sum_{i=1}^{n} \langle G_i z - x_i^k, y_i^k - w_i \rangle \le 0 \qquad \forall (z, w_1, \dots, w_{n-1}) \in \mathcal{S}, \tag{6}$$

where

$$w_n := -\sum_{i=1}^{n-1} G_i^* w_i. \tag{7}$$

The inequality (6) says, in particular, that $\{(x_i^k, y_i^k)\}_{i=1}^n$ defines a function of $(z, w_1, \dots, w_{n-1})$ which is negative in $\mathcal{S}$. Since this function can be proved to be affine (see, e.g., Lemma 3.1 below), it follows from (6) that it defines a semispace in $\boldsymbol{\mathcal{H}}$ containing the extended solution set $\mathcal{S}$.

Based on the exposed above, it follows that the main mechanism behind the idea of projective splitting algorithms is basically: at the iteration $(z^k, w_1^k, \dots, w_{n-1}^k)$, pick, for each $i = 1, \dots, n$, a pair (x_i^k, y_i^k) in the graph of T_i and then update the current iterate $p^k := (z^k, w_1^k, \dots, w_{n-1}^k)$ to $p^{k+1} := (z^{k+1}, w_1^{k+1}, \dots, w_{n-1}^{k+1})$ by projecting p^k onto the semispace defined by the affine function given in the left hand side of (6). Computation of (x_i^k, y_i^k) is in general performed by (inexactly) activating the resolvent $(T_i + I)^{-1}$ operator of each T_i to guarantee, in particular, that the current iterate $(z^k, w_1^k, \dots, w_{n-1}^k)$ belongs to the positive side of the corresponding hyperplane.

In this paper, we propose and study a relative-error inertial-relaxed inexact projective splitting algorithm for solving (1) and, in particular, for solving the convex program (3). Inertial algorithms for solving monotone inclusions of the form $0 \in T(z)$, where T is maximal monotone, were first proposed in [2], and since then developed by different authors and in different directions of research (see, e.g., [3, 7, 8, 10] and references there in). At a current iterate, say p^k, the inertial effect in the iterative process is produced by an extrapolation step of the form (see also Algorithm 1 and Figure 1 below):

$$\widehat{p}^k = p^k + \alpha_k (p^k - p^{k-1}).$$

Since $\alpha_k \geq 0$ controls the magnitude of extrapolation performed in the direction of the vector $p^k - p^{k-1}$, it follows that the asymptotic behavior and size of α_k have a direct influence in the convergence analysis of inertial-type algorithms. An usual sufficient condition [2] imposed on the sequence $\{\alpha_k\}$, with guarantee of weak convergence of $\{p^k\}$, is that $\{\alpha_k\}$ is nondecreasing and $\alpha_k < 1/3$ for all $k \geq 0$. The upper bound $1/3$ has been recently improved in combination with relaxation effects [3, 5].

The main goal of this paper is to develop a projective splitting algorithm-type algorithm for solving (2) with both inertial and relaxation effects and, additionally, with inexact subproblems solution within relative-error criterion. Up to the authors knowledge, this is the first time in the literature that inertial effects are considered in projective splitting-like algorithms. Our main algorithm is Algorithm 2 from Section 3, for which the convergence is studied in Theorems 3.5 and 3.6, under flexible assumptions on the inertial and relaxation parameters. Motivated by the discussion above that (2) is equivalent to the problem of finding a point in the closed and convex set $\mathcal{S}$ as in (5), we first introduce in Section 2 an inertial-relaxed separator-projector method for solving the (feasibility) problem of finding points in closed convex subsets of Hilbert spaces.

The following well-known property will be useful in this paper: for all x, y in a real Hilbert space $\mathcal{H}$ and $t \in \mathbb{R}$, it holds that

$$\|tx + (1-t)y\|^2 = t\|x\|^2 + (1-t)\|y\|^2 - t(1-t)\|x - y\|^2. \tag{8}$$

We shall also use the following inequality:

$$\left\|\sum_{i=1}^{n} x_i\right\|^2 \leq n \sum_{i=1}^{n} \|x_i\|^2. \tag{9}$$

2 An inertial-relaxed separator-projection method

In this section, we propose and study a general separator-projection framework (Algorithm 1) for finding a point in a closed and convex subset of a Hilbert space. The main motivation comes from the fact (as previously discussed in Section 1) that the monotone inclusion problem (2) can be reformulated as the problem of finding a point in the extended solution set $\mathcal{S}$ as in (5). Algorithm 1 will be used in Section 3 to analyze the convergence of the main algorithm proposed in this paper (namely Algorithm 2) for solving (34).

Let $\mathcal{H}$ be a real Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$. We denote the gradient of an affine function $\varphi : \mathcal{H} \to \mathbb{R}$ by the usual notation $\nabla\varphi$ and, in this case, we also write $\varphi(z) = \langle\nabla\varphi, z\rangle + \varphi(0)$ for all $z \in \mathcal{H}$.

Algorithm 1. An inertial-relaxed linear separator-projection method for finding a point in a nonempty closed convex set $\mathcal{S} \subset \mathcal{H}$

(0) Let $p^0 = p^{-1} \in \mathcal{H}$, $\alpha \in [0, 1)$ and $0 < \underline{\beta} < \overline{\beta} < 2$ be given and let $k \leftarrow 0$.

(1) Choose $\alpha_k \in [0, \alpha]$ and define

$$\widehat{p}^k = p^k + \alpha_k(p^k - p^{k-1}). \tag{10}$$

(2) Find an affine function φ_k such that $\nabla\varphi_k \neq 0$ and $\varphi_k(p) \leq 0$ for all $p \in \mathcal{S}$. Choose $\beta_k \in [\underline{\beta}, \overline{\beta}]$ and set

$$p^{k+1} = \widehat{p}^k - \frac{\beta_k \max\{0, \varphi_k(\widehat{p}^k)\}}{\|\nabla\varphi_k\|^2}\nabla\varphi_k. \tag{11}$$

(3) Let $k \leftarrow k + 1$ and go to step 1.

Remarks.

(i) Letting $\widetilde{p}^{k+1}$ be the (orthogonal) projection of $\widehat{p}^k$ onto the semispace $\{p \in \mathcal{H} \mid \varphi_k(p) \leq 0\}$, i.e.,

$$\widetilde{p}^{k+1} = \widehat{p}^k - \frac{\max\{0, \varphi_k(\widehat{p}^k)\}}{\|\nabla\varphi_k\|^2}\nabla\varphi_k \tag{12}$$

and using (11) we conclude that

$$p^{k+1} = \widehat{p}^k + \beta_k(\widetilde{p}^{k+1} - \widehat{p}^k). \tag{13}$$

(ii) Note that (10) and (13) illustrate the different effects promoted in Algorithm 1 by inertia and relaxation, which are respectively controlled by the parameters α_k and β_k. See Figure 1 below.

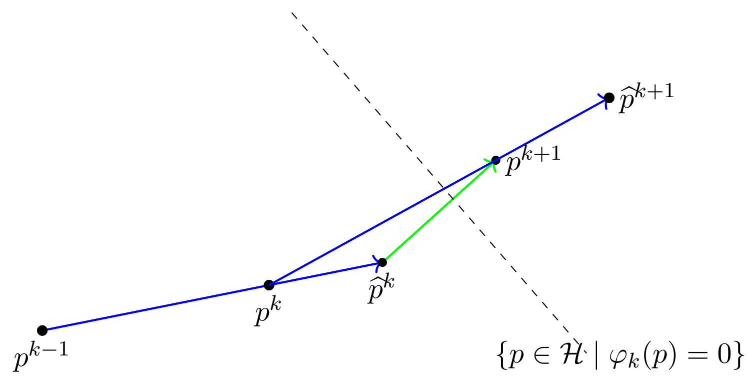


Figure 1: Geometric interpretation of steps (10) and (11) in Algorithm 1. The (overrelaxed) projection step (11) is orthogonal to the separating hyperplane $\{p \in \mathcal{H} \mid \varphi_k(p) = 0\}$, which can differ from the direction between p^{k-1}, p^k, and $\widehat{p}^k$ when $\alpha_k > 0$.

(iii) If $\alpha_k \equiv 0$, in which case $\widehat{p}^k = p^k$ in (10), then it follows that Algorithm 1 reduces to the well-known linear separator-projection method for finding a point in $\mathcal{S} \subset \mathcal{H}$ (see, e.g., [1]).

(iv) As we mentioned early, Algorithm 1 will be used in the next section for analyzing the convergence of Algorithm 2. The main convergence results for Algorithm 1 will be stated in this section, in Theorems 2.2 and 2.3 below.

Next lemma plays the role of Fejér-monotonicity for Algorithm 1 and will be used in the proofs of Theorems 2.2 and 2.3.

