

# A relative-error inexact ADMM splitting algorithm for convex optimization with inertial effects

M. Marques Alves \*

Marina Geremia †

*Dedicated to the memory of Professor Hedy Attouch*

## Abstract

We propose a new relative-error inexact version of the alternating direction method of multipliers (ADMM) for convex optimization. We prove the asymptotic convergence of our main algorithm as well as pointwise and ergodic iteration-complexities for residuals. We also justify the effectiveness of the proposed algorithm through some preliminary numerical experiments on regression problems.

2000 Mathematics Subject Classification: 90C25, 90C06, 49J52.

Key words: Convex optimization, ADMM, inexact, relative-error, inertial algorithms.

## 1 Introduction

Consider the convex problem

$$\underset{x \in \mathcal{H}}{\text{minimize}} \ f(x) + g(Lx), \quad (1)$$

where  $f : \mathcal{H} \rightarrow (-\infty, \infty]$  and  $g : \mathcal{G} \rightarrow (-\infty, \infty]$  are lower semicontinuous proper convex functions and  $L : \mathcal{H} \rightarrow \mathcal{G}$  is a linear operator ( $\mathcal{H}$  and  $\mathcal{G}$  denote finite-dimensional inner product spaces). Problem (1) appears in different contexts in applied mathematics, including optimization, inverse problems, machine learning, among others.

One of the most popular numerical algorithms for solving (1) is the alternating direction method of multipliers (ADMM) [26, 28, 29], which has now attracted a lot of attention from the numerical optimization community (see, e.g., [9, 11, 12, 13, 14, 16, 22, 23, 30, 32, 33, 35, 37, 42, 46]).

In this paper we propose and study a new inexact version of the ADMM allowing relative-error criteria for the solution of the second subproblem (which will appear in the formulation of the proposed algorithm) and promoting inertial effects on iterations.

**Organization of the paper.** The material is organized as follows. In Section 2 we motivate the definition of our main algorithm, review some related works and discuss the main contributions

---

\*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 grant 308036/2021-2.

†Departamento de Matemática, Universidade Federal de Santa Catarina, Florianópolis, Brazil, 88040-900. Departamento de Ensino, Pesquisa e Extensão, Instituto Federal de Santa Catarina (IFSC) (marina.geremia@ifsc.edu.br).

of this paper. In Section 3, we present our main algorithm (Algorithm 1) and some preliminary results that will be needed to study its convergence and iteration-complexity. In Section 4, we study the asymptotic behavior of Algorithm 1 under different assumptions under the inertial parameters involved in the formulation of the method. The main results are Theorems 4.2 and 4.4. In Section 5 we study the iteration-complexities of Algorithm 1. The main results in this section are Theorems 5.2 and 5.4. Numerical experiments will be presented in Section 6. Appendix A contains some auxiliary results.

## 2 Motivation, related works and contributions

**Motivation.** We first note that (1) is clearly equivalent to the separable problem

$$\begin{aligned} & \text{minimize} && f(x) + g(y), \\ & \text{subject to} && Lx - y = 0. \end{aligned} \tag{2}$$

An iteration of the standard ADMM [22] for solving (2) can be described as follows: given a starting point  $(y_0, z_0) \in \mathcal{G}^2$  and a regularization parameter  $\gamma > 0$ , iterate for  $k \geq 0$ :

$$x_{k+1} \in \operatorname{argmin}_{x \in \mathcal{H}} \left\{ f(x) + \langle z_k | Lx - y_k \rangle + \frac{\gamma}{2} \|Lx - y_k\|^2 \right\}, \tag{3}$$

$$y_{k+1} \in \operatorname{argmin}_{y \in \mathcal{G}} \left\{ g(y) + \langle z_k | Lx_{k+1} - y \rangle + \frac{\gamma}{2} \|Lx_{k+1} - y\|^2 \right\}, \tag{4}$$

$$z_{k+1} = z_k + \gamma (Lx_{k+1} - y_{k+1}). \tag{5}$$

We consider here the case in which (3) can be solved exactly and, on the other hand, (4) is supposed to be solved only approximately by some other (inner) algorithm, like, for instance, CG or BFGS, depending on the particular structure of the function  $g(\cdot)$  in (2).

With this in mind, we will introduce a notion of relative-error approximate solution for (4) (more details will be given on Section 3). To this end, first note that (4) is an instance of the general family of minimization problems

$$\operatorname{minimize}_{y \in \mathcal{G}} \left\{ g(y) + \langle z | Lx - y \rangle + \frac{\gamma}{2} \|Lx - y\|^2 \right\}, \tag{6}$$

where  $x \in \mathcal{H}$ ,  $z \in \mathcal{G}$  and  $\gamma > 0$  are given (in the case of (4), we have  $(x, z) = (x_{k+1}, z_k)$ ). Moreover, since the function  $g(\cdot)$  is convex, we have that (6) is also equivalent to the inclusion/equation system for the pair  $(y, v)$ :

$$\begin{cases} v \in \partial g(y), \\ v - z + \gamma(y - Lx) = 0. \end{cases} \tag{7}$$

A formal definition of approximate (inexact) solution of (7) (or, equivalently, (6)) will be given in Definition 3.1 in Section 3; such a notion of approximate solution will allow for errors in both the inclusion and the equation in (7).

**The extended-solution set.** The Fenchel dual of (1) is

$$\underset{z \in \mathcal{G}}{\text{maximize}} -f^*(-L^*z) - g^*(z), \quad (8)$$

where  $f^* : \mathcal{H} \rightarrow (-\infty, \infty]$  and  $g^* : \mathcal{G} \rightarrow (-\infty, \infty]$  denote the Fenchel conjugates of  $f$  and  $g$ , respectively, and  $L^* : \mathcal{G} \rightarrow \mathcal{H}$  denotes the adjoint operator of  $L$ . Under standard regularity conditions [10] on  $f, g$  and  $L$  it is well-known that (1) and (8) are, respectively, equivalent to the (monotone) inclusions

$$0 \in \partial f(x) + L^* \partial g(Lx), \quad (9)$$

and

$$0 \in -L \partial f^*(-L^*z) + \partial g^*(z). \quad (10)$$

We make the blanket assumption:

**Assumption 2.1.** *For the function  $f$  and the operator  $L$  as in (1), the following holds:*

$$\partial(f^* \circ -L^*) = -L \circ \partial f^* \circ -L^*.$$

Several sufficient conditions for Assumption 2.1 to hold true can be found, e.g., in [10]. We will also consider an extended-solution set  $\mathcal{S}$ , attached to the pair of inclusions (9)–(10), defined as

$$\mathcal{S} = \{(z, w) \in \mathcal{G}^2 \mid -w \in \partial(f^* \circ -L^*)(z) \text{ and } w \in \partial g^*(z)\}. \quad (11)$$

Under Assumption 2.1, it is easy to check that if  $(z, w) \in \mathcal{S}$ , then it follows that there exists  $x \in \mathcal{H}$  such that  $x \in \partial f^*(-L^*z)$ ,  $w = Lx$  and  $x$  and  $z$  are solutions of (9) and (10), respectively.

Throughout this work we will assume the following.

**Assumption 2.2.** *We assume the extended solution set  $\mathcal{S}$  as in (11) is nonempty.*

**Inertial algorithms.** Iterative algorithms with inertial effects for monotone inclusions (and related topics in optimization, saddle-point, equilibrium problems, etc) were first proposed in the seminal paper [2] and subsequently developed in various directions of research by different authors and research groups (see, e.g., [3, 5, 6, 7, 8, 12, 17] and references therein). Basically, the main idea consists in at a current iterate, say  $p_k$ , produce an “inertial effect” by a simple extrapolation:

$$\widehat{p}_k = p_k + \alpha_k(p_k - p_{k-1}),$$

where  $\alpha_k \geq 0$ , and then generate the next iterate  $p_{k+1}$  from  $\widehat{p}_k$  instead of  $p_k$  (see (18)–(19) below). Our main algorithm, namely Algorithm 1, will benefit from inertial effects on the iteration; see the comments and remarks following Algorithm 1 for more discussions regarding the effects of inertia.

**Main contributions.** We present a theoretical (asymptotic and iteration-complexity analysis) and computational study of a partially inexact ADMM splitting algorithm for solving (1). Our main algorithm, namely Algorithm 1 below, benefits from the addition of inertial effects; see (18) and (19). The convergence analysis is presented in Theorem 4.2, to which the proof incorporates some elements of [3] and [46]. We also obtained iteration-complexities for the proposed algorithm by showing pointwise  $O(1/\sqrt{k})$  and ergodic  $O(1/k)$  global convergence rates for residuals; see Theorems

5.2 and 5.4 below. We justify the effectiveness of our main algorithm through the realization of numerical experiments on the LASSO problem (see Section 6).

**Related works.** A partially inexact ADMM splitting algorithm was recently proposed and studied in [46]. Paper [1] proposes a partially inexact ADMM for which the first subproblem is supposed to be solved inexactly. The analysis of the main algorithm in [1] is performed by viewing it as a special instance of a non-Euclidean version of the hybrid proximal extragradient method [36]. In contrast to this, analogously to [46], our main algorithm (Algorithm 1 below) assumes the second subproblem is solved inexactly. Moreover, since (22)–(24) below also allows for errors in  $\partial g$ , the error criterion we propose here is potentially more flexible than the corresponding one in [1]. Other relative-error inexact versions of ADMM were also previously studied in [49, 50], but we notice that the convergence results were restricted to the analysis of the dual sequences. We also mention that the relative-error inexact variants of the ADMM from [3, 23] only apply to (1) in the particular case of  $L = I$ , and, additionally, these variants assume the first subproblem to be solved inexactly with the error condition verified only a-posteriori, that is, only after the computation of second subproblem's solution.

**General notation.** We denote by  $\mathcal{H}$  and  $\mathcal{G}$  finite-dimensional inner product spaces with inner product and induced norm denoted, respectively, by  $\langle \cdot | \cdot \rangle$  and  $\|\cdot\| = \sqrt{\langle \cdot | \cdot \rangle}$ . For any set  $\mathcal{X}$  we denote by  $X^n$  the  $n$ -product  $\mathcal{X} \times \cdots \times \mathcal{X}$ . In  $\mathcal{G}^2$ , we will consider the inner product and induced norm defined, respectively, by

$$\langle p | p' \rangle_\gamma := \frac{1}{\gamma} \langle z | z' \rangle + \gamma \langle w | w' \rangle \quad \text{and} \quad \|p\|_\gamma^2 := \langle p | p \rangle_\gamma, \quad (12)$$

where  $p = (z, w), p' = (z', w') \in \mathcal{G}^2$  and  $\gamma > 0$ . More precisely, for  $p = (z, w) \in \mathcal{G}^2$ , the norm of  $p$  is

$$\|p\|_\gamma^2 = \frac{1}{\gamma} \|z\|^2 + \gamma \|w\|^2. \quad (13)$$

An extended-real valued function  $f : \mathcal{H} \rightarrow (-\infty, \infty]$  is said to be *convex* whenever  $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$  for all  $x, y \in \mathcal{H}$  and  $\lambda \in (0, 1)$ , and  $f$  is *proper* if its *effective domain*, denoted by  $\text{dom } f$ , is nonempty. The *Fenchel conjugate* of a proper function  $f : \mathcal{H} \rightarrow (-\infty, \infty]$  is  $f^* : \mathcal{H} \rightarrow (-\infty, \infty]$ , defined at any  $u \in \mathcal{H}$  by  $f^*(u) = \sup_{x \in \mathcal{H}} \{\langle x | u \rangle - f(x)\}$ . The  $\varepsilon$ -*subdifferential* and the *subdifferential* of a convex function  $g : \mathcal{H} \rightarrow (-\infty, \infty]$  at  $x \in \mathcal{H}$  are defined as  $\partial_\varepsilon g(x) := \{u \in \mathcal{H} \mid g(y) \geq g(x) + \langle u | y - x \rangle - \varepsilon \quad \forall y \in \mathcal{H}\}$  and  $\partial g(x) := \partial_0 g(x)$ , respectively. For additional details on standard notations and definitions of convex analysis we refer the reader to the references [10, 41].

