

An inexact version of the symmetric proximal ADMM for solving separable convex optimization

Vando A. Adona * Max L.N. Gonçalves *

June 4, 2020

Abstract

In this paper, we propose and analyze an inexact version of the symmetric proximal alternating direction method of multipliers (ADMM) for solving linearly constrained optimization problems. Basically, the method allows its first subproblem to be solved inexactly in such way that a relative approximate criterion is satisfied. In terms of the iteration number k , we establish global $\mathcal{O}(1/\sqrt{k})$ pointwise and $\mathcal{O}(1/k)$ ergodic convergence rates of the method for a domain of the acceleration parameters, which is consistent with the largest known one in the exact case. Since the symmetric proximal ADMM can be seen as a class of ADMM variants, the new algorithm as well as its convergence rates generalize, in particular, many others in the literature. Numerical experiments illustrating the practical advantages of the method are reported. To the best of our knowledge, this work is the first one to study an inexact version of the symmetric proximal ADMM.

Key words: Symmetric alternating direction method of multipliers, convex program, relative error criterion, pointwise iteration-complexity, ergodic iteration-complexity.

1 Introduction

Throughout this paper, \mathbb{R} , \mathbb{R}^n and $\mathbb{R}^{n \times p}$ denote, respectively, the set of real numbers, the set of n dimensional real column vectors and the set of $n \times p$ real matrices. For any vectors $x, y \in \mathbb{R}^n$, $\langle x, y \rangle$ stands for their inner product and $\|x\| := \sqrt{\langle x, x \rangle}$ stands for the Euclidean norm of x . The space of symmetric positive semidefinite (resp. definite) matrices on $\mathbb{R}^{n \times n}$ is denoted by \mathbb{S}_+^n (resp. \mathbb{S}_{++}^n). Each element $Q \in \mathbb{S}_+^n$ induces a symmetric bilinear form $\langle Q(\cdot), \cdot \rangle$ on $\mathbb{R}^n \times \mathbb{R}^n$ and a seminorm $\|\cdot\|_Q := \sqrt{\langle Q(\cdot), \cdot \rangle}$ on \mathbb{R}^n . The trace and determinant of a matrix P are denoted by $\text{Tr}(P)$ and $\det(P)$, respectively. We use I and $\mathbf{0}$ to stand, respectively, for the identity matrix and the zero matrix with proper dimension throughout the context.

Consider the separable linearly constrained optimization problem

$$\min\{f(x) + g(y) : Ax + By = b\}, \quad (1)$$

where $f: \mathbb{R}^n \rightarrow (-\infty, \infty]$ and $g: \mathbb{R}^p \rightarrow (-\infty, \infty]$ are proper, closed and convex functions, $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{m \times p}$, and $b \in \mathbb{R}^m$. Convex optimization problems with a separable structure such

*IME, Universidade Federal de Goias, Campus II- Caixa Postal 131, CEP 74001-970, Goiânia-GO, Brazil. (E-mails: vandoadona@gmail.com and maxlmg@ufg.br). The work of these authors was supported in part by FAPEG/GO Grant PRONEM-201710267000532, and CNPq Grants 302666/2017-6 and 408123/2018-4.

as (1) appear in many applications areas such as machine learning, compressive sensing and image processing. The augmented Lagrangian method (see, e.g., [7]) attempts to solve (1) directly without taking into account its particular structure. To overcome this drawback, a variant of the augmented Lagrangian method, namely, the alternating direction method of multipliers (ADMM), was proposed and studied in [16, 19]. The ADMM takes full advantage of the special structure of the problem by considering each variable separably in an alternating form and coupling them into the Lagrange multiplier updating; for detailed reviews, see [8, 18].