Lemma 2.1. *Consider the sequences evolved by* Algorithm 1 *and let* $\widetilde{p}^{k+1}$ *be as in* (12)*. For an arbitrary* $p \in \mathcal{S}$*, define*

$$h_k = \|p^k - p\|^2 \qquad \forall k \geq -1. \tag{14}$$

Then the following hold:

(a) *For all* $k \geq 0$*,*

$$h_{k+1} - h_k - \alpha_k(h_k - h_{k-1}) \leq \alpha_k(1 + \alpha_k)\|p^k - p^{k-1}\|^2 - s_{k+1},$$

where

$$s_{k+1} := \beta_k(2 - \beta_k)\|\widehat{p}^k - \widetilde{p}^{k+1}\|^2 \qquad \forall k \geq 0. \tag{15}$$

(b) *For all* $k \geq 0$*,*

$$h_{k+1} - h_k - \alpha_k(h_k - h_{k-1}) \leq \gamma_k\|p^k - p^{k-1}\|^2 - (2 - \overline{\beta})\overline{\beta}^{-1}(1 - \alpha_k)\|p^{k+1} - p^k\|^2, \tag{16}$$

where

$$\gamma_k := 2\left(1 - \overline{\beta}^{-1}\right)\alpha_k^2 + 2\overline{\beta}^{-1}\alpha_k \qquad \forall k \geq 0. \tag{17}$$

Proof. (a) We shall first prove that

$$\|p^{k+1}-p\|^2+\beta_k(2-\beta_k)\|\widehat{p}^k-\widetilde{p}^{k+1}\|^2\leq\|\widehat{p}^k-p\|^2\qquad\forall p\in\mathcal{S},\tag{18}$$

where $\widetilde{p}^{k+1}$ is as in (12), i.e., it is the projection of $\widehat{p}^k$ onto the semispace $\{p\in\mathcal{H}\mid\varphi_k(p)\leq 0\}$. To this end, note first that, for all $p\in\mathcal{S}$,

$$\begin{aligned}\|\widehat{p}^k-p\|^2-\|\widetilde{p}^{k+1}-p\|^2&=\|\widehat{p}^k-\widetilde{p}^{k+1}\|^2+2\langle\widehat{p}^k-\widetilde{p}^{k+1},\widetilde{p}^{k+1}-p\rangle\\&\geq\|\widehat{p}^k-\widetilde{p}^{k+1}\|^2\end{aligned}\tag{19}$$

where we have used (12) and the fact that $\mathcal{S}\subset\{p\in\mathcal{H}\,|\,\varphi_k(p)\leq 0\}$ (see Step 2 of Algorithm 1) to obtain the inequality $\langle\widehat{p}^k-\widetilde{p}^{k+1},\widetilde{p}^{k+1}-p\rangle\geq 0$. Note now that (13) is trivially equivalent to $p^{k+1}=(1-\beta_k)\widehat{p}^k+\beta_k\widetilde{p}^{k+1}$, which in turn combined with the property (8) yields

$$\|p^{k+1}-p\|^2=(1-\beta_k)\|\widehat{p}^k-p\|^2+\beta_k\|\widetilde{p}^{k+1}-p\|^2-\beta_k(1-\beta_k)\|\widehat{p}^k-\widetilde{p}^{k+1}\|^2$$

or, equivalently,

$$\beta_k\left(\|\widehat{p}^k-p\|^2-\|\widetilde{p}^{k+1}-p\|^2\right)=\|\widehat{p}^k-p\|^2-\beta_k(1-\beta_k)\|\widehat{p}^k-\widetilde{p}^{k+1}\|^2-\|p^{k+1}-p\|^2.\tag{20}$$

The desired inequality (18) now follows by multiplying the inequality in (19) by $\beta_k\geq 0$, by combining the resulting inequality with (20) and by using some simple algebraic manipulations.

Now, from (10) we have

$$p^k-p=\frac{1}{1+\alpha_k}(\widehat{p}^k-p)+\frac{\alpha_k}{1+\alpha_k}(p^{k-1}-p)\ \text{ and }\ \widehat{p}^k-p^{k-1}=(1+\alpha_k)(p^k-p^{k-1}).\tag{21}$$

Using (8) and the first identity in (21) we obtain

$$\|p^k-p\|^2=\frac{1}{1+\alpha_k}\|\widehat{p}^k-p\|^2+\frac{\alpha_k}{1+\alpha_k}\|p^{k-1}-p\|^2-\frac{\alpha_k}{(1+\alpha_k)^2}\|\widehat{p}^k-p^{k-1}\|^2,$$

which combined with the second identity in (21) and some algebraic manipulations gives

$$\|\widehat{p}^k-p\|^2=(1+\alpha_k)\|p^k-p\|^2-\alpha_k\|p^{k-1}-p\|^2+\alpha_k(1+\alpha_k)\|p^k-p^{k-1}\|^2.\tag{22}$$

Hence, (a) follows directly from (18), (22) and the definitions of h_k and s_{k+1} in (14) and (15), respectively.

(b) Note that (13) is also trivially equivalent to $\widehat{p}^k-\widetilde{p}^{k+1}=\beta_k^{-1}(\widehat{p}^k-p^{k+1})$, which in turn combined with the definition of s_{k+1} in (15) and the fact that $\beta_k\leq\overline{\beta}$ – see Step 2 of Algorithm 1 – yields

$$s_{k+1}=\beta_k(2-\beta_k)\|\widehat{p}^k-\widetilde{p}^{k+1}\|^2=\left(2\beta_k^{-1}-1\right)\|\widehat{p}^k-p^{k+1}\|^2\geq\left(2\overline{\beta}^{\,-1}-1\right)\|\widehat{p}^k-p^{k+1}\|^2.\tag{23}$$

Using (10), the Cauchy-Schwarz inequality, the Young inequality ($2ab\leq a^2+b^2$ with $a=\|p^{k+1}-p^k\|$ and $b=\|p^k-p^{k-1}\|$) and some algebraic manipulations, we find

$$\begin{aligned}\|\widehat{p}^k-p^{k+1}\|^2&=\|p^{k+1}-p^k\|^2+\alpha_k^2\|p^k-p^{k-1}\|^2-2\alpha_k\langle p^{k+1}-p^k,p^k-p^{k-1}\rangle\\&\geq\|p^{k+1}-p^k\|^2+\alpha_k^2\|p^k-p^{k-1}\|^2-2\alpha_k\|p^{k+1}-p^k\|\|p^k-p^{k-1}\|\\&\geq\|p^{k+1}-p^k\|^2+\alpha_k^2\|p^k-p^{k-1}\|^2-\alpha_k\left(\|p^{k+1}-p^k\|^2+\|p^k-p^{k-1}\|^2\right)\\&=(1-\alpha_k)\left(\|p^{k+1}-p^k\|^2-\alpha_k\|p^k-p^{k-1}\|^2\right).\end{aligned}\tag{24}$$

From (23) and (24) we obtain

$$s_{k+1} \geq \left(2\overline{\beta}^{-1} - 1\right)(1-\alpha_k)\left(\|p^{k+1} - p^k\|^2 - \alpha_k\|p^k - p^{k-1}\|^2\right),$$

which in turn combined with the inequality in (a) and (17), and after some simple manipulations, gives exactly the desired inequality in (b). □

Next is our first result on the (asymptotic) convergence of Algorithm 1. The key assumption is the summability condition (25), for which a sufficient condition, only depending on the parameters α_k and β_k, will be given in Theorem 2.3 – see conditions (26), (27) and Figure 2.

Theorem 2.2 (First result on the convergence of Algorithm 1)**.** *Let $\{p^k\}$, $\{\varphi_k\}$, $\{\widehat{p}^k\}$ and $\{\alpha_k\}$ be generated by* Algorithm 1 *and assume that*

$$\sum_{k=0}^{\infty} \alpha_k \|p^k - p^{k-1}\|^2 < \infty. \tag{25}$$

Then the following hold:

(a) *$\{p^k\}$ and $\{\widehat{p}^k\}$ are bounded sequences.*

(b) *If every weak cluster point of $\{p^k\}$ belongs to $\mathcal{S}$, then $\{p^k\}$ converges weakly to some element in $\mathcal{S}$.*

(c) *We have,*

$$\frac{\max\{0, \varphi_k(\widehat{p}^k)\}}{\|\nabla\varphi_k\|} \to 0.$$

Proof. Defining $\delta_k = \alpha_k(1+\alpha_k)\|p^k - p^{k-1}\|^2$ and using Lemma 2.1(a), we conclude that condition (80) in Lemma A.1 below holds with h_k and s_{k+1} as in (14) and (15), respectively. Hence, using the assumption (25), Lemma A.1(b) and (14), we conclude that

$$\lim_{k\to\infty} \|p^k - p\| \text{ exists for all } p \in \mathcal{S}.$$

This gives, in particular, that $\{p^k\}$ and $\{\widehat{p}^k\}$ are bounded (see (10)) and, after using Lemma A.2 below, that $\{p^k\}$ converges weakly to some element in $\mathcal{S}$ whenever every weak cluster point of $\{p^k\}$ belongs to $\mathcal{S}$. So we have proved (a) and (b).

To prove (c), note first that from (13) we have

$$\frac{\max\{0, \varphi_k(\widehat{p}^k)\}}{\|\nabla\varphi_k\|} = \|\widetilde{p}^{k+1} - \widehat{p}^k\|.$$

Hence, to conclude the proof of (c), it suffices to prove that $\|\widetilde{p}^{k+1} - \widehat{p}^k\| \to 0$. To this end, note that (25) combined with the definition of δ_k above, the fact that $\alpha_k^2 \leq \alpha_k$ and Lemma A.1(a) gives $\sum_{k=0}^{\infty} s_{k+1} < \infty$, with s_{k+1} (for all $k \geq 0$) as in (15), and so $s_{k+1} \to 0$. The desired result now follows form this fact, (15) and the fact that $0 < \underline{\beta} \leq \beta_k \leq \overline{\beta} < 2$ (see Step 2 of Algorithm 1). □

Theorem 2.3 (Second result on the convergence of Algorithm 1)**.** *Let* $\{p^k\}$ *and* $\{\alpha_k\}$ *be generated by* Algorithm 1*. Assume that* $\alpha \in [0,1)$, $\overline{\beta} \in (0,2)$ *and* $\{\alpha_k\}$ *satisfy the following (for some* $\overline{\alpha} > 0$*):*

$$0 \le \alpha_k \le \alpha_{k+1} \le \alpha < \overline{\alpha} < 1 \qquad \forall k \ge 0 \tag{26}$$

and

$$\overline{\beta} = \overline{\beta}(\overline{\alpha}) := \frac{2(\overline{\alpha}-1)^2}{2(\overline{\alpha}-1)^2 + 3\overline{\alpha} - 1}. \tag{27}$$

Then the following hold:

(a) *We have*

$$\sum_{k=0}^{\infty} \|p^k - p^{k-1}\|^2 < \infty. \tag{28}$$

(b) *Under the assumptions* (26) *and* (27)*, if every weak cluster point of* $\{p^k\}$ *belongs to* $\mathcal{S}$*, then* $\{p^k\}$ *converges weakly to some element in* $\mathcal{S}$*.*