### 3 The main algorithm and some preliminary results

Consider the minimization problem (1), i.e.,

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad f(x) + g(Lx), \quad (14)$$

where  $f : \mathcal{H} \rightarrow (-\infty, \infty]$  and  $g : \mathcal{G} \rightarrow (-\infty, \infty]$  are lower semicontinuous proper convex functions and  $L : \mathcal{H} \rightarrow \mathcal{G}$  is a linear operator between finite-dimensional inner product spaces  $\mathcal{H}$  and  $\mathcal{G}$ .

In this section we present our main algorithm, namely Algorithm 1 below. This is a partially inexact (the second block is allowed to be solved inexactly) ADMM with relative-error criterion for the second subproblem. Recall the extended solution set  $\mathcal{S}$  as in (11) and Assumptions 2.1 and 2.2. The three technical lemmas 3.2, 3.3, 3.4 and 3.5 will be used in the subsequent section.

Before presenting our main algorithm, as we discussed in the Introduction, we have to formalize the notion of inexact solution that will be used to compute approximate solution for the second subproblem. Recall that the second subproblem of the standard ADMM (see (4)) belongs to the general family of minimization problems (6), which is, in particular, equivalent to the inclusion/equation system (7) for the pair  $(y, v)$ , i.e.,

$$\begin{cases} v \in \partial g(y), \\ v - z + \gamma(y - Lx) = 0. \end{cases} \quad (15)$$

**Definition 3.1** ( $\sigma$ -approximate solution of (6)). *For  $x \in \mathcal{H}$ ,  $(\hat{z}, \hat{y}) \in \mathcal{G}^2$  and  $\gamma > 0$ , a triple  $(\tilde{y}, v, \varepsilon) \in \mathcal{G} \times \mathcal{G} \times \mathbb{R}_+$  is said to be a  $\sigma$ -approximate solution of (6) (or, equivalently, of (7)) at  $(x, \hat{z}, \hat{y})$  if  $\sigma \in [0, 1]$  and*

$$\begin{cases} v \in \partial_\varepsilon g(\tilde{y}), \\ v - \hat{z} + \gamma(\tilde{y} - Lx) =: e \\ \|e\|^2 + 2\gamma\varepsilon \leq \sigma^2 \min \{ \gamma^2 \|Lx - \hat{y}\|^2, \|v - \hat{z}\|^2 \}. \end{cases} \quad (16)$$

We will also write

$$\tilde{y} \stackrel{\sigma}{\approx} \operatorname{argmin}_{y \in \mathcal{G}} \left\{ g(y) + \langle \hat{z} | Lx - y \rangle + \frac{\gamma}{2} \|Lx - y\|^2 \right\}$$

meaning that there exists  $(v, \varepsilon)$  such that  $(\tilde{y}, v, \varepsilon)$  satisfies (16).

We now make some remarks regarding Definition 3.1:

- (i) Note that if  $\sigma = 0$  in (16), then it follows that  $e = 0$  and  $\varepsilon = 0$ , which is to say that the pair  $(\tilde{y}, v)$  satisfies the inclusion/equation system (7) (recall that  $\partial_0 g = \partial g$ ) and, in particular,  $\tilde{y}$  is an exact solution of (6).
- (ii) The error criterion for (7) as in (16) belongs to the class of *relative-error* criteria for proximal-type algorithms. Different variants of such error conditions have been employed for computing approximate solution for (sub) problems for a wide range of algorithms in monotone inclusions, convex optimization, saddle-point problems, etc (see, e.g., [3, 23, 36, 43, 44, 45]).
- (iii) The error criterion (16) will be used to compute approximate solutions in step 3 of our main algorithm, namely Algorithm 1 below (see (22)–(24)).
- (iv) As an illustrative example, consider the special case of the LASSO problem [47]

$$\min_{x \in \mathbb{R}^d} \left\{ \frac{1}{2} \|Ax - b\|^2 + \nu \|x\|_1 \right\}, \quad (17)$$

where  $A \in \mathbb{R}^{n \times d}$ ,  $b \in \mathbb{R}^n$  and  $\nu > 0$ . Problem (17) is clearly a special instance of (1) in which  $L := I$ ,  $f(x) := \nu \|x\|_1$  and  $g(x) := (1/2) \|Ax - b\|^2$  (see also Section 6 below). In this case, our inclusion/equation system (7) clearly reduces to

$$v = A^*(Ay - b), \quad v - z + \gamma(y - x) = 0,$$

or, in other words, in this special case, (7) is equivalent to the linear system (operator equation)

$$(A^*A + \gamma I)y = A^*b + z + \gamma x.$$

The latter linear system can be solved by the CG algorithm [39], where  $e$  as in (16) will simply denote the residual of the system and the inequality in (16) can be used as a stopping criterion for CG.

Next is our main algorithm.

**Algorithm 1. An inexact inertial ADMM algorithm for solving (1)**

(0) Let  $(z_0, y_0) = (z_{-1}, y_{-1}) \in \mathcal{G}^2$  and let  $\alpha, \sigma \in [0, 1)$ ,  $\tau \in (0, 1)$  and  $\gamma > 0$ . Set  $k = 0$ .

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

$$\widehat{z}_k = z_k + \alpha_k(z_k - z_{k-1}), \quad (18)$$

$$\widehat{y}_k = y_k + \alpha_k(y_k - y_{k-1}). \quad (19)$$

(2) Compute

$$x_k \in \operatorname{argmin}_{x \in \mathcal{H}} \left\{ f(x) + \langle \widehat{z}_k | Lx - \widehat{y}_k \rangle + \frac{\gamma}{2} \|Lx - \widehat{y}_k\|^2 \right\}. \quad (20)$$

(3) Compute

$$\widetilde{y}_k \overset{\sigma}{\approx} \operatorname{argmin}_{y \in \mathcal{G}} \left\{ g(y) + \langle \widehat{z}_k | Lx_k - y \rangle + \frac{\gamma}{2} \|Lx_k - y\|^2 \right\} \quad (21)$$

at  $(x_k, \widehat{z}_k, \widehat{y}_k)$  in the sense of Definition 3.1, i.e., compute  $(\widetilde{y}_k, v_k, \varepsilon_k) \in \mathcal{G} \times \mathcal{G} \times \mathbb{R}_+$  such that

$$v_k \in \partial_{\varepsilon_k} g(\widetilde{y}_k), \quad (22)$$

$$v_k - \widehat{z}_k + \gamma(\widetilde{y}_k - Lx_k) =: e_k, \quad (23)$$

$$\|e_k\|^2 + 2\gamma\varepsilon_k \leq \sigma^2 \min \{ \gamma^2 \|Lx_k - \widehat{y}_k\|^2, \|v_k - \widehat{z}_k\|^2 \}. \quad (24)$$

(4) Set

$$z_{k+1} = \widehat{z}_k + \tau\gamma(Lx_k - \widetilde{y}_k), \quad (25)$$

$$y_{k+1} = (1 - \tau)\widehat{y}_k + \frac{\tau}{\gamma}(\widehat{z}_k + \gamma Lx_k - v_k), \quad (26)$$

$k = k + 1$  and go to step 1.

We now make some remarks concerning Algorithm 1:

- (i) Algorithm 1 is specially designed for instances of (1) in which (20) has a closed-form solution, i.e., for problems in which (20) is easy to solve. In this regard, one example of interest is when  $f(\cdot) = \|\cdot\|_1$  and  $L = I$ , in which case (20) has a unique solution given explicitly by  $x_k = \text{prox}_{\gamma^{-1}\|\cdot\|_1}(\widehat{y}_k - \gamma^{-1}\widehat{z}_k)$ . On the other hand, we assume that the computation of  $\widetilde{y}_k$  as in (21) demands the use of an (inner) algorithm, which the choice of depends on the particular structure of the function  $g$ , and, in this case, one can use (22)–(24) as a stopping criterion for the inner algorithm of choice.
- (ii) Recall that we discussed in the Introduction (see “Related works”) other ADMM-type algorithms related to Algorithm 1.
- (iii) The main results on the convergence and iteration-complexity of Algorithm 1 are Theorems 4.2, 4.4, 5.2 and 5.4 below. Numerical experiments will be presented and discussed in Section 6.
- (iv) The role of the parameter  $0 < \tau < 1$  is to introduce (under) relaxation in the iterative process; see (25) and (26).
- (v) We will also need the sequences  $(\check{z}_k)$  and  $(\check{y}_k)$ , where, for all  $k \geq 0$ ,

$$\check{z}_k := \widehat{z}_k + \gamma(Lx_k - \widetilde{y}_k), \quad \check{y}_k := \frac{1}{\gamma}(\widehat{z}_k + \gamma Lx_k - v_k). \quad (27)$$

Note that  $\check{z}_k = z_{k+1}$  and  $\check{y}_k = y_{k+1}$  if we set  $\tau = 1$  in (25) and (26), respectively.

Next we present four technical lemmas – Lemmas 3.2, 3.3, 3.4 and 3.5 –, which will be useful in the subsequent sections.

**Lemma 3.2.** *Consider the sequences evolved by Algorithm 1, let  $\mathcal{S}$  be as in (11) and let  $(\check{z}_k)$  and  $(\check{y}_k)$  be as in (27). Define*

$$p_k = (z_k, y_k), \quad \widehat{p}_k = (\widehat{z}_k, \widehat{y}_k), \quad \check{p}_k = (\check{z}_k, \check{y}_k) \quad \text{and} \quad \widetilde{p}_k = (v_k, Lx_k) \quad \forall k \geq 0. \quad (28)$$

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

$$p_{k+1} = (1 - \tau)\widehat{p}_k + \tau\check{p}_k.$$

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

$$-Lx_k \in \partial(f^* \circ -L^*)(z'_k),$$

where

$$z'_k := \widehat{z}_k + \gamma(Lx_k - \widehat{y}_k). \quad (29)$$

- (c) For all  $k \geq 0$ ,

$$\check{p}_k - \widehat{p}_k = \left( \gamma(Lx_k - \widetilde{y}_k), \frac{1}{\gamma}(\check{z}'_k - v_k) \right).$$

(d) For all  $k \geq 0$  and  $p = (z, w) \in \mathcal{S}$ ,

$$\langle \check{p}_k - \hat{p}_k \mid p - \tilde{p}_k \rangle_\gamma \geq -\varepsilon_k.$$

(e) For all  $k \geq 0$ ,

$$\hat{p}_k = p_k + \alpha_k(p_k - p_{k-1}).$$

*Proof.* (a) This result is a direct consequence of the definitions of  $p_k$ ,  $\hat{p}_k$  and  $\check{p}_k$  as in (28) – see also (27) – combined with (25) and (26).

(b) First note that from (20) and (29), we obtain  $0 \in \partial f(x_k) + L^* z'_k$ , or, equivalently,  $-L^* z'_k \in \partial f(x_k)$ . As  $(\partial f)^{-1} = \partial f^*$ , the latter inclusion is also equivalent to  $x_k \in \partial f^*(-L^* z'_k)$ , which in turn yields  $-Lx_k \in -L\partial f^*(-L^* z'_k)$ , which by Assumption 2.1 gives item (b).

(c) This follows easily from (27) – (29) and some simple algebraic manipulations.