More recently, a symmetric version of the ADMM was proposed in [23] and since then it has been studied by many authors (see, for example, [6, 9, 17, 29, 30, 32, 33, 34]). The method with proximal terms added (named here as symmetric proximal ADMM) is described as follows: let an initial point $(x_0, y_0, \gamma_0) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$, a penalty parameter $\beta > 0$, two acceleration parameters τ and θ , and two proximal matrices $G \in \mathbb{S}_+^n$ and $H \in \mathbb{S}_+^m$ be given; for $k = 1, 2, \dots$ do

$$x_k \in \arg \min_{x \in \mathbb{R}^n} \left\{ f(x) - \langle \gamma_{k-1}, Ax \rangle + \frac{\beta}{2} \|Ax + By_{k-1} - b\|^2 + \frac{1}{2} \|x - x_{k-1}\|_G^2 \right\}, \quad (2a)$$

$$\gamma_{k-\frac{1}{2}} := \gamma_{k-1} - \tau\beta (Ax_k + By_{k-1} - b), \quad (2b)$$

$$y_k \in \arg \min_{y \in \mathbb{R}^p} \left\{ g(y) - \langle \gamma_{k+\frac{1}{2}}, By \rangle + \frac{\beta}{2} \|Ax_k + By - b\|^2 + \frac{1}{2} \|y - y_{k-1}\|_H^2 \right\}, \quad (2c)$$

$$\gamma_k := \gamma_{k-1} - \theta\beta (Ax_k + By_k - b). \quad (2d)$$

The symmetric proximal ADMM unifies several ADMM variants. For example, it reduces to:

- the standard ADMM when $G = \mathbf{0}$, $H = \mathbf{0}$, $\tau = 0$ and $\theta = 1$;
- the Fortin and Glowinski acceleration version of the proximal ADMM (for short FG-P-ADMM) when $\tau = 0$; see [14, 15, 21];
- the generalized proximal ADMM (for short G-P-ADMM) with the relaxation factor $\alpha := \tau + 1$ when $\theta = 1$; see [1, 11]. The proof of the latter fact can be found, for example, in [23, Remark 5.8];
- the strictly contractive Peaceman–Rachford splitting method studied in [22] when $\tau = \theta \in (0, 1)$. It is worth pointing out that if $\tau = \theta = 1$, $G = \mathbf{0}$ and $H = \mathbf{0}$, the symmetric proximal ADMM corresponds to the standard Peaceman-Rachford splitting method applied to the dual of (1).

As has been observed by some authors (see, e.g., [1, 6, 23]), the use of suitable acceleration parameters τ and θ considerably improves the numerical performances of the ADMM-type algorithms. We also mention that the proximal terms $\|x - x_{k-1}\|_G^2/2$ and $\|y - y_{k-1}\|_H^2/2$ in subproblems (2a) and (2c), respectively, can make them easier to solve or even to have closed-form solution in some applications; see, e.g., [5, 10, 24] for discussion.

In order to ensure the convergence of the symmetric ADMM in [23], the parameters τ and θ were considered into the domain

$$\mathcal{D} := \left\{ (\tau, \theta) \mid \tau \in (-1, 1), \theta \in (0, (1 + \sqrt{5})/2), \tau + \theta > 0, |\tau| < 1 + \theta - \theta^2 \right\}.$$

Later, in the multi-block symmetric ADMM setting, the authors in [6] have extended this convergence domain to

$$\mathcal{K} := \left\{ (\tau, \theta) \mid \tau \leq 1, \tau + \theta > 0, 1 + \tau + \theta - \tau\theta - \tau^2 - \theta^2 > 0 \right\}, \quad (3)$$

by using appropriate proximal terms and by assuming that the matrices associated to the respective multi-block problem have full column rank. Note that, if $\tau = 0$ (resp. $\theta = 1$), the convergence domains in the above regions are equivalent to the classical condition $\theta \in (0, (1 + \sqrt{5})/2)$ (resp. $\tau \in (-1, 1)$) or, in terms of the relaxation factor α , $\alpha \in (0, 2)$ in the FG-P-ADMM (resp. G-P-ADMM). We refer the reader to [1, 21] for some complexity and numerical results of the G-P-ADMM and FG-P-ADMM.

It is well-known that implementations of the ADMM in some applications may be expensive and difficult due to the necessity to solve exactly its two subproblems. For applications in which one subproblem of the ADMM is significantly more challenging to solve than the other one, being necessary, therefore, to use iterative methods to approximately solve it, papers [2] and [3] proposed partially inexact versions of the FG-P-ADMM and G-P-ADMM, respectively, using relative error conditions. Essentially, the proposed schemes allow an inexact solution $\tilde{x}_k \in \mathbb{R}^n$ with residual $u_k \in \mathbb{R}^n$ of subproblem (2a) with $G = I/\beta$, i.e.,