Proof. (a) Define, for all $k \ge 0$,

$$\mu_k = h_k - \alpha_k h_{k-1} + \gamma_k \|p^k - p^{k-1}\|^2 \tag{29}$$

where h_k is as in (14) (for some $p \in \mathcal{S}$) and γ_k is as in (17). Using the assumption (26) and Lemma 2.1(b), we obtain, for all $k \ge 0$,

$$\begin{aligned}
\mu_{k+1} - \mu_k &\le h_{k+1} - \alpha_k h_k + \gamma_{k+1}\|p^{k+1} - p^k\|^2 - h_k + \alpha_k h_{k-1} - \gamma_k \|p^k - p^{k-1}\|^2 && \text{[by (26)]}\\
&= h_{k+1} - h_k - \alpha_k(h_k - h_{k-1}) + \gamma_{k+1}\|p^{k+1} - p^k\|^2 - \gamma_k\|p^k - p^{k-1}\|^2 \\
&\le \left[-\left(2-\overline{\beta}\right)\overline{\beta}^{\,-1}(1-\alpha_k) + \gamma_{k+1}\right]\|p^{k+1} - p^k\|^2 && \text{[by Lemma 2.1(b)]}\\
&\le \left[-\left(2-\overline{\beta}\right)\overline{\beta}^{\,-1}(1-\alpha_{k+1}) + \gamma_{k+1}\right]\|p^{k+1} - p^k\|^2 && \text{[by (26)]}\\
&= -q(\alpha_{k+1})\|p^{k+1} - p^k\|^2 && \text{[by (17) and (31)]}
\end{aligned} \tag{30}$$

where

$$q(\nu) := 2\left(\overline{\beta}^{\,-1} - 1\right)\nu^2 - \left(4\overline{\beta}^{\,-1} - 1\right)\nu + 2\overline{\beta}^{\,-1} - 1, \quad \nu \in \mathbb{R}. \tag{31}$$

Next we will show that $q(\alpha_{k+1})$ admits an uniform lower bound. To this end, note first that (27) and Lemma A.3 below yield

$$\overline{\alpha} = \frac{2(2-\overline{\beta})}{4 - \overline{\beta} + \sqrt{16\overline{\beta} - 7\overline{\beta}^2}},$$

which in turn combined with Lemma A.4 below implies that $q(\overline{\alpha}) = 0$ and $q(\cdot)$ is decreasing in $[0, \overline{\alpha}]$. Thus, in view of (26), we obtain

$$q(\alpha_{k+1}) \ge q(\alpha) > q(\overline{\alpha}) = 0$$

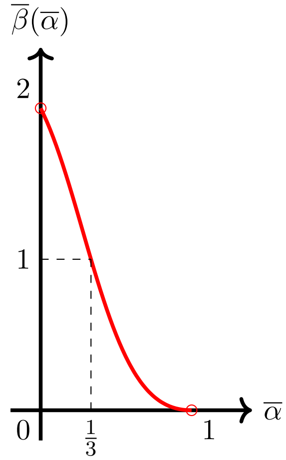


Figure 2: The relaxation parameter upper bound $\overline{\beta}(\overline{\alpha})$ from (27) as a function of inertial step upper bound $\overline{\alpha} > 0$ of (26). Note that $\overline{\beta}(1/3) = 1$, while $\overline{\beta}(\overline{\alpha}) > 1$ whenever $\overline{\alpha} < 1/3$.

and so, in view of (30), it follows that

$$\|p^{k+1} - p^k\|^2 \leq \frac{1}{q(\alpha)}(\mu_k - \mu_{k+1}) \qquad \forall k \geq 0. \tag{32}$$

Hence, for all $k \geq 0$,

$$\begin{aligned}\sum_{j=0}^{k} \|p^{j+1} - p^j\|^2 &\leq \frac{1}{q(\alpha)}(\mu_0 - \mu_{k+1}) \\ &\leq \frac{1}{q(\alpha)}(\mu_0 + \alpha h_k)\end{aligned} \tag{33}$$

where in the second inequality above we also used the fact that $\mu_{k+1} \geq -\alpha h_k$ (in view of (29) and (26)). Therefore, to finish the proof of (a) it is enough to find an upper bound on h_k and use (33). To this end, note that from (32) and (26) we have, for all $k \geq -1$,

$$\begin{aligned}\mu_0 \geq \mu_1 \geq \ldots \geq \mu_{k+1} &= h_{k+1} - \alpha_{k+1} h_k + \gamma_{k+1}\|p^{k+1} - p^k\|^2 \\ &\geq h_{k+1} - \alpha h_k\end{aligned}$$

and so, for all $k \geq -1$,

$$\begin{aligned}h_{k+1} &\leq \alpha^{k+1} h_0 + \left(\sum_{i=0}^{k} \alpha^i\right)\mu_0 \\ &\leq h_0 + \frac{\mu_0}{1-\alpha}\end{aligned}$$

where in the second inequality we also used the fact – from (29) – that $\mu_0 = (1-\alpha_0)h_0 \geq 0$. (b) The result follows trivially from (a), the fact that $\alpha_k \leq 1$ for all $k \geq 0$ and Theorem 2.2(b). □

Remarks.

(i) The proofs of Theorems 2.2 and 2.3 have followed the same outline of the proofs of Theorems 2.4 and 2.5 in [3]. On the other hand, we emphasize that Algorithm 1 proposed in this work is more general that Algorithm 1 from [3], since the latter has been designed to solve inclusions with monotone operators.

(ii) We deduce from conditions (26) and (27) that overrelaxation effects can be achieved in Algorithm 1 at the price of choosing the inertial parameter upper bound $\overline{\alpha}$ strictly smaller than 1/3 (see Figure 2). We also emphasize that the interplay between inertial and relaxation effects has also been investigated, e.g., in [6, 7, 8, 10].

3 A relative-error inertial-relaxed inexact projective splitting algorithm

In this section, we propose and study the asymptotic convergence of a relative-error inertial-relaxed inexact projective splitting algorithm (Algorithm 2). The main convergence results are stated in Theorems 3.5 and 3.6.

We start by considering the monotone inclusion problem (2) (or, equivalently, (4)), i.e., the problem of finding $z \in \mathcal{H}_0$ such that

$$0 \in \sum_{i=1}^{n} G_i^* T_i G_i(z) \tag{34}$$

where $n \geq 2$ and Assumptions (A1)–(A3) of Section 1 are assumed to hold.

Consider the extend solution set (or generalized Kuhn-Tucker set) as in (5) for the problem (34), i.e.:

$$\mathcal{S} := \left\{ (z, w_1, \ldots, w_{n-1}) \in \boldsymbol{\mathcal{H}} \;\middle|\; w_i \in T_i(G_i z),\, i = 1, \ldots, n-1,\, -\sum_{i=1}^{n-1} G_i^* w_i \in T_n(z) \right\}. \tag{35}$$

As we pointed out early, $z \in \mathcal{H}_0$ is a solution of (34) if and only if there exist $w_i \in \mathcal{H}_i$ $(i = 1, \ldots, n-1)$ such that $(z, w_1, \ldots, w_{n-1}) \in \mathcal{S}$. We deduce from Assumption (A3) above that $\mathcal{S}$ is nonempty. Moreover, it follows form [13, Lemma 3] that $\mathcal{S}$ is closed and convex in $\boldsymbol{\mathcal{H}}$ (endowed with inner product and norm as in (1)). As a consequence, one can apply the framework (Algorithm 1) of Section 2 for $\mathcal{S}$ as in (35) and the Hilbert space $\boldsymbol{\mathcal{H}}$ with the inner product and norm as in (1). The resulting scheme is Algorithm 2, which, in particular, will be shown in Proposition 3.2 to be a special instance of Algorithm 1.

Since Step 2 of Algorithm 1 demands the construction of an (nonconstant) affine function φ_k such that $\varphi_k(p) \leq 0$ for all $p \in \mathcal{S}$, next we discuss the construction of such φ_k satisfying the latter inequality for $\mathcal{S}$ as in (35).

Motivated by (6) and (7), for $y_i^k \in T_i(x_i^k)$ $(i = 1, \ldots, n)$, we define $\varphi_k : \boldsymbol{\mathcal{H}} \to \mathbb{R}$ by

$$\varphi_k(\underbrace{z, w_1, \ldots, w_{n-1}}_{p}) = \sum_{i=1}^{n-1} \langle G_i z - x_i^k, y_i^k - w_i \rangle + \langle G_n z - x_n^k, y_n^k + \sum_{i=1}^{n-1} G_i^* w_i \rangle. \tag{36}$$

We shall also use the fact, from (36) and (7), that

$$\varphi_k(p) = \sum_{i=1}^{n} \langle G_i z - x_i^k, y_i^k - w_i \rangle. \tag{37}$$

Note that the construction above depends on the computation of pairs (x_i^k, y_i^k) in the graph of T_i, for each $i = 1, \ldots, n$, which can be computed by inexact evaluation (with relative-error tolerance) of the resolvent $J_{T_i} = (T_i + I)^{-1}$ of T_i (see Step 2 of Algorithm 2).

Next lemma presents some properties of φ_k which will be useful in this paper.

Lemma 3.1. ([13, Lemma 4]) *Let $\varphi_k(\cdot)$ and $\mathcal{S}$ be as in* (36) *and* (35)*, respectively. The following hold:*

(a) *φ_k is affine on $\boldsymbol{\mathcal{H}}$.*

(b) *$\varphi_k(p) \leq 0$ for all $p \in \mathcal{S}$.*

(c) *The gradient of φ_k with respect to the inner product $\langle \cdot, \cdot \rangle_\gamma$ as in* (1) *is*

$$\nabla \varphi_k = \left(\frac{1}{\gamma} \left(\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k \right), x_1^k - G_1 x_n^k, \ldots, x_{n-1}^k - G_{n-1} x_n^k \right). \tag{38}$$

(d) *If $\nabla \varphi_k = 0$, then $(x_n^k, y_1^k, \ldots, y_{n-1}^k) \in \mathcal{S}$.*

As a direct consequence of Lemma 3.1(c) and (1), we have

$$\|\nabla \varphi_k\|_\gamma^2 = \gamma^{-1} \left\| \sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k \right\|^2 + \sum_{i=1}^{n-1} \|x_i^k - G_i x_n^k\|^2. \tag{39}$$

Next we present the main algorithm of this paper. As we mentioned before, it consists of a relative-error inertial-relaxed inexact projective splitting method for solving (34).