(d) As  $(\partial_{\varepsilon_k} g)^{-1} = \partial_{\varepsilon_k} g^*$  – see, e.g., [51, p. 85, Theorem 2.4.4(iv)] –, we have that (22) is equivalent to the inclusion

$$\tilde{y}_k \in \partial_{\varepsilon_k} g^*(v_k). \quad (30)$$

As  $p = (z, w) \in \mathcal{S}$ , according to the definition of  $\mathcal{S}$  in (11), we have  $-w \in \partial(f^* \circ -L^*)(z)$  and  $w \in \partial g^*(z)$ . The latter inclusions combined with item (b) and (30) and the monotonicity of  $\partial(f^* \circ -L^*)$  and  $\partial g^*$  yield

$$\langle z'_k - z \mid w - Lx_k \rangle \geq 0 \quad \text{and} \quad \langle z - v_k \mid w - \tilde{y}_k \rangle \geq -\varepsilon_k. \quad (31)$$

Now using (28), item (c), (31) and the definition of  $\langle \cdot \mid \cdot \rangle_\gamma$  as in (12) we find

$$\begin{aligned} \langle \check{p}_k - \hat{p}_k \mid p - \tilde{p}_k \rangle_\gamma &= \frac{1}{\gamma} \langle \gamma(Lx_k - \tilde{y}_k) \mid z - v_k \rangle + \gamma \left\langle \frac{1}{\gamma} (z'_k - v_k) \mid w - Lx_k \right\rangle \\ &= \langle Lx_k - \tilde{y}_k \mid z - v_k \rangle + \langle z'_k - z \mid w - Lx_k \rangle + \langle z - v_k \mid w - Lx_k \rangle \\ &= \langle z - v_k \mid w - \tilde{y}_k \rangle + \langle z'_k - z \mid w - Lx_k \rangle \\ &\geq -\varepsilon_k. \end{aligned}$$

(e) This follows directly from (18), (19) and the definition of  $\hat{p}_k$  as in (28).  $\square$

**Lemma 3.3.** *Consider the sequences evolved by Algorithm 1 and let  $(\check{p}_k)$ ,  $(\hat{p}_k)$  and  $(\tilde{p}_k)$  be as in (28). For all  $k \geq 0$ ,*

- (a)  $\|\tilde{p}_k - \check{p}_k\|_\gamma^2 = \frac{1}{\gamma} \left( \|e_k\|^2 + \|v_k - \hat{z}_k\|^2 \right).$
- (b)  $\|\check{p}_k - \hat{p}_k\|_\gamma \leq 2\|\tilde{p}_k - \hat{p}_k\|_\gamma.$



*Proof.* (a) Direct use of (13), (23), (27) and (28) gives

$$\begin{aligned}
\|\tilde{p}_k - \check{p}_k\|_\gamma^2 &= \frac{1}{\gamma} \|v_k - \check{z}_k\|^2 + \gamma \|Lx_k - \check{y}_k\|^2 \\
&= \frac{1}{\gamma} \underbrace{\|v_k - \hat{z}_k + \gamma(\tilde{y}_k - Lx_k)\|}_{e_k}^2 + \gamma \|Lx_k - [\gamma^{-1}(\hat{z}_k + \gamma Lx_k - v_k)]\|^2 \\
&= \frac{1}{\gamma} (\|e_k\|^2 + \|v_k - \hat{z}_k\|^2).
\end{aligned}$$

(b) In view of (24), (28) and item (a),

$$\begin{aligned}
\|\tilde{p}_k - \check{p}_k\|_\gamma^2 &= \frac{1}{\gamma} (\|e_k\|^2 + \|v_k - \hat{z}_k\|^2) \\
&\leq \frac{1}{\gamma} (\sigma^2 \|\gamma(Lx_k - \hat{y}_k)\|^2 + \|v_k - \hat{z}_k\|^2) \\
&\leq \frac{1}{\gamma} \|v_k - \hat{z}_k\|^2 + \gamma \|Lx_k - \hat{y}_k\|^2 \\
&= \|\tilde{p}_k - \hat{p}_k\|_\gamma^2.
\end{aligned}$$

Hence, using the triangle inequality,

$$\|\check{p}_k - \hat{p}_k\|_\gamma \leq \|\tilde{p}_k - \check{p}_k\|_\gamma + \|\tilde{p}_k - \hat{p}_k\|_\gamma \leq 2\|\tilde{p}_k - \hat{p}_k\|_\gamma.$$

□

**Lemma 3.4.** *Consider the sequences evolved by Algorithm 1 and let  $(\hat{p}_k)$ ,  $(\check{p}_k)$  and  $(\tilde{p}_k)$  be as in (28).*

(a) *For all  $k \geq 0$  and  $p \in \mathcal{G}^2$ ,*

$$\|p - \hat{p}_k\|_\gamma^2 - \|p - \check{p}_k\|_\gamma^2 = \|\tilde{p}_k - \hat{p}_k\|_\gamma^2 - \|\tilde{p}_k - \check{p}_k\|_\gamma^2 + 2\langle \check{p}_k - \hat{p}_k | p - \tilde{p}_k \rangle_\gamma.$$

(b) *For all  $k \geq 0$  and  $p \in \mathcal{G}^2$ ,*

$$\|p - \hat{p}_k\|_\gamma^2 - \|p - \check{p}_k\|_\gamma^2 \geq \gamma(1 - \sigma^2) \|Lx_k - \hat{y}_k\|^2 + 2[\varepsilon_k + \langle \check{p}_k - \hat{p}_k | p - \tilde{p}_k \rangle_\gamma].$$

(c) *For all  $k \geq 0$  and  $p = (z, w) \in \mathcal{S}$ ,*

$$\|p - \hat{p}_k\|_\gamma^2 - \|p - \check{p}_k\|_\gamma^2 \geq \gamma(1 - \sigma^2) \|Lx_k - \hat{y}_k\|^2.$$

(d) *For all  $k \geq 0$  and  $p = (z, w) \in \mathcal{S}$ ,*

$$\|p - \hat{p}_k\|_\gamma^2 - \|p - p_{k+1}\|_\gamma^2 \geq \tau\gamma(1 - \sigma^2) \|Lx_k - \hat{y}_k\|^2 + (1 - \tau)\tau \|\check{p}_k - \hat{p}_k\|_\gamma^2.$$

(e) For all  $k \geq 0$  and  $p = (z, w) \in \mathcal{S}$ ,

$$\begin{aligned} \|p - \widehat{p}_k\|_\gamma^2 - \|p - p_{k+1}\|_\gamma^2 &\geq \tau(1 - \tau)(1 - \sigma)^2 \|\widetilde{p}_k - \widehat{p}_k\|_\gamma^2 \\ &\geq \frac{(1 - \tau)(1 - \sigma)^2}{4\tau} \|p_{k+1} - \widehat{p}_k\|_\gamma^2. \end{aligned}$$

*Proof.* (a) The desired result follows directly from the well-known identity  $\|a - b\|_\gamma^2 - \|a - c\|_\gamma^2 = \|d - b\|_\gamma^2 - \|d - c\|_\gamma^2 + 2\langle c - b | a - d \rangle_\gamma$  with  $a = p$ ,  $b = \widehat{p}_k$ ,  $c = \check{p}_k$  and  $d = \widetilde{p}_k$ .

(b) Using (13) and the definitions of  $(\widetilde{p}_k)$  and  $(\widehat{p}_k)$  as in (28) we get

$$\|\widetilde{p}_k - \widehat{p}_k\|_\gamma^2 = \frac{1}{\gamma} \|v_k - \widehat{z}_k\|^2 + \gamma \|Lx_k - \widehat{y}_k\|^2,$$

which in turn combined with Lemma 3.3(a) yields

$$\|\widetilde{p}_k - \widehat{p}_k\|_\gamma^2 - \|\widetilde{p}_k - \check{p}_k\|_\gamma^2 = \gamma \|Lx_k - \widehat{y}_k\|^2 - \frac{1}{\gamma} \|e_k\|^2.$$

From (24), item (a), the latter identity and some algebraic manipulations,

$$\begin{aligned} \|p - \widehat{p}_k\|_\gamma^2 - \|p - \check{p}_k\|_\gamma^2 &= \gamma \|Lx_k - \widehat{y}_k\|^2 - \frac{1}{\gamma} \|e_k\|^2 + 2\langle \check{p}_k - \widehat{p}_k | p - \widetilde{p}_k \rangle_\gamma \\ &= \gamma \|Lx_k - \widehat{y}_k\|^2 - \frac{1}{\gamma} (\|e_k\|^2 + 2\gamma \varepsilon_k) + 2[\varepsilon_k + \langle \check{p}_k - \widehat{p}_k | p - \widetilde{p}_k \rangle_\gamma] \\ &\geq \gamma(1 - \sigma^2) \|Lx_k - \widehat{y}_k\|^2 + 2[\varepsilon_k + \langle \check{p}_k - \widehat{p}_k | p - \widetilde{p}_k \rangle_\gamma], \end{aligned}$$

which finishes the proof of (b).

(c) This is a direct consequence of Lemma 3.2(d) and item (b) above.

(d) Using Lemma 3.2(a) and the identity  $\|(1 - \tau)a + \tau b\|_\gamma^2 = (1 - \tau)\|a\|_\gamma^2 + \tau\|b\|_\gamma^2 - (1 - \tau)\tau\|a - b\|_\gamma^2$  with  $a = p - \widehat{p}_k$  and  $b = p - \check{p}_k$ , we obtain

$$\begin{aligned} \|p - p_{k+1}\|_\gamma^2 &= \|(1 - \tau)(p - \widehat{p}_k) + \tau(p - \check{p}_k)\|_\gamma^2 \\ &= (1 - \tau)\|p - \widehat{p}_k\|_\gamma^2 + \tau\|p - \check{p}_k\|_\gamma^2 - (1 - \tau)\tau\|\check{p}_k - \widehat{p}_k\|_\gamma^2. \end{aligned}$$

Now by multiplying the inequality in item (c) by  $\tau > 0$ , using the latter identity and some simple algebraic manipulations, we find the desired result.

(e) Note first that using (24) and the triangle inequality, we find

$$\begin{aligned} \|v_k - \widehat{z}_k\| &\leq \underbrace{\|v_k - \widehat{z}_k + \gamma(\widetilde{y}_k - Lx_k)\|}_{e_k} + \|\gamma(\widetilde{y}_k - Lx_k)\| \\ &\leq \sigma \|v_k - \widehat{z}_k\| + \|\gamma(\widetilde{y}_k - Lx_k)\|, \end{aligned}$$

so that

$$\|\gamma(\widetilde{y}_k - Lx_k)\| \geq (1 - \sigma) \|v_k - \widehat{z}_k\|,$$

which in turn combined with (13), (27) and (28) yields

$$\begin{aligned}
\|\check{p}_k - \widehat{p}_k\|_\gamma^2 &= \frac{1}{\gamma} \|\check{z}_k - \widehat{z}_k\|^2 + \gamma \|\check{y}_k - \widehat{y}_k\|^2 \\
&\geq \frac{1}{\gamma} \|\check{z}_k - \widehat{z}_k\|^2 \\
&= \frac{1}{\gamma} \|\gamma(\widehat{y}_k - Lx_k)\|^2 \\
&\geq \frac{1}{\gamma} (1 - \sigma)^2 \|v_k - \widehat{z}_k\|^2.
\end{aligned} \tag{32}$$

Now using (12), (28), item (d) above and (32),

$$\begin{aligned}
\|p - \widehat{p}_k\|_\gamma^2 - \|p - p_{k+1}\|_\gamma^2 &\geq \tau\gamma(1 - \sigma^2) \|Lx_k - \widehat{y}_k\|^2 + (1 - \tau)\tau \|\check{p}_k - \widehat{p}_k\|_\gamma^2 \\
&\geq \tau\gamma(1 - \sigma^2) \|Lx_k - \widehat{y}_k\|^2 + (1 - \tau)\tau \frac{1}{\gamma} (1 - \sigma)^2 \|v_k - \widehat{z}_k\|^2 \\
&\geq \tau(1 - \tau)(1 - \sigma)^2 \left[ \frac{1}{\gamma} \|v_k - \widehat{z}_k\|^2 + \gamma \|Lx_k - \widehat{y}_k\|^2 \right] \\
&= \tau(1 - \tau)(1 - \sigma)^2 \|\widetilde{p}_k - \widehat{p}_k\|_\gamma^2.
\end{aligned} \tag{33}$$