$$u_k \in \partial f(\tilde{x}_k) - A^* \tilde{\gamma}_k, \quad (4)$$

such that the relative error condition

$$\|\tilde{x}_k - x_{k-1} + \beta u_k\|^2 \leq \tilde{\sigma} \|\tilde{\gamma}_k - \gamma_{k-1}\|^2 + \hat{\sigma} \|\tilde{x}_k - x_{k-1}\|^2, \quad (5)$$

is satisfied, where

$$\tilde{\gamma}_k := \gamma_{k-1} - \beta(A\tilde{x}_k + By_{k-1} - b), \quad x_k := x_{k-1} - \beta u_k,$$

and $\tilde{\sigma}$ and $\hat{\sigma}$ are two error tolerance parameters. Recall that the ε -subdifferential of a convex function $h : \mathbb{R}^n \rightarrow \mathbb{R}$ is defined by

$$\partial_\varepsilon h(x) := \{u \in \mathbb{R}^n : h(\tilde{x}) \geq h(x) + \langle u, \tilde{x} - x \rangle - \varepsilon, \quad \forall \tilde{x} \in \mathbb{R}^n\}, \quad \forall x \in \mathbb{R}^n.$$

When $\varepsilon = 0$, then $\partial_0 h(x)$ is denoted by $\partial h(x)$ and is called the subdifferential of f at x . Note that the inclusion in (4) is based on the first-order optimality condition for (2a). For the inexact FG-P-ADMM in [2], the domain of the acceleration factor θ was

$$\theta \in \left(0, \frac{1 - 2\tilde{\sigma} + \sqrt{(1 - 2\tilde{\sigma})^2 + 4(1 - \tilde{\sigma})}}{2(1 - \tilde{\sigma})}\right), \quad (6)$$

whereas, for the inexact G-P-ADMM in [3], the domain of the acceleration factor τ was

$$\tau \in (-1, 1 - \tilde{\sigma}) \quad (\text{or, in term of the relaxation factor } \alpha, \alpha \in (0, 2 - \tilde{\sigma})). \quad (7)$$

If $\tilde{\sigma} = 0$, then (6) and (7) reduce, respectively, to the standard conditions $\theta \in (0, (1 + \sqrt{5})/2)$ and $\tau \in (-1, 1)$. Other inexact ADMMs with relative and/or absolute error condition were proposed in [4, 12, 13, 25, 35]. It is worth pointing out that, as observed in [12], approximation criteria based on relative error are more interesting from a computational viewpoint than those based on absolute error.

Therefore, the main goal of this work is to present an inexact version of the symmetric proximal ADMM (2) in which, similarly to [2, 3], the solution of the first subproblem can be computed in an approximate way such that a relative error condition is satisfied. From the theoretical viewpoint, the global $\mathcal{O}(1/\sqrt{k})$ pointwise convergence rate is shown, which ensures, in particular, that for a given

tolerance $\rho > 0$, the algorithm generates a ρ -approximate solution $(\tilde{x}, y, \tilde{\gamma})$ with residual (u, v, w) of the Lagrangian system associated to (1), i.e.,

$$u \in \partial f(\tilde{x}) - A^* \tilde{\gamma}, \quad v \in \partial g(y) - B^* \tilde{\gamma}, \quad w = A\tilde{x} + By - b,$$

and

$$\max\{\|u\|, \|v\|, \|w\|\} \leq \rho,$$

in at most $\mathcal{O}(1/\rho^2)$ iterations. The global $\mathcal{O}(1/k)$ ergodic convergence rate is also established, which implies, in particular, that a ρ -approximate solution $(\tilde{x}^a, y^a, \tilde{\gamma}^a)$ with residuals (u^a, v^a, w^a) and (ε^a, ζ^a) of the Lagrangian system associated to (1), i.e.,

$$u^a \in \partial_{\varepsilon^a} f(\tilde{x}^a) - A^* \tilde{\gamma}^a, \quad v^a \in \partial_{\zeta^a} g(y^a) - B^* \tilde{\gamma}^a, \quad w^a = A\tilde{x}^a + By^a - b,$$

and

$$\max\{\|u^a\|, \|v^a\|, \|w^a\|, \varepsilon^a, \zeta^a\} \leq \rho,$$

is obtained in at most $\mathcal{O}(1/\rho)$ iterations by means of the ergodic sequence. The analysis of the method is established without any assumptions on A and B . Moreover, the new convergence domain of τ and θ reduces to (3), except for the case $\tau = 1$ and $\theta \in (-1, 1)$, in the exact setting (see Remark 2.1(a)). From the applicability viewpoint, we report numerical experiments in order to illustrate the efficiency of the method for solving real-life applications. To the best of our knowledge, this work is the first one to study an inexact version of the symmetric proximal ADMM.