Algorithm 2. A relative-error inertial-relaxed inexact projective splitting algorithm for solving (34)

(0) Let $(z^{-1}, w_1^{-1}, \ldots, w_{n-1}^{-1}) = (z^0, w_1^0, \ldots, w_{n-1}^0) \in \boldsymbol{\mathcal{H}}$, $0 \leq \alpha, \sigma < 1$, $0 < \underline{\beta} \leq \overline{\beta} < 2$ and $\gamma > 0$ be given; let $k \leftarrow 0$.

(1) Choose $\alpha_k \in [0, \alpha]$ and let

$$\widehat{z}^k = z^k + \alpha_k (z^k - z^{k-1}), \tag{40}$$

$$\widehat{w}_i^k = w_i^k + \alpha_k (w_i^k - w_i^{k-1}), \quad i = 1, \ldots, n-1, \tag{41}$$

$$\widehat{w}_n^k = -\sum_{i=1}^{n-1} G_i^* \widehat{w}_i^k. \tag{42}$$

(2) Choose scalars $\rho_i^k > 0$ and compute (x_i^k, y_i^k) $(i = 1, \ldots, n)$ satisfying

$$\begin{cases} y_i^k \in T_i(x_i^k), \quad x_i^k + \rho_i^k y_i^k = G_i \widehat{z}^k + \rho_i^k \widehat{w}_i^k + e_i^k, \\[2ex] \|e_i^k\|^2 \leq \sigma^2 \left(\|G_i \widehat{z}^k - x_i^k\|^2 + \|\rho_i^k (\widehat{w}_i^k - y_i^k)\|^2 \right). \end{cases} \tag{43}$$

(3) If $x_i^k = G_i x_n^k$ $(i = 1, \ldots, n-1)$ and $\displaystyle\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k = 0$, then STOP. Otherwise, define

$$\varphi_k(\widehat{p}^k) = \sum_{i=1}^{n-1} \langle G_i \widehat{z}^k - x_i^k, y_i^k - \widehat{w}_i^k \rangle + \langle G_n \widehat{z}^k - x_n^k, y_n^k + \sum_{i=1}^{n-1} G_i^* \widehat{w}_i^k \rangle, \tag{44}$$

$$\theta_k = \frac{\max\{0, \varphi_k(\widehat{p}^k)\}}{\gamma^{-1} \|\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k\|^2 + \sum_{i=1}^{n-1} \|x_i^k - G_i x_n^k\|^2}. \tag{45}$$

(4) Choose some relaxation parameter $\beta_k \in [\underline{\beta}, \overline{\beta}]$ and define

$$z^{k+1} = \widehat{z}^k - \gamma^{-1} \beta_k \theta_k \left(\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k \right), \tag{46}$$

$$w_i^{k+1} = \widehat{w}_i^k - \beta_k \theta_k \left(x_i^k - G_i x_n^k \right), \quad i = 1, \ldots, n-1. \tag{47}$$

(5) Let $k \leftarrow k + 1$ and go to step 1.

Remarks.

(i) Similarly to Algorithm 1 of Section 2, Algorithm 2 also promotes inertial and relaxation effects, controlled by the parameters α_k and β_k, respectively. The inertial (extrapolation) step is performed in (40) and (41), while the relaxed projective step is given in (46) and (47) (in the context of Algorithm 1, see Figure 1 of Section 2). Conditions on the choice of the upper bounds α and $\overline{\beta}$, as well as on the sequence of extrapolation parameters $\{\alpha_k\}$, to guarantee the convergence of Algorithm 2 will be given in Theorem 3.6.

(ii) Direct substitution of (41) into (42) gives that, similarly to $\widehat{w}_i^k$ for $i = 1, \ldots, n-1$, $\widehat{w}_n^k$ also satisfies

$$\widehat{w}_n^k = w_n^k + \alpha_k (w_n^k - w_n^{k-1}), \tag{48}$$

where

$$w_n^k := -\sum_{i=1}^{n-1} G_i^* w_i^k, \qquad \forall k \geq 0. \tag{49}$$

(iii) The computation of (x_i^k, y_i^k) in (43) can be performed inexactly within a relative-error tolerance controlled by the parameter $\sigma \in [0, 1)$. In practice, the error condition in (43) is used as a stopping-criterion for some computational procedure (e.g., conjugate gradient algorithm) applied to (inexactly) solving the related inclusion (for $i = 1, \ldots, n$)

$$0 \in \rho_i^k T_i(x) + x - (G_i \widehat{z}^k + \rho_i^k \widehat{w}_i^k)$$

until the error-condition in (43) is satisfied for the first time. Note also that (x_i^k, y_i^k) is given explicitly by $x_i^k = J_{\rho_i^k T_i}(G_i \widehat{z}^k + \rho_i^k \widehat{w}_i^k)$ and $y_i^k = \frac{G_i \widehat{z}^k - x_i^k}{\rho_i^k} + \widehat{w}_i^k$ whenever the resolvent $J_{\rho_i^k T_i} = (\rho_i^k T_i + I)^{-1}$ of T_i is assumed to be easily computed and $\sigma = 0$ in (43). For the particular case of the minimization problem (3), the computation of (x_i^k, y_i^k) reduces to the (inexact) computation of the proximity operator $\operatorname{prox}_{\rho_i^k f_i}$, i.e., in this case

$$x_i^k \approx \arg\min_{z \in \mathcal{H}_0} \left\{ f_i(z) + \frac{1}{2\rho_i^k} \| z - (G_i \widehat{z}^k + \rho_i^k \widehat{w}_i^k) \|^2 \right\}. \tag{50}$$

See also Section 4 for an additional discussion in the context of LASSO problems.

(iv) It follows from Lemma 3.1, items (c) and (d), that $(x_n^k, y_1^k, \ldots, y_{n-1}^k)$ belongs to the extended solution set $\mathcal{S}$ whenever Algorithm 2 stops at Step 3. In particular, in this case, x_n^k is a solution of (34).

From now one in this paper, we assume that Algorithm 2 *generates infinite sequences, i.e., we assume that it never stops at* Step 3.

(v) We also emphasize that if $\alpha_k \equiv 0$ in Algorithm 2, then it reduces to the projective splitting algorithm (or some of its variants) originated in [12] and later developed in different directions in, e.g., [1, 13, 14, 15]. The advantages and flexibility of projective splitting algorithms (beyond inertial effects) when compared to other proximal-splitting strategies are also extensively discussed in the latter references.

Next we show that Algorithm 2 (under the assumption that it never stops at Step 3; see Remark (iv) above) is a special instance of Algorithm 1 for finding a point in $\mathcal{S}$ as in (35) in the Hilbert space $\boldsymbol{\mathcal{H}}$ endowed with the inner product and norm as in (1).

Proposition 3.2. *Assume that* Algorithm 2 *does not stop at* Step 3*, let* $\{z^k\}$*,* $\{w_1^k\}, \dots, \{w_{n-1}^k\}$ *be generated by* Algorithm 2*, let* $\{\varphi_k\}$ *be as in* (36) *and define*

$$p^k = (z^k, w_1^k, \dots, w_{n-1}^k) \qquad \forall k \geq -1. \tag{51}$$

Then the following hold:

(a) *For all* $k \geq 0$*,*

$$\nabla\varphi_k \neq 0 \ \ \textit{and} \ \ \varphi_k(p) \leq 0 \qquad \forall p \in \mathcal{S},$$

where $\mathcal{S}$ *is as in* (35)*.*

(b) *For all* $k \geq 0$*,*

$$p^{k+1} = \widehat{p}^k - \frac{\beta_k \max\{0, \varphi_k(\widehat{p}^k)\}}{\|\nabla\varphi_k\|_\gamma^2}\nabla\varphi_k \ \text{ and } \ \widehat{p}^k = (\widehat{z}^k, \widehat{w}_i^k, \dots, \widehat{w}_{n-1}^k), \tag{52}$$

where $\widehat{p}^k$ *is as in* (10)*.*

As a consequence of (a) *and* (b) *above, it follows that* Algorithm 2 *is a special instance of* Algorithm 1 *for finding a point in the extended solution set* $\mathcal{S}$ *as in* (35)*.*

Proof. (a) The fact that $\nabla\varphi_k \neq 0$ follows from the assumption that Algorithm 2 does not stop at Step 3 and Lemma 3.1(c). Using now Lemma 3.1(b) and the inclusions in (43), we conclude that $\varphi_k(p) \leq 0$ for all $p \in \mathcal{S}$.

(b) The second identity in (52) follows from (10), (51), (40) and (41). On the other hand, the first identity in (52) is a direct consequence of (45)–(47), (38), (39) and the second identity in (52).

Finally, the last statement of the proposition is a consequence of items (a) and (b) as well as of Algorithm 1's definition. □

Since Algorithm 2 is a special instance of Algorithm 1 of Section 2, it follows from Theorems 2.2(b) and 2.3(b), under the assumptions (25) and (26)–(27), respectively, that to prove the convergence of Algorithm 2 it suffices to check that every weak cluster point of Algorithm 2 belongs to $\mathcal{S}$ as in (35). This will be done in Proposition 3.4(e), but before we need the lemma below.