To prove the second inequality, one can use Lemmas 3.2(a) and 3.3(b) to conclude that  $\|p_{k+1} - \widehat{p}_k\|_\gamma \leq 2\tau \|\widetilde{p}_k - \widehat{p}_k\|_\gamma$  and then apply it in (33).  $\square$

**Lemma 3.5.** *Consider the sequences evolved by Algorithm 1 and let  $(p_k)$  and  $(\widehat{p}_k)$  be as in (28). Then, for all  $k \geq 0$ ,*

$$\|\widehat{p}_k - p\|_\gamma^2 = (1 + \alpha_k) \|p_k - p\|_\gamma^2 - \alpha_k \|p_{k-1} - p\|_\gamma^2 + \alpha_k(1 + \alpha_k) \|p_k - p_{k-1}\|_\gamma^2 \quad \forall p \in \mathcal{G}^2.$$

*Proof.* Recall that from Lemma 3.2(e) we have

$$\widehat{p}_k = p_k + \alpha_k(p_k - p_{k-1}), \tag{34}$$

which is clearly equivalent to

$$p_k - p = \frac{1}{1 + \alpha_k} (\widehat{p}_k - p) + \frac{\alpha_k}{1 + \alpha_k} (p_{k-1} - p).$$

Now using the well-known identity  $\|tx + (1 - t)y\|_\gamma^2 = t\|x\|_\gamma^2 + (1 - t)\|y\|_\gamma^2 - t(1 - t)\|x - y\|_\gamma^2$  with  $t = 1/(1 + \alpha_k)$ ,  $x = \widehat{p}_k - p$  and  $y = p_{k-1} - p$ , we find

$$\|p_k - p\|_\gamma^2 = \frac{1}{1 + \alpha_k} \|\widehat{p}_k - p\|_\gamma^2 + \frac{\alpha_k}{1 + \alpha_k} \|p_{k-1} - p\|_\gamma^2 - \frac{\alpha_k}{(1 + \alpha_k)^2} \|\widehat{p}_k - p_{k-1}\|_\gamma^2,$$

which, in turn, when combined with the fact that  $\widehat{p}_k - p_{k-1} = (1 + \alpha_k)(p_k - p_{k-1})$  – see (34) – and after some simple algebraic manipulations it yields

$$\|\widehat{p}_k - p\|_\gamma^2 = (1 + \alpha_k) \|p_k - p\|_\gamma^2 - \alpha_k \|p_{k-1} - p\|_\gamma^2 + \alpha_k(1 + \alpha_k) \|p_k - p_{k-1}\|_\gamma^2.$$

$\square$

## 4 Asymptotic convergence of Algorithm 1

In this section, we study the asymptotic convergence of Algorithm 1. The main results are Theorems 4.2 and 4.4.

**Lemma 4.1.** *Consider the sequences evolved by Algorithm 1 and, for an arbitrary  $p = (z, w) \in \mathcal{S}$ , define*

$$h_k = \|p_k - p\|_\gamma^2 \quad \forall k \geq -1. \quad (35)$$

Then  $h_0 = h_{-1}$  and, for all  $k \geq 0$ ,

$$h_{k+1} - h_k - \alpha_k(h_k - h_{k-1}) + \tau(1 - \tau)(1 - \sigma)^2 \|\tilde{p}_k - \hat{p}_k\|_\gamma^2 \leq \alpha_k(1 + \alpha_k) \|p_k - p_{k-1}\|_\gamma^2,$$

i.e.,  $(h_k)$  satisfies the assumptions of Lemma A.1 below, where, for all  $k \geq 0$ ,

$$s_{k+1} := \tau(1 - \tau)(1 - \sigma)^2 \|\tilde{p}_k - \hat{p}_k\|_\gamma^2, \quad (36)$$

$$\delta_k := \alpha_k(1 + \alpha_k) \|p_k - p_{k-1}\|_\gamma^2. \quad (37)$$

*Proof.* The fact that  $h_0 = h_{-1}$  follows directly from the fact that  $p_0 = p_{-1}$  (see step 0 in Algorithm 1 and the definition of  $p_k$  as in (28)). On the other hand, from Lemma 3.5 and the definition of  $h_k$  as in (35),

$$\|\hat{p}_k - p\|_\gamma^2 = (1 + \alpha_k) \underbrace{\|p_k - p\|_\gamma^2}_{h_k} - \alpha_k \underbrace{\|p_{k-1} - p\|_\gamma^2}_{h_{k-1}} + \alpha_k(1 + \alpha_k) \|p_k - p_{k-1}\|_\gamma^2.$$

The desired result now follows from the above displayed equation, Lemma 3.4(e) and the definition of  $h_k$  as in (35).  $\square$

Next we present our first result on the convergence of Algorithm 1 when  $k \rightarrow +\infty$ .

**Theorem 4.2** (First result on the asymptotic convergence of Algorithm 1). *Consider the sequences evolved by Algorithm 1 and let  $\emptyset \neq \mathcal{S}$  be as in (11). Assume that*

$$\sum_{k=0}^{\infty} \alpha_k \|p_k - p_{k-1}\|_\gamma^2 < \infty, \quad (38)$$

where  $(p_k)$  is as in (28). Then there exists  $(z_\infty, w_\infty) \in \mathcal{S}$  such that

$$z_k \rightarrow z_\infty \quad \text{and} \quad y_k \rightarrow w_\infty. \quad (39)$$

Additionally, we also have

$$v_k \rightarrow z_\infty, \quad Lx_k \rightarrow w_\infty \quad \text{and} \quad \tilde{y}_k \rightarrow w_\infty. \quad (40)$$

*Proof.* We start by making a few remarks. First, from (38) and the fact that  $\alpha_k(1 + \alpha_k) \leq 2\alpha_k$  (because  $0 \leq \alpha_k < 1$ ), we conclude that  $\sum_{k=0}^{\infty} \delta_k < \infty$ , where  $\delta_k$  is as in (37), which, in turn, combined with Lemmas 4.1 and A.1 (below) gives

$$\lim_{k \rightarrow \infty} h_k \text{ exists and } \sum_{k=1}^{\infty} s_k < \infty, \quad (41)$$

where  $h_k$  and  $s_{k+1}$  are as in (35) and (36), respectively. Using (36) and the second statement in (41) we also obtain  $\|\tilde{p}_k - \hat{p}_k\|_{\gamma}^2 \rightarrow 0$ , which in turn when combined with the definitions of  $\tilde{p}_k$  and  $\hat{p}_k$  – as in (28) –, (13), (23) and (24) yields

$$v_k - \hat{z}_k \rightarrow 0, \quad Lx_k - \hat{y}_k \rightarrow 0, \quad \tilde{y}_k - Lx_k \rightarrow 0 \quad \text{and} \quad \varepsilon_k \rightarrow 0. \quad (42)$$

Second, from (38) and the fact that  $\alpha_k^2 \leq \alpha_k$ , we obtain

$$\lim_{k \rightarrow \infty} \alpha_k \|z_k - z_{k-1}\| = \lim_{k \rightarrow \infty} \alpha_k \|y_k - y_{k-1}\| = 0,$$

which, in turn, when combined with the definitions of  $\hat{z}_k$  and  $\hat{y}_k$  as in (18) and (19) yields

$$\hat{z}_k - z_k \rightarrow 0 \quad \text{and} \quad \hat{y}_k - y_k \rightarrow 0. \quad (43)$$

Now, let  $p_k = (z_k, y_k)$  be as in (28). Note that using the first statement in (41), the definition of  $h_k$  as in (35) and Lemma A.2 below, it follows that to prove the convergence of  $(p_k)$  to some element in  $\mathcal{S}$  – and hence the statement in (39) – it suffices to show that every cluster point of  $(p_k)$  belongs to  $\mathcal{S}$ . To this end, let  $p_{\infty} = (z_{\infty}, y_{\infty}) \in \mathcal{G}^2$  be a cluster point of  $(p_k)$  (we know from (41) and (35) that  $(p_k)$  is bounded), i.e., let  $z_{\infty}$  and  $y_{\infty}$  be cluster points of  $(z_k)$  and  $(y_k)$ , respectively. Then let also  $(k_j)$  be an increasing sequence of indexes such that

$$z_{k_j} \rightarrow z_{\infty} \quad \text{and} \quad y_{k_j} \rightarrow y_{\infty}. \quad (44)$$

In view of (43) and (44), we have

$$\hat{z}_{k_j} \rightarrow z_{\infty} \quad \text{and} \quad \hat{y}_{k_j} \rightarrow y_{\infty}, \quad (45)$$

which, in particular, when combined with (42) gives

$$v_{k_j} \rightarrow z_{\infty}, \quad Lx_{k_j} \rightarrow y_{\infty}, \quad \tilde{y}_{k_j} \rightarrow y_{\infty} \quad \text{and} \quad \varepsilon_{k_j} \rightarrow 0. \quad (46)$$

From (29), the second statement in (42) (with  $k = k_j$ ) and the first statement in (45) we also obtain  $z'_{k_j} \rightarrow z_{\infty}$ , which combined with Lemma 3.2(b) (with  $k = k_j$ ), the fact that the graph of  $\partial(f^* \circ -L^*)$  is closed and the second statement in (46) yields  $-y_{\infty} \in \partial(f^* \circ -L^*)(z_{\infty})$ . As a consequence, according to the definition of  $\mathcal{S}$  as in (11), to prove that  $(z_{\infty}, y_{\infty}) \in \mathcal{S}$ , it remains to verify that  $y_{\infty} \in \partial g^*(z_{\infty})$ . To this end, recall first that from (22) (with  $k = k_j$ ) we know that  $v_{k_j} \in \partial_{\varepsilon_{k_j}} g(\tilde{y}_{k_j})$ , which is equivalent to  $\tilde{y}_{k_j} \in \partial_{\varepsilon_{k_j}} g^*(v_{k_j})$ . Combining the latter inclusion with the first, third and fourth statements in (46) as well as with the closedness of the graph of the  $\varepsilon$ -subdifferential of  $g^*$ , we obtain the desired result, namely  $y_{\infty} \in \partial g^*(z_{\infty})$ .

Altogether, we have proved that every cluster point of  $(p_k)$  belongs to  $\mathcal{S}$  and so, as we explained above, it guarantees that  $(p_k)$  converges to some element in  $\mathcal{S}$ , i.e., here we finish the proof of (39).

Finally, the proof of (40) follows trivially from (39), (42) and (43).  $\square$

**Remark:** As we discussed in the Introduction (following Assumption 2.1), under standard regularity conditions on (1), the result on  $(z_\infty, w_\infty)$  as in Theorem 4.2, gives that there exists  $x_\infty \in \mathcal{H}$  such that  $x_\infty \in \partial f^*(-L^*z_\infty)$ ,  $w_\infty = Lx_\infty$  and  $x_\infty$  and  $z_\infty$  are solutions of (9) and (10), respectively. Moreover, the second statements in (39) and (40) give, in particular, that  $Lx_k - y_k \rightarrow 0$ .

We will consider the following two sufficient conditions on the sequences  $(\alpha_k)$  and/or  $(p_k)$  to ensure (38) holds – see (12) and the definition of  $p_k$  as in (28) – :

**Assumption A:** for some  $0 < \theta < 1$  and  $k_0 \geq 1$ ,

$$\alpha_k \leq \min \left\{ \alpha, \frac{\theta^k}{\gamma^{-1}\|z_k - z_{k-1}\|^2 + \gamma\|y_k - y_{k-1}\|^2} \right\}, \quad \forall k \geq k_0; \quad (47)$$

here we adopt the convention  $1/0 = \infty$ .