The paper is organized as follows. Section 2 presents the inexact symmetric proximal ADMM as well as its pointwise and ergodic convergence rates. Section 3 is devoted to the numerical study of the proposed method. Some concluding remarks are given in Section 4.

2 Inexact symmetric proximal ADMM

This section describes and investigates an inexact version of the symmetric proximal ADMM for solving (1). Essentially, the method allows its first subproblem to be solved inexactly in such way that a relative error condition is satisfied. In particular, the new algorithm as well as its iteration-complexity results generalize many others in the literature.

We begin by formally stating the inexact algorithm.

Algorithm 1: An inexact symmetric proximal ADMM

Step 0. Let an initial point $(x_0, y_0, \gamma_0) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$, a penalty parameter $\beta > 0$, two error tolerance parameters $\tilde{\sigma}, \hat{\sigma} \in [0, 1)$, and two proximal matrices $G \in \mathbb{S}_{++}^n$ and $H \in \mathbb{S}_{++}^p$ be given. Choose the acceleration parameters τ and θ such that $(\tau, \theta) \in \mathcal{R}_{\tilde{\sigma}}$ where

$$\mathcal{R}_{\tilde{\sigma}} = \left\{ (\tau, \theta) \mid \begin{array}{l} \tau \in (-1, 1 - \tilde{\sigma}), \quad \tau + \theta > 0, \quad \text{and} \\ (1 - \tau^2)(2 - \tau - \theta - \tilde{\sigma}) - (1 - \theta)^2(1 - \tau - \tilde{\sigma}) > 0 \end{array} \right\}, \quad (8)$$

and set $k = 1$.

Step 1. Compute $(\tilde{x}_k, u_k) \in \mathbb{R}^n \times \mathbb{R}^n$ such that

$$u_k \in \partial f(\tilde{x}_k) - A^* \tilde{\gamma}_k, \quad (9)$$

and

$$\|\tilde{x}_k - x_{k-1} + G^{-1}u_k\|_G^2 \leq \frac{\tilde{\sigma}}{\beta} \|\tilde{\gamma}_k - \gamma_{k-1}\|^2 + \hat{\sigma} \|\tilde{x}_k - x_{k-1}\|_G^2, \quad (10)$$

where

$$\tilde{\gamma}_k := \gamma_{k-1} - \beta(A\tilde{x}_k + By_{k-1} - b). \quad (11)$$

Step 2. Set

$$\gamma_{k-\frac{1}{2}} := \gamma_{k-1} - \tau\beta(A\tilde{x}_k + By_{k-1} - b). \quad (12)$$

Step 3. Compute an optimal solution $y_k \in \mathbb{R}^p$ of the subproblem

$$\min_{y \in \mathbb{R}^p} \left\{ g(y) - \langle \gamma_{k-\frac{1}{2}}, By \rangle + \frac{\beta}{2} \|A\tilde{x}_k + By - b\|^2 + \frac{1}{2} \|y - y_{k-1}\|_H^2 \right\}. \quad (13)$$

Step 4. Set

$$x_k := x_{k-1} - G^{-1}u_k, \quad \gamma_k := \gamma_{k-\frac{1}{2}} - \theta\beta(A\tilde{x}_k + By_k - b), \quad (14)$$

and $k \leftarrow k + 1$, and go to step 1.

We now make some relevant comments of our approach.