Lemma 3.3. *Consider the sequences evolved by* Algorithm 2*, let* $\widehat{p}^k = (\widehat{z}^k, \widehat{w}_i^k, \dots, \widehat{w}_{n-1}^k)$ *and let* $\widehat{w}_n^k$ *be as in* (42)*. Assume that, for* $i = 1, \dots, n$*,*

$$0 < \underline{\rho} \leq \rho_i^k \leq \overline{\rho} < \infty \qquad \forall k \geq 0. \tag{53}$$

Then the following hold:

(a) *For all* $k \geq 0$*,*

$$\varphi_k(\widehat{p}^k) \geq \frac{(1-\sigma^2)\min\left\{\overline{\rho}^{\,-1}, \overline{\rho}\right\}}{2}\sum_{i=1}^n \left(\|G_i\widehat{z}^k - x_i^k\|^2 + \|\widehat{w}_i^k - y_i^k\|^2\right) \geq 0. \tag{54}$$

(b) *There exists a constant* $c > 0$ *such that, for all* $k \geq 0$*,*

$$\frac{\varphi_k(\widehat{p}^k)^2}{c\|\nabla\varphi_k\|_\gamma^2} \geq \varphi_k(\widehat{p}^k) \geq c\,\|\nabla\varphi_k\|_\gamma^2. \tag{55}$$

Proof. (a) Using the identity $\langle a, b\rangle = (1/2)\left(\|a+b\|^2 - \|a\|^2 - \|b\|^2\right)$ with $a = x_i^k - G_i\widehat{z}^k$ and $b = \rho_i^k(y_i^k - \widehat{w}_i^k)$, and some algebraic manipulations, we obtain, for $i = 1, \ldots, n$,

$$\begin{aligned}\langle x_i^k - G_i\widehat{z}^k, \rho_i^k(y_i^k - \widehat{w}_i^k)\rangle &= \frac{1}{2}\left(\|\underbrace{x_i^k - G_i\widehat{z}^k + \rho_i^k(y_i^k - \widehat{w}_i^k)}_{=e_i^k}\|^2 - \|G_i\widehat{z}^k - x_i^k\|^2 - \|\rho_i^k(\widehat{w}_i^k - y_i^k)\|^2\right)\\ &\leq \frac{-(1-\sigma^2)}{2}\left(\|G_i\widehat{z}^k - x_i^k\|^2 + \|\rho_i^k(\widehat{w}_i^k - y_i^k)\|^2\right),\end{aligned}$$

where we also used the error condition in (43). Note now that the desired result follows by dividing the latter inequality by $-\rho_i^k$ and by using (37) and assumption (53).

(b) First note that using the property (9), (42) and the assumption that $G_n = I$, we obtain

$$\left\|\sum_{i=1}^n G_i^* y_i^k\right\|^2 = \left\|\sum_{i=1}^n G_i^*(\widehat{w}_i^k - y_i^k)\right\|^2 \leq n\left(\max_{i=1,\ldots,n}\|G_i^*\|^2\right)\sum_{i=1}^n \|\widehat{w}_i^k - y_i^k\|^2. \tag{56}$$

On the other hand, using the inequality $\|a+b\|^2 \leq 2(\|a\|^2 + \|b\|^2)$, (again) the fact that $G_n = I$ and some algebraic manipulations, we find

$$\begin{aligned}\sum_{i=1}^{n-1}\|x_i^k - G_i x_n^k\|^2 &= \sum_{i=1}^{n-1}\|x_i^k - G_i\widehat{z}^k + G_i(\widehat{z}^k - x_n^k)\|^2\\ &\leq 2\sum_{i=1}^{n-1}\left(\|G_i\widehat{z}^k - x_i^k\|^2 + \|G_i(\widehat{z}^k - x_n^k)\|^2\right)\\ &\leq 2\left(\sum_{i=1}^{n-1}\|G_i\widehat{z}^k - x_i^k\|^2 + (n-1)\max_{i=1,\ldots,n-1}\{\|G_i\|^2\}\|\widehat{z}^k - x_n^k\|^2\right)\\ &\leq 2\max\left\{1, (n-1)\max_{i=1,\ldots,n-1}\{\|G_i\|^2\}\right\}\sum_{i=1}^n\|G_i\widehat{z}^k - x_i^k\|^2.\end{aligned} \tag{57}$$

We know from (39) (and the fact that $G_n = I$) that

$$\|\nabla\varphi_k\|_\gamma^2 = \gamma^{-1}\left\|\sum_{i=1}^n G_i^* y_i^k\right\|^2 + \sum_{i=1}^{n-1}\|x_i^k - G_i x_n^k\|^2,$$

which combined with (56), (57) and (54) yields the second inequality in (55), for some constant $c > 0$. To finish the proof, note that the first inequality in (55) is a direct consequence of the second one. □

Proposition 3.4. *Consider the sequences evolved by* Algorithm 2 *and let* $\{w_n^k\}$ *and* $\{p^k\}$ *be as in* (49) *and* (51)*, respectively. Assume that*

$$\sum_{k=0}^\infty \alpha_k\|p^k - p^{k-1}\|_\gamma^2 < \infty \tag{58}$$

and, for $i = 1, \ldots, n$,

$$0 < \underline{\rho} \le \rho_i^k \le \overline{\rho} < \infty \qquad \forall k \ge 0. \tag{59}$$

Then,

(a) *We have,* $\varphi_k(\widehat{p}^k) \to 0$ *and* $\|\nabla \varphi_k\|_\gamma \to 0$.

(b) *We have,* $\sum_{i=1}^n G_i^* y_i^k \to 0$ *and* $x_i^k - G_i x_n^k \to 0$ $(i = 1, \ldots, n-1)$.

(c) *For each* $i = 1, \ldots, n$*, we have* $\|G_i \widehat{z}^k - x_i^k\| \to 0$ *and* $\|\widehat{w}_i^k - y_i^k\| \to 0$.

(d) *For each* $i = 1, \ldots, n$*, we have* $\|G_i z^k - x_i^k\| \to 0$ *and* $\|w_i^k - y_i^k\| \to 0$.

(e) *Every weak cluster point of* $\{p^k\}$ *belongs to* $\mathcal{S}$*, where* $\mathcal{S}$ *is as in* (35).

Proof. (a) Using the last statement in Proposition 3.2, Theorem 2.2(c) and the fact from (54) that $\varphi_k(\widehat{p}^k) \ge 0$, we obtain

$$\frac{\varphi_k(\widehat{p}^k)}{\|\nabla \varphi_k\|_\gamma} \to 0,$$

which after taking limit in (55) gives the desired result in item (a).

(b) This follows from the second limit in item (a) combined with (39) (and the fact that $G_n = I$).

(c) This follows from the first limit in item (a) and (54).

(d) Using the triangle inequality, the identity (40), (51) and (1), we find

$$\begin{aligned} \|G_i z^k - x_i^k\| &\le \|z^k - \widehat{z}^k\| \|G_i\| + \|G_i \widehat{z}^k - x_i^k\| \\ &= \alpha_k \|z^k - z^{k-1}\| \|G_i\| + \|G_i \widehat{z}^k - x_i^k\| \\ &\le \sqrt{\gamma^{-1}} \sqrt{\alpha_k}\, \|p^k - p^{k-1}\|_\gamma \|G_i\| + \|G_i \widehat{z}^k - x_i^k\|, \quad i = 1, \ldots, n, \end{aligned} \tag{60}$$

where we also used the fact that $\alpha_k \le \sqrt{\alpha_k}$ (because $0 \le \alpha_k < 1$). Using a similar reasoning, we also find

$$\|y_i^k - w_i^k\| \le \sqrt{\alpha_k}\, \|p^k - p^{k-1}\|_\gamma + \|y_i^k - \widehat{w}_i^k\|, \qquad i = 1, \ldots, n-1. \tag{61}$$

Note also that, using (42), (41), (49), the property (9), the fact that $\alpha_k^2 \le \alpha_k$ and (1), we obtain

$$\begin{aligned} \frac{1}{2}\|y_n^k - w_n^k\|^2 &\le \|\widehat{w}_n^k - w_n^k\|^2 + \|y_n^k - \widehat{w}_n^k\|^2 \\ &\le (n-1) \max_{i=1,\ldots,n-1} \{\|G_i^*\|^2\} \left(\sum_{i=1}^{n-1} \|\widehat{w}_i^k - w_i^k\|^2 \right) + \|y_n^k - \widehat{w}_n^k\|^2 \\ &= (n-1) \max_{i=1,\ldots,n-1} \{\|G_i^*\|^2\} \left(\sum_{i=1}^{n-1} \alpha_k^2 \|w_i^{k-1} - w_i^k\|^2 \right) + \|y_n^k - \widehat{w}_n^k\|^2 \\ &\le (n-1) \max_{i=1,\ldots,n-1} \{\|G_i^*\|^2\} \left(\sum_{i=1}^{n-1} \alpha_k \|w_i^{k-1} - w_i^k\|^2 \right) + \|y_n^k - \widehat{w}_n^k\|^2 \\ &\le (n-1) \max_{i=1,\ldots,n-1} \{\|G_i^*\|^2\} \alpha_k \|p^k - p^{k-1}\|_\gamma^2 + \|y_n^k - \widehat{w}_n^k\|^2. \end{aligned} \tag{62}$$

To finish the proof of (d), combine (60)–(62) with item (c) and assumption (58) (which, in particular, implies that $\alpha_k\|p^k - p^{k-1}\|_\gamma^2 \to 0$).