**Assumption B:**  $(\alpha, \sigma, \tau) \in [0, 1) \times [0, 1) \times (0, 1)$  and the sequence  $(\alpha_k)$  satisfy

$$0 \leq \alpha_k \leq \alpha_{k+1} \leq \alpha < \beta < 1 \quad \forall k \geq 0, \quad (48)$$

where

$$\beta := \frac{2\eta}{1 + 2\eta + \sqrt{1 + 8\eta}} \quad (49)$$

and

$$\eta := \frac{(1 - \tau)(1 - \sigma)^2}{4\tau}. \quad (50)$$

**Lemma 4.3.** *Under the Assumption B on Algorithm 1, define the quadratic real function*

$$q(t) := (\eta - 1)t^2 - (1 + 2\eta)t + \eta \quad \forall t \in \mathbb{R}. \quad (51)$$

*Then,  $q(\alpha) > 0$  and, for every  $p = (z, w) \in \mathcal{S}$ ,*

$$\sum_{j=0}^k \|p_j - p_{j-1}\|_\gamma^2 \leq \frac{2\|p_0 - p\|_\gamma^2}{(1 - \alpha)q(\alpha)} \quad \forall k \geq 1. \quad (52)$$

*Proof.* Note first that combining Lemmas 4.1 and 3.4(e) (second inequality) and (50) we obtain, for all  $k \geq 0$ ,

$$h_{k+1} - h_k - \alpha_k(h_k - h_{k-1}) + \eta\|p_{k+1} - \hat{p}_k\|_\gamma^2 \leq \alpha_k(1 + \alpha_k)\|p_k - p_{k-1}\|_\gamma^2, \quad (53)$$

where  $h_k$  is as in (35). On the other hand, using Lemma 3.2(e), the Cauchy-Schwarz inequality and the Young inequality  $2ab \leq a^2 + b^2$  with  $a := \|p_{k+1} - p_k\|_\gamma$  and  $b := \|p_k - p_{k-1}\|_\gamma$  we find

$$\begin{aligned} \|p_{k+1} - \hat{p}_k\|_\gamma^2 &= \|p_{k+1} - p_k\|_\gamma^2 + \alpha_k^2\|p_k - p_{k-1}\|_\gamma^2 - 2\alpha_k\langle p_{k+1} - p_k | p_k - p_{k-1} \rangle_\gamma \\ &\geq \|p_{k+1} - p_k\|_\gamma^2 + \alpha_k^2\|p_k - p_{k-1}\|_\gamma^2 - \alpha_k(2\|p_{k+1} - p_k\|_\gamma\|p_k - p_{k-1}\|_\gamma) \\ &\geq (1 - \alpha_k)\|p_{k+1} - p_k\|_\gamma^2 - \alpha_k(1 - \alpha_k)\|p_k - p_{k-1}\|_\gamma^2. \end{aligned} \quad (54)$$

Using (53), (54) and some simple algebraic manipulations we find

$$h_{k+1} - h_k - \alpha_k(h_k - h_{k-1}) + \eta(1 - \alpha_k)\|p_{k+1} - p_k\|_\gamma^2 \leq \gamma_k\|p_k - p_{k-1}\|_\gamma^2, \quad (55)$$

where, for all  $k \geq 0$ ,

$$\gamma_k := (1 - \eta)\alpha_k^2 + (1 + \eta)\alpha_k. \quad (56)$$

Define

$$\mu_0 := (1 - \alpha_0)h_0 \geq 0 \quad \text{and} \quad \mu_k := h_k - \alpha_{k-1}h_{k-1} + \gamma_k\|p_k - p_{k-1}\|_\gamma^2 \quad \forall k \geq 1, \quad (57)$$

where  $h_k$  is as in (35). Using (51), the assumption that  $(\alpha_k)$  is nondecreasing – see (48) – and (55)–(57) we obtain, for all  $k \geq 1$ ,

$$\begin{aligned} \mu_k - \mu_{k-1} &\leq [h_k - h_{k-1} - \alpha_{k-1}(h_{k-1} - h_{k-2}) - \gamma_{k-1}\|p_{k-1} - p_{k-2}\|_\gamma^2] + \gamma_k\|p_k - p_{k-1}\|_\gamma^2 \\ &\leq [\gamma_k - \eta(1 - \alpha_k)]\|p_k - p_{k-1}\|_\gamma^2 \\ &= -[(\eta - 1)\alpha_k^2 - (1 + 2\eta)\alpha_k + \eta]\|p_k - p_{k-1}\|_\gamma^2 \\ &= -q(\alpha_k)\|p_k - p_{k-1}\|_\gamma^2. \end{aligned} \quad (58)$$

Note now that  $0 < \beta < 1$  as in (49) is either the smallest or the largest root of the quadratic function  $q(\cdot)$ . Hence, from (48), for all  $k \geq 0$ ,

$$q(\alpha_k) \geq q(\alpha) > q(\beta) = 0.$$

The above inequalities combined with (58) yield

$$\|p_k - p_{k-1}\|_\gamma^2 \leq \frac{1}{q(\alpha)}(\mu_{k-1} - \mu_k), \quad \forall k \geq 1, \quad (59)$$

which combined with (48) and the definition of  $\mu_k$  as in (57) gives

$$\begin{aligned} \sum_{j=0}^k \|p_j - p_{j-1}\|_\gamma^2 &\leq \frac{1}{q(\alpha)}(\mu_0 - \mu_k), \\ &\leq \frac{1}{q(\alpha)}(\mu_0 + \alpha h_{k-1}) \quad \forall k \geq 1. \end{aligned} \quad (60)$$

Note now that using (48), (57) and (59) we also find

$$\begin{aligned} \mu_0 &\geq \dots \geq \mu_k = h_k - \alpha_{k-1}h_{k-1} + \gamma_k\|p_k - p_{k-1}\|_\gamma^2 \\ &\geq h_k - \alpha h_{k-1}, \quad \forall k \geq 1, \end{aligned}$$

and so

$$h_k \leq \alpha^k h_0 + \frac{\mu_0}{1 - \alpha} \leq h_0 + \frac{\mu_0}{1 - \alpha} \quad \forall k \geq 0. \quad (61)$$

Hence, (52) follows directly from (60), (61), the definition of  $\mu_0$  as in (57) and the definition of  $h_0$  as in (35).  $\square$

**Theorem 4.4** (Second result on the asymptotic convergence of Algorithm 1). *Under the assumptions A or B on the sequence  $(\alpha_k)$ , all the conclusions of Theorem 4.2 hold true.*

*Proof.* The proof follows from Theorem 4.2 (see (38)), Assumptions A and B above and Lemma 4.3.  $\square$

## 5 Global convergence rates of Algorithm 1

In this section, we study global convergence rates for Algorithm 1. We obtain (global) pointwise  $O(1/\sqrt{k})$  and ergodic  $O(1/k)$  rates for residuals; see Theorems 5.2 and 5.4 below.

**Lemma 5.1.** *Consider the sequences evolved by Algorithm 1 and assume that*

Assumption B *holds.*

*Let  $(p_k)$ ,  $(\tilde{p}_k)$  and  $(\hat{p}_k)$  be as in (28) and let also  $q(\cdot)$  be as in (51). Then, for all  $p = (z, w) \in \mathcal{S}$ ,*

$$\|p_k - p\|_\gamma^2 + \tau(1 - \tau)(1 - \sigma)^2 \sum_{j=0}^k \|\tilde{p}_j - \hat{p}_j\|_\gamma^2 \leq \left(1 + \frac{2\alpha(1 + \alpha)}{(1 - \alpha)^2 q(\alpha)}\right) \|p_0 - p\|_\gamma^2.$$

*Proof.* From Lemmas 4.1 and A.1(a) (below),

$$h_k + \sum_{j=1}^k s_j \leq h_0 + \frac{1}{1 - \alpha} \sum_{j=0}^{k-1} \delta_j,$$

where  $h_k$ ,  $s_k$  and  $\delta_k$  are as in (35), (36) and (37), respectively. Then, in view of (52),

$$\begin{aligned} \|p_k - p\|_\gamma^2 + \tau(1 - \tau)(1 - \sigma)^2 \sum_{j=0}^k \|\tilde{p}_j - \hat{p}_j\|_\gamma^2 &\leq \|p_0 - p\|_\gamma^2 + \frac{1}{1 - \alpha} \sum_{j=0}^{k-1} \alpha_j(1 + \alpha_j) \|p_j - p_{j-1}\|_\gamma^2 \\ &\leq \left(1 + \frac{2\alpha(1 + \alpha)}{(1 - \alpha)^2 q(\alpha)}\right) \|p_0 - p\|_\gamma^2. \end{aligned}$$

□

**Theorem 5.2** (Pointwise global convergence rates of Algorithm 1). *Consider the sequences evolved by Algorithm 1 and assume that*

Assumption B *holds.*

*Let  $(z'_k)$  be as in (29) and let  $d_0$  denote the distance of  $p_0 = (z_0, y_0)$  to the solution set  $\mathcal{S}$  as in (11). Then, for every  $k \geq 0$ , there exists  $0 \leq i \leq k$  such that*

$$\left\{ \begin{array}{l} -Lx_i \in \partial(f^* \circ -L^*)(z'_i), \quad \tilde{y}_i \in \partial_{\varepsilon_i} g^*(v_i), \\ \gamma \|Lx_i - \tilde{y}_i\|^2 + \frac{1}{\gamma} \|z'_i - v_i\|^2 \leq \frac{2Cd_0^2}{k}, \\ \varepsilon_i \leq \frac{\sigma^2 Cd_0^2}{2k}, \end{array} \right. \quad (62)$$

where

$$C := \frac{1}{\tau(1 - \tau)(1 - \sigma)^2} \left(1 + \frac{2\alpha(1 + \alpha)}{(1 - \alpha)^2 q(\alpha)}\right). \quad (63)$$



*Proof.* Let  $p^* = (z^*, w^*) \in \mathcal{S}$  be such that  $d_0 = \|p_0 - p^*\|_\gamma$ . From Lemma 5.1 (with  $p = p^*$ ) and the definition of  $C > 0$  as in (63),

$$\sum_{j=0}^k \|\tilde{p}_j - \hat{p}_j\|_\gamma^2 \leq C d_0^2. \quad (64)$$

From (13) and Lemmas 3.2(c) and 3.3(b),

$$\frac{\gamma}{2} \|Lx_k - \tilde{y}_k\|^2 + \frac{1}{2\gamma} \|z'_k - v_k\|^2 = \frac{1}{2} \|\tilde{p}_k - \hat{p}_k\|_\gamma^2 \leq \|\tilde{p}_k - \hat{p}_k\|_\gamma^2. \quad (65)$$

Due to (13), (24) and the definitions of  $\tilde{p}_k$  and  $\hat{p}_k$  as in (28) we also find

$$\frac{2\varepsilon_k}{\sigma^2} \leq \gamma \|Lx_k - \hat{y}_k\|^2 + \frac{1}{\gamma} \|v_k - \hat{z}_k\|^2 = \|\tilde{p}_k - \hat{p}_k\|_\gamma^2. \quad (66)$$

Hence, from (64) – (66),

$$\sum_{j=0}^k \Delta_j \leq C d_0^2, \quad (67)$$

where

$$\Delta_j := \max \left\{ \frac{\gamma}{2} \|Lx_j - \tilde{y}_j\|^2 + \frac{1}{2\gamma} \|z'_j - v_j\|^2, \frac{2\varepsilon_j}{\sigma^2} \right\}, \quad j = 0, \dots, k. \quad (68)$$

The two inequalities in (62) follow by choosing  $i \in \{0, \dots, k\}$  such that  $\Delta_i \leq \Delta_j$  for all  $j = 0, \dots, k$  and using (67) and the definition of  $\Delta_i$  as in (68). To finish the proof of the theorem, note that the inclusions in (62) follow directly from (22) (combined with the fact that  $(\partial_{\varepsilon_k} g)^{-1} = \partial_{\varepsilon_k} g^*$ ) and Lemma 3.2(b).  $\square$