Remark 2.1. (a) Clearly, it follows from the definition in (8) that if $(\tau, \theta) \in \mathcal{R}_{\tilde{\sigma}}$, then $\theta < 2$, $(1 - \tau - \tilde{\sigma}) > 0$ and $(2 - \tau - \theta - \tilde{\sigma}) > 0$. Moreover, the third condition in (8) can be rewritten as

$$(1 + \tau + \theta - \tau\theta - \tau^2 - \theta^2)(1 - \tau) + (\tau^2 - 2\theta + \theta^2)\tilde{\sigma} > 0.$$

If $\tilde{\sigma} = 0$, then $\tau \in (-1, 1)$. Hence, it follows from the above inequality that \mathcal{R}_0 reduces to the region \mathcal{K} in (3) with $\tau \neq 1$. The regions \mathcal{R}_0 , $\mathcal{R}_{0.3}$ and $\mathcal{R}_{0.6}$ are illustrated in Fig. 1(a), 1(b), and 1(c), respectively. Note that for some suitable choice of $(\tilde{\sigma}, \tau)$, the stepsize θ can be even chosen greater than $(1 + \sqrt{5})/2 \approx 1.618$.

(b) If the inaccuracy parameters $\tilde{\sigma}$ and $\hat{\sigma}$ are zeros, from (10) and the first equality in (14), we obtain $\tilde{x}_k = x_k$ and $u_k = G(x_{k-1} - x_k)$. Hence, in view of the definition of $\tilde{\gamma}_k$ in (11) and the

inclusion in (9), it follows that computing x_k is equivalent to solve exactly the subproblem in (2a). Therefore, we can conclude that Algorithm 1 recovers its exact version.

(c) In order to simplify the updated formula of x_k in (14) and the relative error condition in (10), a trivial choice for the proximal matrix G would be I/β .

(d) If $\tau = 0$, then $(\tau, \theta) \in \mathcal{R}_{\tilde{\sigma}}$ corresponds to

$$\theta \in \left(0, \frac{1 - 2\tilde{\sigma} + \sqrt{(1 - 2\tilde{\sigma})^2 + 4(1 - \tilde{\sigma})}}{2(1 - \tilde{\sigma})}\right), \quad (15)$$

and hence Algorithm 1 with $G = I/\beta$ reduces to the partially inexact proximal ADMM studied in [2]. Note also that if $\tilde{\sigma} = 0$ (exact case), then (15) turns out to be the classical condition $\theta \in (0, (1 + \sqrt{5})/2)$ for the FG-P-ADMM; see [21].

(e) If $\theta = 1$, then $(\tau, \theta) \in \mathcal{R}_{\tilde{\sigma}}$ corresponds to $\tau \in (-1, 1 - \tilde{\sigma})$. By setting $\alpha := \tau + 1$, it is possible to prove (see, e.g., [23, Remark 5.8]) that Algorithm 1 with $G = I/\beta$ reduces to the inexact generalized proximal ADMM in [3]. Furthermore, if $\tilde{\sigma} = 0$, then the condition on τ becomes the standard condition $\tau \in (-1, 1)$ (or, in term of the relaxation factor α , $\alpha \in (0, 2)$) for the G-P-ADMM; see [1].

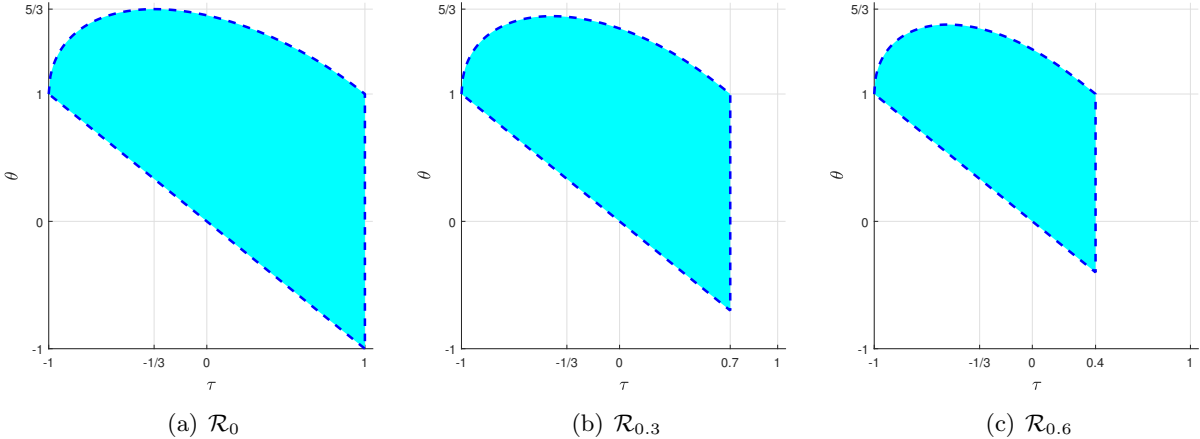


Figure 1: Some instances of $\mathcal{R}_{\tilde{\sigma}}$.