(e) Let $p^\infty := (z^\infty, w_1^\infty, \cdots, w_{n-1}^\infty) \in \boldsymbol{\mathcal{H}}$ be a weak cluster point of $\{p^k\}$ (by Proposition 3.2 and Theorem 2.2(a), we have that $\{p^k\}$ is bounded) and let $\{p^{k_j}\}$ be a subsequence of $\{p^k\}$ such that $p^{k_j} \rightharpoonup p^\infty$, i.e.,

$$z^{k_j} \rightharpoonup z^\infty \ \text{ and } \ w_i^{k_j} \rightharpoonup w_i^\infty, \quad i = 1, \ldots, n-1. \tag{63}$$

Using item (d), (63) and the fact that $G_n = I$ (see Assumption (A2)), we obtain

$$x_n^{k_j} \rightharpoonup z^\infty \ \text{ and } \ y_i^{k_j} \rightharpoonup w_i^\infty, \quad i = 1, \ldots, n-1. \tag{64}$$

Now define the maximal monotone operators $A : \mathcal{H}_0 \rightrightarrows \mathcal{H}_0$, $B : \mathcal{H}_1 \times \cdots \times \mathcal{H}_{n-1} \rightrightarrows \mathcal{H}_1 \times \cdots \times \mathcal{H}_{n-1}$ and the bounded linear operator $G : \mathcal{H}_0 \to \mathcal{H}_1 \times \cdots \times \mathcal{H}_{n-1}$ by

$$A := T_n, \quad B := T_1 \times \cdots \times T_{n-1} \ \text{ and } \ G := (G_1, \ldots, G_{n-1}). \tag{65}$$

Using the above definitions of A and B and the inclusions in (43), we have

$$a^{k_j} \in A(r^{k_j}) \ \text{ and } \ b^{k_j} \in B(s^{k_j}), \tag{66}$$

where

$$a^{k_j} := y_n^{k_j}, \quad r^{k_j} := x_n^{k_j}, \quad b^{k_j} := (y_1^{k_j}, \ldots, y_{n-1}^{k_j}) \ \text{ and } \ s^{k_j} := (x_1^{k_j}, \ldots, x_{n-1}^{k_j}). \tag{67}$$

Moreover, (67) and (64) yield

$$r^{k_j} \rightharpoonup r^\infty \ \text{ and } \ b^{k_j} \rightharpoonup b^\infty, \tag{68}$$

where

$$r^\infty := z^\infty \ \text{ and } \ b^\infty := (w_1^\infty, \cdots, w_{n-1}^\infty). \tag{69}$$

Note now that using (67), the fact that $G^*(w_1, \ldots, w_{n-1}) = \sum_{i=1}^{n-1} G_i^* w_i$, for all $(w_1, \cdots, w_{n-1}) \in \mathcal{H}_1 \times \cdots \times \mathcal{H}_{n-1}$, the fact that $G_n = I$ and the first limit in item (b), we find

$$a^{k_j} + G^* b^{k_j} = \sum_{i=1}^{n} G_i^* y_i^{k_j} \to 0. \tag{70}$$

Using now the second limit in item (b) combined with (67) and the definition of G in (65), we obtain

$$G r^{k_j} - s^{k_j} \to 0. \tag{71}$$

Using Lemma A.5 below combined with (66), (68), (70) and (71) we conclude that

$$b^\infty \in B(G r^\infty) \ \text{ and } \ -G^* b^\infty \in A(r^\infty),$$

which, in turn, combined with (65) and (69) implies that

$$w_i^\infty \in T_i(G_i z^\infty), \ \ i = 1, \cdots, n-1, \ \ -\sum_{i=1}^{n-1} G_i^* w_i^\infty \in T_n(z^\infty),$$

which means exactly (see (35)) that $p^\infty = (z^\infty, w_1^\infty, \cdots, w_{n-1}^\infty) \in \mathcal{S}$. Hence, we conclude that every weak cluster point of $\{p^k\}$ belongs to $\mathcal{S}$. □

Next is the first result on the asymptotic convergence of Algorithm 2.

Theorem 3.5 (First result on the convergence of Algorithm 2)**.** *Consider the sequences evolved by* Algorithm 2 *and let* $\{p^k\}$ *be as in* (51)*. Assume that conditions* (58) *and* (59) *of* Proposition 3.4 *hold, i.e., assume that*

$$\sum_{k=0}^{\infty} \alpha_k \|p^k - p^{k-1}\|_\gamma^2 < \infty \tag{72}$$

and, for $i = 1, \ldots, n$,

$$0 < \underline{\rho} \le \rho_i^k \le \overline{\rho} < \infty \qquad \forall k \ge 0. \tag{73}$$

Then, there exists $(z^\infty, w_1^\infty, \cdots, w_{n-1}^\infty) \in \mathcal{S}$ *such that* $z^k \rightharpoonup z^\infty$ *and* $w_i^k \rightharpoonup w_i^\infty$*, for* $i = 1, \ldots, n-1$*. Furthermore,* $x_i^k \rightharpoonup G_i z^\infty$ *and* $y_i^k \rightharpoonup w_i^\infty$*, for* $i = 1, \ldots, n$*, where* w_n^k *is as in* (49)*.*

Proof. In view of Propositions 3.2 and 3.4(e) and Theorem 2.2(b) one concludes that that $\{p^k\}$ converges weakly to some $p^\infty := (z^\infty, w_1^\infty, \cdots, w_{n-1}^\infty)$ in $\mathcal{S}$ as in (35). Using the definition of p^k in (51) one easily concludes that $z^k \rightharpoonup z^\infty$ and $w_i^k \rightharpoonup w_i^\infty$, for $i = 1, \ldots, n-1$, which in turn combined with Proposition 3.4(d) implies that $x_i^k \rightharpoonup G_i z^\infty$ and $y_i^k \rightharpoonup w_i^\infty$, for $i = 1, \ldots, n$. □

Next theorem shows the convergence of Algorithm 2 under certain assumptions on α, $\overline{\beta}$ and the sequence $\{\alpha_k\}$ (see the remarks below).

Theorem 3.6 (Second result on the convergence of Algorithm 2)**.** *Consider the sequences evolved by* Algorithm 2 *and assume that* $\alpha \in [0, 1)$, $\overline{\beta} \in (0, 2)$ *and* $\{\alpha_k\}$ *satisfy* (*for some* $\overline{\alpha} > 0$) *the conditions* (26) *and* (27) *of* Theorem 2.3*, i.e.,*

$$0 \le \alpha_k \le \alpha_{k+1} \le \alpha < \overline{\alpha} < 1 \qquad \forall k \ge 0 \tag{74}$$

and

$$\overline{\beta} = \overline{\beta}(\overline{\alpha}) := \frac{2(\overline{\alpha} - 1)^2}{2(\overline{\alpha} - 1)^2 + 3\overline{\alpha} - 1}. \tag{75}$$

Assume also that condition (73) *holds, i.e., assume that, for* $i = 1, \ldots, n$,

$$0 < \underline{\rho} \le \rho_i^k \le \overline{\rho} < \infty \qquad \forall k \ge 0. \tag{76}$$

Then, the same conclusions of Theorem 3.5 *hold, i.e., there exists* $(z^\infty, w_1^\infty, \cdots, w_{n-1}^\infty) \in \mathcal{S}$ *such that* $z^k \rightharpoonup z^\infty$ *and* $w_i^k \rightharpoonup w_i^\infty$*, for* $i = 1, \ldots, n-1$*. Furthermore,* $x_i^k \rightharpoonup G_i z^\infty$ *and* $y_i^k \rightharpoonup w_i^\infty$*, for* $i = 1, \ldots, n$*, where* w_n^k *is as in* (49)*.*

Proof. In view of Propositions 3.2 and 3.4(e) and Theorem 2.3(b) one concludes that that $\{p^k\}$ converges weakly to some $p^\infty := (z^\infty, w_1^\infty, \cdots, w_{n-1}^\infty)$ in $\mathcal{S}$ as in (35). The rest of the proof follows the same argument used in Theorem 3.5's proof. □

Remarks.

(i) We emphasize that the conditions on α, $\overline{\beta}$ and on the sequence $\{\alpha_k\}$ above are exactly the same of Theorem 2.3, namely (26) and (27). See also the second remark following Theorem 2.3 and Figure 2 for a discussion of the interplay between inertial and overrelaxation parameters.

(ii) Note that, since $(z^\infty, w_1^\infty, \cdots, w_{n-1}^\infty) \in \mathcal{S}$ in Theorem 3.6, it follows that the weak limit $z^\infty \in \mathcal{H}_0$ is a solution of the monotone inclusion problem (34).

4 Numerical experiments on LASSO Problems

In this section, we present simple numerical experiments on ℓ_1–regularized least square problems

$$\min_{x\in\mathbb{R}^d}\left\{\frac{1}{2}\|Qx-b\|_2^2+\lambda\|x\|_1\right\}, \tag{77}$$

where $Q\in\mathbb{R}^{m\times d}$, $b\in\mathbb{R}^m$ and $\lambda\geq 0$. Let $\mathcal{R}=\{R_1,\dots,R_r\}$ be an arbitrary partition [1] of $\{1,\dots,m\}$ and, for $i=1,\dots,r$, let $Q_i\in\mathbb{R}^{|R_i|\times d}$ be the submatrix of Q with rows corresponding to indices in R_i and similarly let $b_i\in\mathbb{R}^{|R_i|}$ be the corresponding subvector of b. Then, problem (77) is equivalent to the minimization problem

$$\min_{x\in\mathbb{R}^d}\left\{\sum_{i=1}^r\frac{1}{2}\|Q_ix-b_i\|_2^2+\lambda\|x\|_1\right\},$$

which, in turn, is clearly equivalent to the monotone inclusion problem

$$0\in\sum_{i=1}^r Q_i^T(Q_ix-b_i)+\partial(\lambda\|x\|_1). \tag{78}$$

On the other hand, (78) is a special instance of the monotone inclusion problem (34) with $z\leftarrow x$, $n=r+1$, $G_i=I$ $(i=1,\dots,n)$,

$$T_i(x)=Q_i^T(Q_ix-b_i)\ (i=1,\dots,n-1)\ \text{ and }\ T_n(x)=\partial(\lambda\|x\|_1).$$

In this section, we shall apply Algorithm 2 for solving (78) (and, in particular, (77)) with the following choice of parameters (see Steps 0, 1, 2 and 4 of Algorithm 2):

$$\alpha_k\equiv\alpha=0.1,\ \ \sigma=0.99,\ \ \gamma=1,\ \ \rho_i^k\equiv 1\ \text{ and }\ \ \beta_k\equiv\underline{\beta}=\overline{\beta}=1.5519.$$

The value 1.5519 is computed from (75) with $\overline{\alpha}=0.17>\alpha$. Following [13], we stop the algorithm using the stopping criterion

$$\frac{|F(z^k)-F^*|}{F^*}\leq 10^{-4}, \tag{79}$$

where $F(\cdot)$ denotes the objective function in (77) and F^* is the optimal value of the problem estimated by running Algorithm 2 at least 10^4 iterations and taking the minimum objective value.