For the sequences generated by Algorithm 1 and  $(z'_k)$  as in (29), define the ergodic means

$$\begin{aligned} x_k^a &:= \frac{1}{k+1} \sum_{j=0}^k x_j, & \tilde{y}_k^a &:= \frac{1}{k+1} \sum_{j=0}^k \tilde{y}_j, \\ z_k'^a &:= \frac{1}{k+1} \sum_{j=0}^k z'_j, & v_k^a &:= \frac{1}{k+1} \sum_{j=0}^k v_j. \end{aligned} \quad (69)$$

Define also, for all  $k \geq 0$ ,

$$\begin{aligned} \delta_k^a &:= \frac{1}{k+1} \sum_{j=0}^k \langle z'_j \mid L(x_k^a - x_j) \rangle, \\ \varepsilon_k^a &:= \frac{1}{k+1} \sum_{j=0}^k [\varepsilon_j + \langle \tilde{y}_j \mid v_j - v_k^a \rangle]. \end{aligned} \quad (70)$$

**Lemma 5.3.** Consider the sequences evolved by Algorithm 1, let  $(x_k^a)$ ,  $(\tilde{y}_k^a)$ ,  $(z_k'^a)$  and  $(v_k^a)$  be as in (69) and let  $(\delta_k^a)$  and  $(\varepsilon_k^a)$  be as in (70). Let also  $(\check{p}_k)$ ,  $(\hat{p}_k)$  and  $(\tilde{p}_k)$  be as in (28). For all  $k \geq 0$ ,

$$(a) \quad \delta_k^a, \varepsilon_k^a \geq 0 \quad \text{and} \quad -Lx_k^a \in \partial_{\delta_k^a} (f^* \circ -L^*) (z_k'^a), \quad \tilde{y}_k^a \in \partial_{\varepsilon_k^a} g^*(v_k^a).$$

$$(b) \quad \delta_k^a + \varepsilon_k^a = \frac{1}{k+1} \sum_{j=0}^k [\varepsilon_j + \langle \check{p}_j - \hat{p}_j | \tilde{p}_k^a - \tilde{p}_j \rangle_\gamma], \quad \text{where}$$

$$\tilde{p}_k^a := \frac{1}{k+1} \sum_{j=0}^k \tilde{p}_j = (v_k^a, Lx_k^a). \quad (71)$$

*Proof.* (a) The desired result follows from [36, Theorem 2.3] and the inclusions in (22) and in Lemma 3.2(b).

(b) In view of Lemma 3.2(c), for  $j = 0, \dots, k$ , we have  $\check{p}_j - \hat{p}_j = \left( \gamma(Lx_j - \tilde{y}_j), \frac{1}{\gamma} (z_j' - v_j) \right)$  and so by using the definition of  $(\tilde{p}_j)$  and (71) we get

$$\begin{aligned} \sum_{j=0}^k \langle \check{p}_j - \hat{p}_j | \tilde{p}_k^a - \tilde{p}_j \rangle_\gamma &= \sum_{j=0}^k \left[ \frac{1}{\gamma} \langle \gamma(Lx_j - \tilde{y}_j) | v_k^a - v_j \rangle + \gamma \langle \frac{1}{\gamma} (z_j' - v_j) | Lx_k^a - Lx_j \rangle \right] \\ &= \sum_{j=0}^k [\langle Lx_j - \tilde{y}_j | v_k^a - v_j \rangle + \langle z_j' - v_j | L(x_k^a - x_j) \rangle] \\ &= \sum_{j=0}^k [\langle \tilde{y}_j | v_j - v_k^a \rangle + \langle Lx_j | v_k^a \rangle + \langle z_j' | L(x_k^a - x_j) \rangle - \langle v_j | Lx_k^a \rangle] \\ &= \sum_{j=0}^k [\langle \tilde{y}_j | v_j - v_k^a \rangle + \langle z_j' | L(x_k^a - x_j) \rangle]. \end{aligned}$$

The desired result now follows by adding the two equations in (70) and using the latter identity.  $\square$

**Theorem 5.4** (Ergodic global convergence rates of Algorithm 1). Consider the sequences evolved by Algorithm 1, let  $(x_k^a)$ ,  $(\tilde{y}_k^a)$ ,  $(z_k'^a)$  and  $(v_k^a)$  be as in (69) and let  $(\delta_k^a)$  and  $(\varepsilon_k^a)$  be as in (70). Let also  $d_0$  denote the distance of  $p_0 = (z_0, y_0)$  to the solution set  $\mathcal{S}$  as in (11). Assume that  $\alpha_k \equiv \alpha$  and that

Assumption B holds.

Then, for all  $k \geq 0$ ,  $\delta_k^a, \varepsilon_k^a \geq 0$  and

$$\left\{ \begin{array}{l} -Lx_k^a \in \partial_{\delta_k^a} (f^* \circ -L^*) (z_k'^a), \quad \tilde{y}_k^a \in \partial_{\varepsilon_k^a} g^*(v_k^a), \\ \gamma \|Lx_k^a - \tilde{y}_k^a\|_\gamma^2 + \frac{1}{\gamma} \|z_k'^a - v_k^a\|_\gamma^2 \leq \frac{D^2 d_0^2}{k^2}, \\ \delta_k^a + \varepsilon_k^a \leq \frac{1}{k} \left( \frac{\alpha(1+\alpha)}{\tau(1-\alpha)q(\alpha)} + D(1+2\sqrt{3})\sqrt{C} \right) d_0^2, \end{array} \right. \quad (72)$$

where  $C > 0$  is as in (63) and

$$D := \frac{1+\alpha}{\tau} \left( 1 + \sqrt{1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}} \right). \quad (73)$$

*Proof.* Note first that the inclusions in (72) follow from Lemma 5.3(a). Now let  $p^* = (z^*, w^*) \in \mathcal{S}$  be such that  $d_0 = \|p_0 - p^*\|_\gamma$ . Using Lemma 3.2[(a) and (e)] and the assumption  $\alpha_k \equiv \alpha$  we find  $\tau(\widehat{p}_k - \check{p}_k) = p_k - p_{k+1} + \alpha(p_k - p_{k-1})$ , for all  $k \geq 0$ , and so (recall that  $p_0 = p_{-1}$ )

$$\tau \left\| \sum_{j=0}^k (\widehat{p}_j - \check{p}_j) \right\|_\gamma = \|p_0 - p_{k+1} + \alpha(p_k - p_0)\|_\gamma \leq \|p_0 - p_{k+1}\|_\gamma + \alpha\|p_k - p_0\|_\gamma. \quad (74)$$

In view of Lemma 5.1 (with  $p = p^*$ ), the definitions of  $d_0$  and  $D > 0$ , and the triangle inequality,

$$\begin{aligned} \|p_k - p_0\|_\gamma &\leq \|p_k - p^*\|_\gamma + \|p^* - p_0\|_\gamma \\ &\leq \left( 1 + \sqrt{1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}} \right) d_0 \\ &= \frac{\tau D d_0}{1+\alpha} \quad \forall k \geq 0, \end{aligned} \quad (75)$$

which combined with (73) and (74) yields

$$\left\| \sum_{j=0}^k (\widehat{p}_j - \check{p}_j) \right\|_\gamma \leq D d_0. \quad (76)$$

Recall that from Lemma 3.2(c) we have

$$\check{p}_k - \widehat{p}_k = \left( \gamma(Lx_k - \widetilde{y}_k), \frac{1}{\gamma}(z'_k - v_k) \right),$$

and so from the definitions of ergodic means as in (69), we find

$$\frac{1}{k+1} \sum_{j=0}^k (\check{p}_j - \widehat{p}_j) = \left( \gamma(Lx_k^a - \widetilde{y}_k^a), \frac{1}{\gamma}(z_k'^a - v_k^a) \right).$$

Hence, in view of (13) and (76),

$$\gamma \|Lx_k^a - \widetilde{y}_k^a\|_\gamma^2 + \frac{1}{\gamma} \|z_k'^a - v_k^a\|_\gamma^2 = \frac{1}{(k+1)^2} \left\| \sum_{j=0}^k (\check{p}_j - \widehat{p}_j) \right\|_\gamma^2 \leq \frac{D^2 d_0^2}{k^2},$$

which gives the first inequality in (72).

Now let's prove the second inequality in (72). To this end, let  $p = (z, w) \in \mathcal{G}^2$  and first note that from Lemma 3.4(b),

$$\|p - \widehat{p}_k\|_\gamma^2 - \|p - \check{p}_k\|_\gamma^2 \geq 2[\varepsilon_k + \langle \check{p}_k - \widehat{p}_k | p - \widetilde{p}_k \rangle_\gamma]. \quad (77)$$

By Lemma 3.2(a) and the identity  $\|(1-\tau)a + \tau b\|_\gamma^2 = (1-\tau)\|a\|_\gamma^2 + \tau\|b\|_\gamma^2 - (1-\tau)\tau\|a-b\|_\gamma^2$  with  $a = p - \widehat{p}_k$  and  $b = p - \check{p}_k$ ,

$$\|p - p_{k+1}\|_\gamma^2 = (1-\tau)\|p - \widehat{p}_k\|_\gamma^2 + \tau\|p - \check{p}_k\|_\gamma^2 - (1-\tau)\tau\|\check{p}_k - \widehat{p}_k\|_\gamma^2.$$

Now multiplying (77) by  $\tau > 0$  and using the latter identity we get

$$\begin{aligned} \|p - \widehat{p}_k\|_\gamma^2 - \|p - p_{k+1}\|_\gamma^2 &\geq 2\tau [\varepsilon_k + \langle \check{p}_k - \widehat{p}_k | p - \widetilde{p}_k \rangle_\gamma] + (1-\tau)\tau\|\check{p}_k - \widehat{p}_k\|_\gamma^2 \\ &\geq 2\tau [\varepsilon_k + \langle \check{p}_k - \widehat{p}_k | p - \widetilde{p}_k \rangle_\gamma]. \end{aligned} \quad (78)$$

Note that from Lemma 3.5 the assumption  $\alpha_k \equiv \alpha$  we obtain

$$\|p - \widehat{p}_k\|_\gamma^2 = (1+\alpha)\|p - p_k\|_\gamma^2 - \alpha\|p - p_{k-1}\|_\gamma^2 + \alpha(1+\alpha)\|p_k - p_{k-1}\|_\gamma^2. \quad (79)$$

Making the substitution of (79) into (78) and after some simple algebra, we find (now replacing the index  $k \geq 0$  by  $j \geq 0$ ),

$$\begin{aligned} \|p - p_j\|_\gamma^2 - \|p - p_{j+1}\|_\gamma^2 + \alpha(1+\alpha)\|p_j - p_{j-1}\|_\gamma^2 &\geq 2\tau [\varepsilon_j + \langle \check{p}_j - \widehat{p}_j | p - \widetilde{p}_j \rangle_\gamma] \\ &\quad - \alpha [\|p - p_j\|_\gamma^2 - \|p - p_{j-1}\|_\gamma^2]. \end{aligned}$$