Throughout the paper, we make the following standard assumption.

Assumption 1. There exists a solution $(x^*, y^*, \gamma^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ of the Lagrangian system

$$0 \in \partial f(x) - A^* \gamma, \quad 0 \in \partial g(y) - B^* \gamma, \quad 0 = Ax + By - b, \quad (16)$$

associated to (1).

In order to establish pointwise and ergodic convergence rates for Algorithm 1, we first show in Section 2.1 that the algorithm can be seen as an instance of a general proximal point method. With this fact in hand, we will be able to present convergence rates of Algorithm 1 in Section 2.2. It should be mentioned that the analysis of Algorithm 1 is much more complicated, since it involves two acceleration parameters τ and θ .

2.1 Auxiliar results

Our goal in this section is to show that Algorithm 1 can be seen as an instance of the hybrid proximal extragradient (HPE) framework in [21] (see also [1, 2]). More specifically, it will be proven that there exists a scalar $\sigma \in [\hat{\sigma}, 1)$ such that

$$M(z_{k-1} - z_k) \in T(\tilde{z}_k), \quad \|\tilde{z}_k - z_k\|_M^2 + \eta_k \leq \sigma \|\tilde{z}_k - z_{k-1}\|_M^2 + \eta_{k-1}, \quad \forall k \geq 1, \quad (17)$$

where $z_k := (x_k, y_k, \gamma_k)$ and $\tilde{z}_k := (\tilde{x}_k, y_k, \tilde{\gamma}_k)$, and the matrix M , the operator T and the sequence $\{\eta_k\}$ will be specified later. As a consequence of the latter fact, the pointwise convergence rate to be presented in the next section could be derived from [21, Theorem 3.3]. However, since its proof follows easily from (17), we present it here for completeness and convenience of the reader. On the other hand, although the ergodic convergence rate in the next section is related to [21, Theorem 3.4], its proof does not follow immediately from the latter theorem.

The proof of (17) is extensive and nontrivial. We begin by defining and establishing some properties of the matrix M and the operator T .

Proposition 2.2. *Consider the operator T and the matrix M defined as*

$$T(x, y, \gamma) = \begin{bmatrix} \partial f(x) - A^* \gamma \\ \partial g(y) - B^* \gamma \\ Ax + By - b \end{bmatrix}, \quad M = \begin{bmatrix} G & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & H + \frac{(\tau - \tau\theta + \theta)\beta}{\tau + \theta} B^* B & -\frac{\tau}{\tau + \theta} B^* \\ \mathbf{0} & -\frac{\tau}{\tau + \theta} B & \frac{1}{(\tau + \theta)\beta} I \end{bmatrix}. \quad (18)$$

Then, T is maximal monotone and M is symmetric positive semidefinite.

Proof. Note that T can be decomposed as $T = \tilde{T} + \hat{T}$, where

$$\tilde{T}(z) := (\partial f(x), \partial g(y), -b) \quad \text{and} \quad \hat{T}(z) := Dz, \quad \text{with} \quad D := \begin{bmatrix} \mathbf{0} & \mathbf{0} & -A^* \\ \mathbf{0} & \mathbf{0} & -B^* \\ A & B & \mathbf{0} \end{bmatrix}.$$

Thus, since f and g are convex functions, the operators ∂f and ∂g are maximal monotone (see [27]) and, hence, the operator \tilde{T} is maximal monotone as well. In addition, since D is skew-symmetric, \hat{T} is also maximal monotone. Therefore, we obtain that T is maximal monotone.