At each iteration k of Algorithm 2, we used two different strategies for computing (x_i^k,y_i^k) $(i=1,\dots,n)$ satisfying (43): for $i=1,\dots,n-1$, in which case $T_i(x)=Q_i^T(Q_ix-b_i)$, we implemented the standard conjugate gradient (CG) algorithm for computing $x=x_i^k$ as an approximate solution of the linear system

$$(Q_i^TQ_i+I)x=\widehat{z}^k+\widehat{w}_i^k+Q_i^Tb_i$$

until the satisfaction of the relative-error condition in (43) with $y_i^k:=T_i(x_i^k)$ by the residual $e_i^k:=T_i(x_i^k)+x_i^k-(\widehat{z}^k+\widehat{w}_i^k)$. On the other hand, for $i=n$, in which case $T_n(x)=\partial(\lambda\|x\|_1)$, we set $x_i^k=\operatorname{prox}_{\lambda\|\cdot\|_1}(\widehat{z}^k+\widehat{w}_i^k)$ and $y_i^k=(\widehat{z}^k+\widehat{w}_i^k)-x_i^k$ (in this case, $e_i^k=0$).

Data sets. We implemented Algorithm 2 using the following data sets:

- Four randomly generated instances of (77): RandomA, RandomB, RandomC and RandomD. We used the Matlab command "randn" to generate $Q\in\mathbb{R}^{m\times d}$, and $b\in\mathbb{R}^m$ with $b_j\in\{0,1\}$ $(j=1,\dots,m)$ where $b=(b_1,\dots,b_j,\dots,b_m)$ (see Table 1).

[1] $R_i\neq\emptyset$ $(i=1,\dots,r)$, $R_i\cap R_j=\emptyset$ for $i\neq j$ and $\cup_{i=1}^r R_i=\{1,\dots,m\}$.

Table 1: Dimensions of $Q \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$, size r of the partition $\mathcal{R}$ of $\{1, \dots, m\}$ and number of rows of each submatrix of Q on four randomly generated instances of (77)

	m	d	r	$\mid R_i \mid$
RandomA	1000	1000	10	100 $(i = 1, \dots, 10)$
RandomB	5000	100	20	250 $(i = 1, \dots, 20)$
RandomC	50000	100	250	200 $(i = 1, \dots, 250)$
RandomD	100000	100	325	307 $(i = 1, \dots, 324)$ 532 $(i = 325)$

- Five data sets (real examples) from the UCI Machine Learning Repository [11]: the blog feedback dataset (BlogFeedback) [2] , communities and crime dataset (Crime) [3] , DrivFace dataset (DrivFace) [4] , Single-Pixed Camera (Mug32) [5] and Breast Cancer Wisconsin (Diagnostic) dataset (Wisconsin) [6] (see Table 2).

We also used $\lambda = 0.1\|Q^T b\|_\infty$ (see [9]) in (77). Table 3 shows the number of outer iterations, and Table 4 shows runtimes in seconds. Figures 3 and 4 show the same results graphically (see the stopping criterion (79)).

A Auxiliary results

The following lemma was essentially proved by Alvarez and Attouch in [2, Theorem 2.1] (see also [4, Lemma A.4]).

Lemma A.1. *Let the sequences* $\{h_k\}$, $\{s_k\}$, $\{\alpha_k\}$ *and* $\{\delta_k\}$ *in* $[0, +\infty)$ *and* $\alpha \in \mathbb{R}$ *be such that* $h_0 = h_{-1}$, $0 \le \alpha_k \le \alpha < 1$ *and*

$$h_{k+1} - h_k + s_{k+1} \le \alpha_k(h_k - h_{k-1}) + \delta_k \qquad \forall k \ge 0. \tag{80}$$

The following hold:

(a) *For all* $k \ge 1$,

$$h_k + \sum_{j=1}^{k} s_j \le h_0 + \frac{1}{1-\alpha} \sum_{j=0}^{k-1} \delta_j. \tag{81}$$

[2] https://archive.ics.uci.edu/ml/datasets/BlogFeedback.
[3] http://archive.ics.uci.edu/ml/datasets/communities+and+crime.
[4] https://archive.ics.uci.edu/ml/datasets/DrivFace.
[5] see [3].
[6] https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic).

Table 2: Dimensions of $Q \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$, size r of the partition $\mathcal{R}$ of $\{1, \dots, m\}$ and number of rows of each submatrix of Q on five real examples (from the UCI Machine Learning Repository [11]) of (77)

	m	d	r	$\mid R_i \mid$
BlogFeedback	60021	280	175	343 $(i = 1, \dots, 174)$
				339 $(i = 175)$
Crime	1994	121	10	200 $(i = 1, \dots, 9)$
				194 $(i = 10)$
DrivFace	606	6400	6	101 $(i = 1, \dots, 6)$
Mug32	410	1024	4	100 $(i = 1, 2, 3)$
				110 $(i = 4)$
Wisconsin	198	30	3	66 $(i = 1, 2, 3)$

Table 3: Outer iterations for LASSO problems

Problem	PS $_{iteration1}$	PS_in_rel $_{iteration2}$	$\frac{iteration2}{iteration1}$
BlogFeedback	2968	**2342**	0.7891
Crime	**211**	216	1.0237
DrivFace	2008	**585**	0.2913
Mug32	203	**192**	0.9458
Wisconsin	211	**210**	0.9952
RandomA	219	**185**	0.8447
RandomB	23	**21**	0.9131
RandomC	408	**151**	0.3701
RandomD	507	**278**	0.5483
Geometric mean	337.04	**231.98**	0.6883

Figure 3: Comparison of performance in LASSO problems

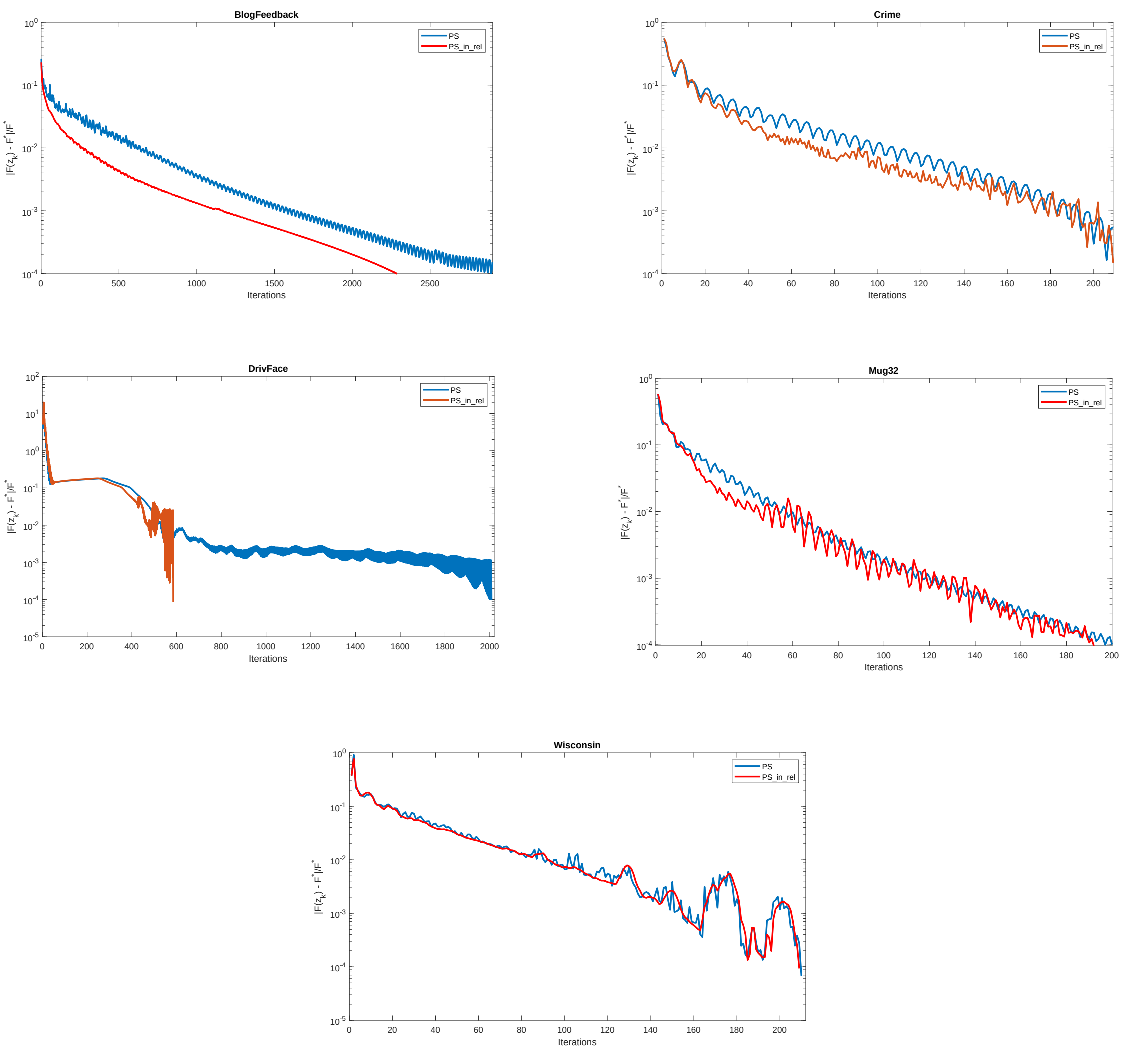
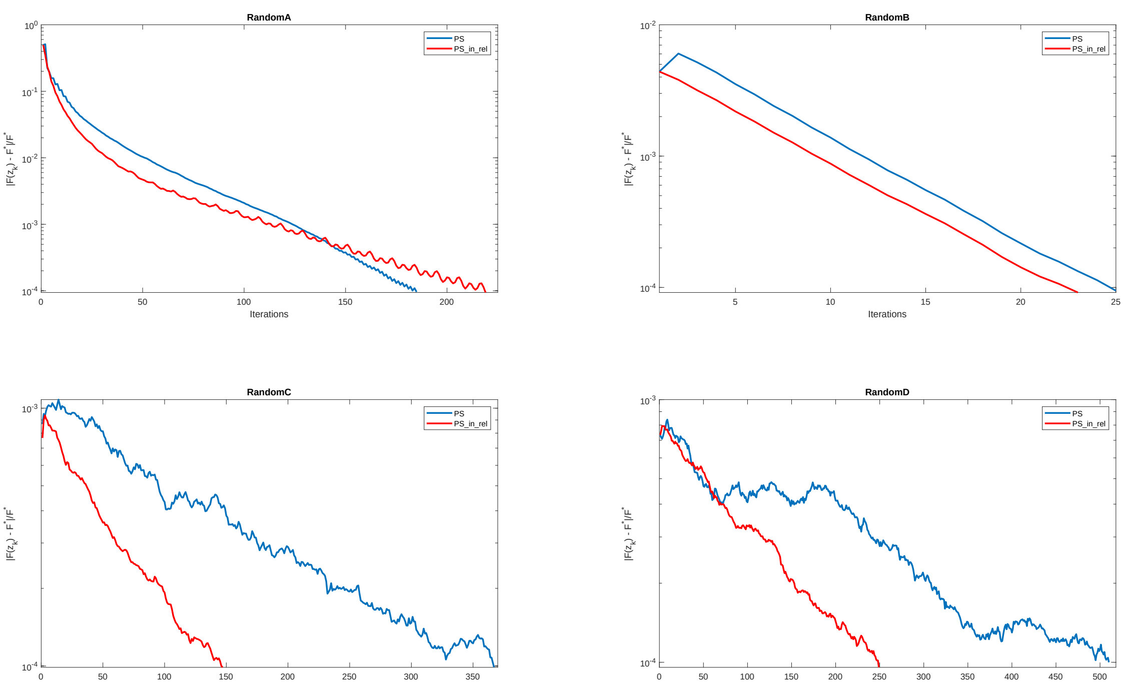