Summing up the latter inequality from  $j = 0, \dots, k$  and with  $p = \widetilde{p}_k^a$  — see (71) —, and using Lemma 5.3(b),

$$\begin{aligned} \|\widetilde{p}_k^a - p_0\|_\gamma^2 - \|\widetilde{p}_k^a - p_{k+1}\|_\gamma^2 + \alpha(1+\alpha) \sum_{j=0}^k \|p_j - p_{j-1}\|_\gamma^2 \\ \geq 2\tau \sum_{j=0}^k [\varepsilon_j + \langle \check{p}_j - \widehat{p}_j | \widetilde{p}_k^a - \widetilde{p}_j \rangle_\gamma] - \alpha [\|\widetilde{p}_k^a - p_k\|_\gamma^2 - \|\widetilde{p}_k^a - p_0\|_\gamma^2] \\ = 2\tau(k+1)(\delta_k^a + \varepsilon_k^a) - \alpha [\|\widetilde{p}_k^a - p_k\|_\gamma^2 - \|\widetilde{p}_k^a - p_0\|_\gamma^2], \end{aligned}$$

which combined with (52) (with  $p = p^*$ ) and the definition of  $d_0$  yields

$$\begin{aligned} 2\tau(k+1)(\delta_k^a + \varepsilon_k^a) - \frac{2\alpha(1+\alpha)d_0^2}{(1-\alpha)q(\alpha)} &\leq [\|\widetilde{p}_k^a - p_0\|_\gamma^2 - \|\widetilde{p}_k^a - p_{k+1}\|_\gamma^2] \\ &\quad + \alpha [\|\widetilde{p}_k^a - p_k\|_\gamma^2 - \|\widetilde{p}_k^a - p_0\|_\gamma^2]. \end{aligned}$$

Now using the inequality  $\|a\|_\gamma^2 - \|b\|_\gamma^2 \leq 2\|a\|_\gamma\|a-b\|_\gamma$  (in both terms in the right-hand side of the latter inequality) and (75) we find

$$\begin{aligned} 2\tau(k+1)(\delta_k^a + \varepsilon_k^a) - \frac{2\alpha(1+\alpha)d_0^2}{(1-\alpha)q(\alpha)} &\leq 2\|\widetilde{p}_k^a - p_0\|_\gamma\|p_{k+1} - p_0\|_\gamma + 2\alpha\|\widetilde{p}_k^a - p_k\|_\gamma\|p_k - p_0\|_\gamma \\ &\leq \frac{2\tau D d_0}{1+\alpha} \left( \|\widetilde{p}_k^a - p_0\|_\gamma + \alpha\|\widetilde{p}_k^a - p_k\|_\gamma \right) \\ &\leq 2\tau D d_0 \max\{\|\widetilde{p}_k^a - p_0\|_\gamma, \|\widetilde{p}_k^a - p_k\|_\gamma\}. \end{aligned} \quad (80)$$

Now define, for all  $k \geq 0$ ,

$$\check{p}_k^a := \frac{1}{k+1} \sum_{j=0}^k \check{p}_j. \quad (81)$$

Using (24), (28), Lemma 3.3(a), (71), (81) as well as the convexity of  $\|\cdot\|_\gamma^2$  we find

$$\begin{aligned} \|\tilde{p}_k^a - \check{p}_k^a\|_\gamma^2 &\leq \frac{1}{k+1} \sum_{j=0}^k \|\tilde{p}_j - \check{p}_j\|_\gamma^2 \\ &= \frac{1}{k+1} \sum_{j=0}^k \left[ \frac{1}{\gamma} \|e_j\|^2 + \frac{1}{\gamma} \|v_j - \hat{z}_j\|^2 \right] \\ &\leq \frac{1}{k+1} \sum_{j=0}^k \left[ \gamma \|Lx_j - \hat{y}_j\|^2 + \frac{1}{\gamma} \|v_j - \hat{z}_j\|^2 \right] \\ &= \frac{1}{k+1} \sum_{j=0}^k \|\tilde{p}_j - \hat{p}_j\|_\gamma^2 \\ &\leq \frac{Cd_0^2}{k+1} \\ &\leq Cd_0^2, \end{aligned} \quad (82)$$

where we used Lemma 5.1 (with  $p = p^*$ ) and the definition of  $C > 0$  as in (63). Similarly, using (81), the triangle inequality and the well-known inequality  $\|a + b\|_\gamma^2 \leq 2(\|a\|_\gamma^2 + \|b\|_\gamma^2)$ , we get, for all  $\ell \geq 1$ ,

$$\begin{aligned} \|p_\ell - \check{p}_k^a\|_\gamma^2 &= \left\| \frac{1}{k+1} \sum_{j=0}^k (p_\ell - \check{p}_j) \right\|_\gamma^2 \\ &\leq \frac{1}{k+1} \sum_{j=0}^k \|p_\ell - \check{p}_j\|_\gamma^2 \\ &\leq \frac{2}{k+1} \sum_{j=0}^k \left( \|p_\ell - p_{j+1}\|_\gamma^2 + \|p_{j+1} - \check{p}_j\|_\gamma^2 \right) \\ &= \frac{2}{k+1} \sum_{j=0}^k \|p_\ell - p_{j+1}\|_\gamma^2 + \frac{2}{k+1} \sum_{j=0}^k \|p_{j+1} - \check{p}_j\|_\gamma^2. \end{aligned} \quad (83)$$

Using (again) Lemma 5.1 (with  $p = p^*$ ) as well as the definition of  $C > 0$  as in (63), we find, for all  $\ell, j \geq 0$ ,

$$\|p_\ell - p_{j+1}\|_\gamma^2 \leq 2 \left( \|p_\ell - p^*\|_\gamma^2 + \|p_{j+1} - p^*\|_\gamma^2 \right) \leq 2\tau(1-\tau)(1-\sigma)^2 Cd_0^2. \quad (84)$$

Recall also that from Lemma 3.2(b) we have  $p_{j+1} - \check{p}_j = (1 - \tau)(\hat{p}_j - \check{p}_j)$  ( $0 \leq j \leq k$ ) and so from Lemma 3.3(a), for all  $j \geq 0$ ,

$$\|p_{j+1} - \check{p}_j\|_\gamma = (1 - \tau)\|\hat{p}_j - \check{p}_j\|_\gamma \leq 2(1 - \tau)\|\tilde{p}_j - \hat{p}_j\|_\gamma,$$

which then yields (by Lemma 5.1 (with  $p = p^*$ ) and the definition of  $C > 0$ )

$$\sum_{j=0}^k \|p_{j+1} - \check{p}_j\|_\gamma^2 = 4(1 - \tau)^2 \sum_{j=0}^k \|\tilde{p}_j - \hat{p}_j\|_\gamma^2 \leq 4(1 - \tau)^2 C d_0^2. \quad (85)$$

Putting it all together, from (83) – (85) we obtain, for all  $\ell \geq 0$ ,

$$\begin{aligned} \|p_\ell - \check{p}_k^a\|_\gamma^2 &\leq 4\tau(1 - \tau)(1 - \sigma)^2 C d_0^2 + \frac{8}{k+1}(1 - \tau)^2 C d_0^2 \\ &= 4(1 - \tau) \left( \tau(1 - \sigma^2) + \frac{2}{k+1}(1 - \tau) \right) C d_0^2 \\ &\leq 12C d_0^2. \end{aligned} \quad (86)$$

Using now the triangle inequality, (82) and (86) we find, for all  $\ell \geq 0$ ,

$$\|p_\ell - \tilde{p}_k^a\|_\gamma \leq \|p_\ell - \check{p}_k^a\|_\gamma + \|\check{p}_k^a - \tilde{p}_k^a\|_\gamma \leq (1 + 2\sqrt{3})\sqrt{C}d_0. \quad (87)$$

Finally, using (80) and (87) (with  $\ell = 0$  and  $\ell = k$ ),

$$2\tau(k+1)(\delta_k^a + \varepsilon_k^a) - \frac{2\alpha(1 + \alpha)d_0^2}{(1 - \alpha)q(\alpha)} \leq 2\tau D(1 + 2\sqrt{3})\sqrt{C}d_0^2$$

so that

$$\delta_k^a + \varepsilon_k^a \leq \frac{1}{k} \left( \frac{\alpha(1 + \alpha)}{\tau(1 - \alpha)q(\alpha)} + D(1 + 2\sqrt{3})\sqrt{C} \right) d_0^2.$$

□

## 6 Numerical Experiments

This section presents some numerical experiments on the LASSO problem, which is an instance of the minimization problem (1). We compared Algorithm 1 from this paper with and without inertial effects; they are called *Inexact ADMM* and *Inexact inertial ADMM*, respectively. We implemented both algorithms in Matlab R2021a and, for both algorithms and all problem classes, used the same stopping criterion, namely

$$\text{dist}_\infty(0, \partial f(x_k) + \partial g(x_k)) \leq \varepsilon, \quad (88)$$

where  $\text{dist}_\infty(0, S) := \inf\{\|s\|_\infty \mid s \in S\}$  and  $\varepsilon$  is a tolerance parameter set to  $10^{-6}$ .

The inertial parameter  $\alpha_k$  (as in step 1 of Algorithm 1) is updated according to the rule (47) with  $\theta = 0.99$  and  $k_0 = 1$ . More precisely, we choose  $\alpha_k$  as

$$\alpha_k = \min \left\{ \alpha, \frac{\theta^k}{\gamma^{-1}\|z_k - z_{k-1}\|^2 + \gamma\|y_k - y_{k-1}\|^2} \right\}, \quad \forall k \geq 1,$$

where  $0 \leq \alpha < 1$ . The source codes are available under request (marina.geremia@ifsc.edu.br).

**The LASSO problem.** We perform numerical experiments on the LASSO problem (as already discussed in (17)), namely

$$\min_{x \in \mathbb{R}^d} \left\{ \frac{1}{2} \|Ax - b\|^2 + \nu \|x\|_1 \right\}, \quad (89)$$

where  $A \in \mathbb{R}^{n \times d}$ ,  $b \in \mathbb{R}^n$  and  $\nu > 0$ . For the data matrix  $A$  and the vector  $b$ , we used five categories of non-artificial datasets (available at the UCI Machine Learning Repository, <https://archive.ics.uci.edu>):

*BlogFeedback*: This category consists of one standard microarray datasets that contain features extracted from a blog post from [15]. This problem is called *blogFeedback* (with  $n = 60021$  and  $d = 280$ ).

*Breast Cancer Wisconsin (Prognostic)*: This category consists of one prognostic Wisconsin breast cancer database from [48], which has a dense matrix  $A$ . Each row represents follow-up data for one breast cancer case. This problem is called *Wisconsin* (with  $n = 198$  and  $d = 33$ ).

*DriveFace*: This category comprises a single standard microarray dataset containing image sequences of individuals driving in real-world scenarios from [34]. This problem is called *DriveFace* (with  $n = 606$  and  $d = 6400$ ) and has a dense matrix  $A$ .

*Gene expression*: This category consists of six standard cancer DNA microarray datasets from [19], which have dense and wide matrices  $A$ , with the number of rows  $n \in [42, 102]$  and the number of columns  $d \in [2000, 6033]$ . These problems are called *brain* (with  $n = 42$  and  $d = 5597$ ), *colon* (with  $n = 62$  and  $d = 2000$ ), *leukemia* (with  $n = 72$  and  $d = 3571$ ), *lymphoma* (with  $n = 62$  and  $d = 4026$ ), *prostate* (with  $n = 102$  and  $d = 6033$ ) and *srbc* (with  $n = 63$  and  $d = 2308$ ).

*Single-Pixel camera*: This category consists of four compressed image sensing datasets from [21], which have dense and wide matrices  $A$ , with  $n \in \{410, 1638\}$  and  $d \in \{1024, 4096\}$ . These problems are called *Ball64\_singlepixcam* (with  $n = 1638$  and  $d = 4096$ ), *Logo64\_singlepixcam* (with  $n = 1638$  and  $d = 4096$ ), *Mug32\_singlepixcam* (with  $n = 410$  and  $d = 1024$ ) and *Mug128\_singlepixcam* (with  $n = 410$  and  $d = 1024$ ).