Table 4: LASSO runtimes in seconds

Problem	PS $time1$	PS_in_relerr $time2$	$\frac{time2}{time1}$
BlogFeedback	207.18	**130.44**	0.6296
Crime	0.85	**0.78**	0.9176
DrivFace	133.19	**37.11**	0.2786
Mug32	1.36	**1.18**	0.8676
Wisconsin	0.15	**0.11**	0.7333
randomA	2.45	**1.69**	0.6898
randomB	0.25	**0.13**	0.52
randomC	10.53	**4.08**	0.3875
randomD	20.09	**11.31**	0.5629
Geometric mean	3.79	**2.57**	0.6793

Figure 4: Comparison of performance in LASSO problems



(b) *If* $\sum_{k=0}^{\infty} \delta_k < +\infty$, *then* $\lim_{k\to\infty} h_k$ *exist, i.e., the sequence* $\{h_k\}$ *converges to some element in* $[0, +\infty)$.

Lemma A.2 (Opial [16])**.** *Let* $\mathcal{H}$ *be a real Hilbert space, let* $\emptyset \neq \mathcal{S} \subset \mathcal{H}$ *and let* $\{p^k\}$ *be a sequence in* $\mathcal{H}$ *such that every weak cluster point of* $\{p^k\}$ *belongs to* $\mathcal{S}$ *and* $\lim_{k\to\infty} \|p^k - p\|$ *exists for every* $p \in \mathcal{S}$. *Then* $\{p^k\}$ *converges weakly to a point in* $\mathcal{S}$.

Lemma A.3. ([3, Lemma A.2]) *The inverse function of the scalar map*

$$(0,2) \ni \beta \mapsto \frac{2(2-\beta)}{4-\beta+\sqrt{16\beta-7\beta^2}} \in (0,1)$$

is given by

$$(0,1) \ni \overline{\alpha} \mapsto \frac{2(\overline{\alpha}-1)^2}{2(\overline{\alpha}-1)^2+3\overline{\alpha}-1} \in (0,2).$$

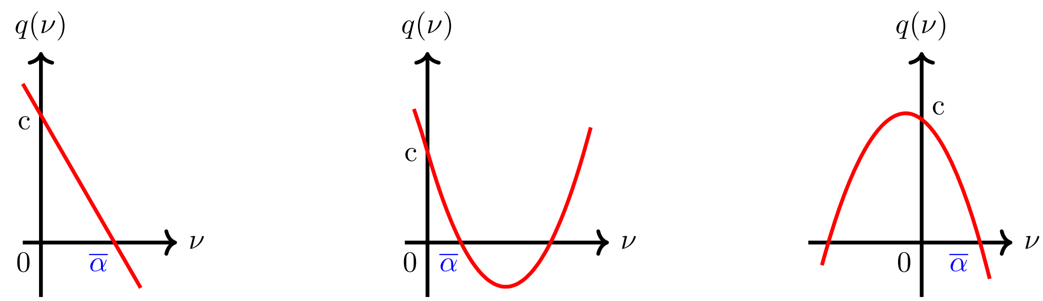
Lemma A.4. ([3, Lemma A.3]) *Let* $\mathbb{R} \ni \nu \mapsto q(\nu) := a\nu^2 - b\nu + c$ *be a real function and assume that* $b, c > 0$ *and* $b^2 - 4ac > 0$. *Define*

$$\overline{\alpha} := \frac{2c}{b+\sqrt{b^2-4ac}} > 0. \tag{82}$$

(i) *If* $a = 0$, *then* $q(\cdot)$ *is a decreasing affine function and* $\overline{\alpha} > 0$ *as in* (82) *is its unique root* (*see* Figure 5(a)).

(ii) *If* $a > 0$ (*resp.* $a < 0$), *then* $q(\cdot)$ *is a convex* (*resp. concave*) *quadratic function and* $\overline{\alpha} > 0$ *as in* (82) *is its smallest* (*resp. largest*) *root* (*see* Figure 5(b) *and* Figure 5(c), *resp.*).

In both cases (i) *and* (ii), $\overline{\alpha} > 0$ *as in* (82) *is a root of* $q(\cdot)$, *and* $q(\cdot)$ *is decreasing in the interval* $[0, \overline{\alpha}]$ (*see* Figure 5).



(a) $a = 0$ (b) $a > 0$ (c) $a < 0$

Figure 5: Possible cases for the real function $q(\cdot)$ in Lemma A.4.

The lemma below was proved (with a different notation) in [1, Proposition 2.4].

Lemma A.5. *Let* $\mathcal{H}$ *and* $\mathcal{G}$ *be real Hilbert spaces, let* $A : \mathcal{H} \rightrightarrows \mathcal{H}$ *and* $B : \mathcal{G} \rightrightarrows \mathcal{G}$ *be maximal monotone operators and let* $G : \mathcal{H} \to \mathcal{G}$ *be a bounded linear operator. Let also* $a^k \in A(r^k)$ *and* $b^k \in B(s^k)$ *be such that* $r^k \rightharpoonup r^\infty$ *and* $b^k \rightharpoonup b^\infty$, *for some* $r^\infty \in \mathcal{H}$ *and* $b^\infty \in \mathcal{G}$. *If,* $a^k + G^* b^k \to 0$ *and* $Gr^k - s^k \to 0$, *then* $b^\infty \in B(Gr^\infty)$ *and* $-G^* b^\infty \in A(r^\infty)$.

References

[1] A. Alotaibi, P.L. Combettes, and N. Shahzad. Solving coupled composite monotone inclusions by successive fejér approximations of their kuhn-tucker set. *SIAM J. Optim.*, 24(8):2076–2095, 2014.

[2] F. Alvarez and H. Attouch. An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Anal.*, 9(1-2):3–11, 2001. Wellposedness in optimization and related topics (Gargnano, 1999).

[3] M. Marques Alves, J. Eckstein, M. Geremia, and J.G. Melo. Relative-error inertial-relaxed inexact versions of Douglas-Rachford and ADMM splitting algorithms. Preprint arXiv:1904.10502 [math.OC], 2019.

[4] M. Marques Alves and R.T. Marcavillaca. On inexact relative-error hybrid proximal extragradient, forward-backward and Tseng's modified forward-backward methods with inertial effects,. *Set-Valued and Variational Analysis,* to appear, 2019.

[5] H. Attouch and A. Cabot. Convergence of a relaxed inertial proximal algorithm for maximally monotone operators. Preprint hal-01708905, 2018.

[6] H. Attouch and A. Cabot. Convergence of damped inertial dynamics governed by regularized maximally monotone operators. *J. Differential Equations*, 264(12):7138–7182, 2018.

[7] H. Attouch and J. Peypouquet. Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators. *Math. Program.*, 174(1-2, Ser. B):391–432, 2019.

[8] R. I. Boţ, E. R. Csetnek, and C. Hendrich. Inertial Douglas-Rachford splitting for monotone inclusion problems. *Appl. Math. Comput.*, 256:472–487, 2015.

[9] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, 2011.

[10] P. L. Combettes and L. E. Glaudin. Quasi-nonexpansive iterations on the affine hull of orbits: from Mann's mean value algorithm to inertial methods. *SIAM J. Optim.*, 27(4):2356–2380, 2017.

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

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

[13] P.R. Johnstone and J. Eckstein. Projective splitting with forward steps: Asynchronous and block-iterative operator splitting. Preprint arXiv:1803.07043v5, 2018.

[14] P.R. Johnstone and J. Eckstein. Convergence rates for projective splitting. *SIAM J. Optim.*, 29(3):1931–1957, 2019.

[15] P.R. Johnstone and J. Eckstein. Single-forward-step projective splitting:exploiting cocoercivity. Preprint arXiv:1902.09025v1, 2019.

[16] Z. Opial. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Amer. Math. Soc.*, 73:591–597, 1967.