We implemented both algorithms *Inexact ADMM* and *Inexact inertial ADMM* in Matlab R2021a, combined with a CG procedure to approximately solve the subproblems (21); see also the fourth remark following Definition 3.1. As usual (see, e.g., [3]), we solved the (easy) subproblem (20) by using the standard-soft thresholding operator. We also set  $(\alpha, \sigma, \tau, \gamma) = (0.33, 0.99, 0.999, 1)$  and  $(\sigma, \tau, \gamma) = (0.99, 0.999, 1)$  for *Inexact inertial ADMM* and *Inexact ADMM*, respectively. Moreover, as in [13], we set the regularization parameter  $\nu$  as  $0.1 \|A^T b\|_\infty$ , and scaled the vector  $b$  and the columns of matrix  $A$  to have  $\ell_2$  unit norm.

Table 1 shows the number of outer iterations required by each algorithm on each problem instance, the cumulative total number of inner iterations required by the CG algorithm for solving (21) and runtimes in seconds demanded by each algorithm to achieve the prescribed tolerance as in (88). From Table 1, we see that *Inexact inertial ADMM* outperforms *Inexact ADMM* on average by about 30%, 25% and 25% on “Outer iterations”, “Total inner iterations” and “Runtimes”, respectively.

Table 1: Comparison of performance in the LASSO problem

	Inexact ADMM			Inexact inertial ADMM					
	Outer iterations ( <i>outer1</i> )	Total inner iterations ( <i>inner1</i> )	Time ( <i>time1</i> )	Outer iterations ( <i>outer2</i> )	Total inner iterations ( <i>inner2</i> )	Time ( <i>time2</i> )	$\frac{outer2}{outer1}$	$\frac{inner2}{inner1}$	$\frac{time2}{time1}$
Brain	2923	14007	3.97	2120	11112	3.11	0.7253	0.7933	0.7834
Colon	505	2818	0.39	347	1866	0.27	0.6871	0.6622	0.6923
Leukemia	764	3695	0.83	544	2861	0.62	0.7121	0.7743	0.7469
Lymphoma	1101	6091	1.45	862	5119	1.13	0.7829	0.8404	0.7793
Prostate	2006	8328	4.13	1414	6561	3.46	0.7049	0.7878	0.8378
Srbct	511	3554	0.55	346	2293	0.38	0.6771	0.6452	0.6909
Ball64	313	490	8.77	216	325	5.93	0.6901	0.6633	0.6762
Logo64	316	495	8.47	221	359	5.98	0.6994	0.7253	0.7061
Mug32	134	282	0.09	88	197	0.06	0.6567	0.6986	0.6667
Mug128	955	1163	248.04	822	986	212.21	0.8607	0.8478	0.8555
DrivFace	2682	18727	165.47	1803	13284	115.36	0.6723	0.7094	0.6972
Wisconsin	285	605	0.07	182	456	0.05	0.6386	0.7537	0.7143
blogFeedback	386	1108	46.79	317	938	38.84	0.8212	0.8466	0.8301
Geometric mean	662.68	2187.47	3.21	473.79	1633.23	2.38	0.7149	0.7466	0.7414

## 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  $h_0 = h_{-1}$ ,  $0 \leq \alpha_k \leq \alpha < 1$  and*

$$h_{k+1} - h_k + s_{k+1} \leq \alpha_k(h_k - h_{k-1}) + \delta_k \quad \forall k \geq 0. \quad (90)$$

*The following hold:*

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

$$h_k + \sum_{j=1}^k s_j \leq h_0 + \frac{1}{1-\alpha} \sum_{j=0}^{k-1} \delta_j. \quad (91)$$

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

**Lemma A.2** (Opial [40]). *Let  $\mathcal{H}$  be a finite dimensional inner product space, let  $\emptyset \neq \mathcal{S} \subset \mathcal{H}$  and let  $\{p_k\}$  be a sequence in  $\mathcal{H}$  such that every cluster point of  $\{p_k\}$  belongs to  $\mathcal{S}$  and  $\lim_{k \rightarrow \infty} \|p_k - p\|$  exists for every  $p \in \mathcal{S}$ . Then  $\{p_k\}$  converges to a point in  $\mathcal{S}$ .*

## References

- [1] V. A. Adona, M. L. N. Gonçalves, and J. G. Melo. A partially inexact proximal alternating direction method of multipliers and its iteration-complexity analysis. *J. Optim. Theory Appl.*, 182(2):640–666, 2019.



- [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.
- [3] M. M. Alves, J. Eckstein, M. Geremia, and J.G. Melo. Relative-error inertial-relaxed inexact versions of Douglas-Rachford and ADMM splitting algorithms. *Comput. Optim. Appl.*, 75(2):389–422, 2020.
- [4] M. M. 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 Var. Anal.*, 28(2):301–325, 2020.
- [5] H. Attouch. Fast inertial proximal ADMM algorithms for convex structured optimization with linear constraint. *Minimax Theory Appl.*, 6(1):1–24, 2021.
- [6] H. Attouch and A. Cabot. Convergence of a relaxed inertial proximal algorithm for maximally monotone operators. *Math. Program.*, 184(1-2, Ser. A):243–287, 2020.
- [7] H. Attouch, A. Cabot, Z. Chbani, and H. Riahi. Inertial forward-backward algorithms with perturbations: application to Tikhonov regularization. *J. Optim. Theory Appl.*, 179(1):1–36, 2018.
- [8] 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.
- [9] H. Attouch and M. Soueycatt. Augmented Lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces. Applications to games, PDE’s and control. *Pac. J. Optim.*, 5(1):17–37, 2009.
- [10] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York, 2011.
- [11] M. Benning, F. Knoll, C.-B. Schönlieb, and T. Valkonen. Preconditioned ADMM with nonlinear operator constraint. In Lorena Bociu, Jean-Antoine Désidéri, and Abderrahmane Habbal, editors, *System Modeling and Optimization*, pages 117–126, Cham, 2016. Springer International Publishing.
- [12] R. I. Boţ and E. R. Csetnek. ADMM for monotone operators: convergence analysis and rates. *Adv. Comput. Math.*, 45(1):327–359, 2019.
- [13] 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.
- [14] K. Bredies and H. Sun. A proximal point analysis of the preconditioned alternating direction method of multipliers. *J. Optim. Theory Appl.*, 173(3):878–907, 2017.
- [15] K. Buza. BlogFeedback. *UCI Machine Learning Repository*, 2014.
- [16] L. Chen, X. Li, D. Sun, and K.-C. Toh. On the equivalence of inexact proximal ALM and ADMM for a class of convex composite programming. *Math. Program.*, 185(1-2, Ser. A):111–161, 2021.

- [17] 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.
- [18] W. Deng and W. Yin. On the global and linear convergence of the generalized alternating direction method of multipliers. *J. Sci. Comput.*, 66(3):889–916, 2016.
- [19] M. Dettling and P. Bühlmann. Finding predictive gene groups from microarray data. *J. Multivariate Anal.*, 90(1):106–131, 2004.
- [20] D. Dua and C. Graff. UCI machine learning repository, 2017.
- [21] M. F. Duarte, M. A. Davenport, Dharmpal T., J. N. Laska, Ting S., K. F. Kelly, and R. G. Baraniuk. Single-pixel imaging via compressive sampling: Building simpler, smaller, and less-expensive digital cameras. *IEEE Signal Processing Magazine*, 25(2):83–91, 2008.
- [22] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.*, 55(3):293–318, 1992.
- [23] J. Eckstein and W. Yao. Relative-error approximate versions of Douglas-Rachford splitting and special cases of the ADMM. *Math. Program.*, 170(2, Ser. A):417–444, 2018.
- [24] J. Eckstein and W. Yao. Relative-error approximate versions of Douglas-Rachford splitting and special cases of the ADMM. *Math. Program.*, 170(2):417–444, 2018.
- [25] R.-E. Fan, P.-H. Chen, and C.-J. Lin. Working set selection using second order information for training support vector machines. *Journal of Machine Learning Research*, 6:1889–1918, 12 2005.
- [26] M. Fortin and R. Glowinski. On decomposition-coordination methods using an augmented Lagrangian. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian methods: Applications to the numerical solution of boundary-value problems*, volume 15 of *Studies in Mathematics and its Applications*, pages 97–146. North-Holland, Amsterdam, 1983.
- [27] J. Franklin. The elements of statistical learning: data mining, inference and prediction. *Math. Intelligencer*, 27(2):83–85, 2005.
- [28] D. Gabay. Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian methods: Applications to the numerical solution of boundary-value problems*, volume 15 of *Studies in Mathematics and its Applications*, pages 299–331. North-Holland, Amsterdam, 1983.
- [29] R. Glowinski and A. Marroco. Sur l’approximation, par éléments finis d’ordre 1, et la résolution, par pénalisation-dualité, d’une classe de problèmes de Dirichlet non linéaires. *C. R. Acad. Sci. Paris Sér. A*, 278:1649–1652, 1974.
- [30] R. Glowinski, S. J. Osher, and W. Yin, editors. *Splitting methods in communication, imaging, science, and engineering*. Scientific Computation. Springer, Cham, 2016.
- [31] I. Guyon, S. Gunn, A. Ben-Hur, and G. Dror. Result analysis of the nips 2003 feature selection challenge. volume 17, 01 2004.

- [32] W. W. Hager and H. Zhang. Convergence rates for an inexact ADMM applied to separable convex optimization. *Comput. Optim. Appl.*, 77(3):729–754, 2020.
- [33] B. He and X. Yuan. On the  $O(1/n)$  convergence rate of the Douglas-Rachford alternating direction method. *SIAM J. Numer. Anal.*, 50(2):700–709, 2012.
- [34] A. Hernández-Sabat, A. López, and K. Díaz-Chito. DrivFace. *UCI Machine Learning Repository*, 2016.
- [35] D. A. Lorenz and Q. Tran-Dinh. Non-stationary Douglas-Rachford and alternating direction method of multipliers: adaptive step-sizes and convergence. *Comput. Optim. Appl.*, 74(1):67–92, 2019.
- [36] R. D. C. Monteiro and B. F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM J. Optim.*, 20(6):2755–2787, 2010.
- [37] R. D. C. Monteiro and B. F. Svaiter. Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. *SIAM J. Optim.*, 23(1):475–507, 2013.
- [38] A. Y. Ng. Feature selection,  $L_1$  vs.  $L_2$  regularization, and rotational invariance. In *Proceedings, Twenty-First International Conference on Machine Learning, ICML*, pages 615–622, 2004.
- [39] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2006.
- [40] Z. Opial. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Amer. Math. Soc.*, 73:591–597, 1967.
- [41] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.
- [42] R. Shefi and M. Teboulle. Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization. *SIAM J. Optim.*, 24(1):269–297, 2014.
- [43] M. V. Solodov and B. F. Svaiter. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal.*, 7(4):323–345, 1999.
- [44] M. V. Solodov and B. F. Svaiter. A hybrid projection-proximal point algorithm. *J. Convex Anal.*, 6(1):59–70, 1999.
- [45] M. V. Solodov and B. F. Svaiter. A unified framework for some inexact proximal point algorithms. *Numer. Funct. Anal. Optim.*, 22(7-8):1013–1035, 2001.
- [46] B. F. Svaiter. A partially inexact ADMM with  $o(1/n)$  asymptotic convergence rate,  $\mathcal{O}(1/n)$  complexity, and immediate relative error tolerance. *Optimization*, 70(10):2061–2080, 2021.
- [47] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58(1):267–288, 1996.
- [48] W. Wolberg, W. Street, and O. Mangasarian. Breast Cancer Wisconsin (Prognostic). *UCI Machine Learning Repository*, 1995.

- [49] J. Xie. On inexact ADMMs with relative error criteria. *Comput. Optim. Appl.*, 71(3):743–765, 2018.
- [50] J. Xie, A. Liao, and X. Yang. An inexact alternating direction method of multipliers with relative error criteria. *Optim. Lett.*, 11(3):583–596, 2017.
- [51] C. Zălinescu. *Convex analysis in general vector spaces*. World Scientific Publishing Co., Inc., River Edge, NJ, 2002.