Now, it is evident that M is symmetric and, using the inequality of Cauchy-Schwarz, for every $z = (x, y, \gamma) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$

$$\begin{aligned} \langle Mz, z \rangle &= \|x\|_G^2 + \|y\|_H^2 + \frac{(\tau - \tau\theta + \theta)\beta}{\tau + \theta} \|By\|^2 - \frac{2\tau}{\tau + \theta} \langle By, \gamma \rangle + \frac{1}{(\tau + \theta)\beta} \|\gamma\|^2 \\ &\geq \frac{(\tau - \tau\theta + \theta)\beta}{\tau + \theta} \|By\|^2 - \frac{2|\tau|}{\tau + \theta} \|By\| \|\gamma\| + \frac{1}{(\tau + \theta)\beta} \|\gamma\|^2 = \langle Pw, w \rangle, \end{aligned} \quad (19)$$

where $w := (\|\gamma\|, \|By\|)$ and

$$P := \begin{bmatrix} \frac{1}{(\tau + \theta)\beta} & -\frac{|\tau|}{\tau + \theta} \\ -\frac{|\tau|}{\tau + \theta} & \frac{(\tau - \tau\theta + \theta)\beta}{\tau + \theta} \end{bmatrix}.$$

From step 0 of Algorithm 1, we obtain

$$P_{1,1} = \frac{1}{(\tau + \theta)\beta} > 0, \quad \text{and} \quad \det(P) = \frac{(1 - \tau)(\tau + \theta)}{(\tau + \theta)^2} > 0,$$

Therefore, P is symmetric positive definite and, hence, the statement on M follows now from (19). \square

We next establish a technical result.

Lemma 2.3. *Consider the sequences $\{p_k\}$ and $\{q_k\}$ defined by*

$$p_k = B(y_k - y_{k-1}), \quad q_k = -\beta(A\tilde{x}_k + By_k - b), \quad \forall k \geq 1. \quad (20)$$

Then, for every $k \geq 1$, the following equalities hold:

$$\tilde{\gamma}_k - \gamma_{k-1} = \beta p_k + q_k, \quad \tilde{\gamma}_k - \gamma_k = (1 - \tau)\beta p_k + (1 - \tau - \theta)q_k, \quad (21)$$

$$\gamma_k - \gamma_{k-1} = \tau\beta p_k + (\tau + \theta)q_k. \quad (22)$$

Proof. From the definition of $\tilde{\gamma}_k$ in (11), we have

$$\tilde{\gamma}_k - \gamma_{k-1} = \beta B(y_k - y_{k-1}) - \beta(A\tilde{x}_k + By_k - b),$$

which, in view of (20), proves the first identity in (21). Now, using (11), (12) and the definition of γ_k in (14) we get

$$\begin{aligned} \tilde{\gamma}_k - \gamma_k &= \gamma_{k-1} - \gamma_{k-\frac{1}{2}} - \beta(A\tilde{x}_k + By_{k-1} - b) + \theta\beta(A\tilde{x}_k + By_k - b) \\ &= -(1 - \tau)\beta(A\tilde{x}_k + By_{k-1} - b) + \theta\beta(A\tilde{x}_k + By_k - b) \\ &= (1 - \tau)\beta B(y_k - y_{k-1}) - (1 - \tau - \theta)\beta(A\tilde{x}_k + By_k - b). \end{aligned}$$

This equality, together with (20), implies the second identity in (21). Again using the definitions of $\gamma_{k-\frac{1}{2}}$ and γ_k in (12) and (14), respectively, we obtain

$$\begin{aligned} \gamma_k - \gamma_{k-1} &= -\theta\beta(A\tilde{x}_k + By_k - b) - \tau\beta(A\tilde{x}_k + By_{k-1} - b) \\ &= \tau\beta B(y_k - y_{k-1}) - (\tau + \theta)\beta(A\tilde{x}_k + By_k - b), \end{aligned}$$

which, combined with (20), yields (22). \square

We next show that the inclusion in (17) holds.

Theorem 2.4. *For every $k \geq 1$, the following estimatives hold:*

$$G(x_{k-1} - x_k) \in \partial f(\tilde{x}_k) - A^*\tilde{\gamma}_k, \quad (23)$$

$$\left(H + \frac{(\tau - \tau\theta + \theta)\beta}{\tau + \theta}B^*B\right)(y_{k-1} - y_k) - \frac{\tau}{\tau + \theta}B^*(\gamma_{k-1} - \gamma_k) \in \partial g(y_k) - B^*\tilde{\gamma}_k, \quad (24)$$

$$-\frac{\tau}{\tau + \theta}B(y_{k-1} - y_k) + \frac{1}{(\tau + \theta)\beta}(\gamma_{k-1} - \gamma_k) = A\tilde{x}_k + By_k - b. \quad (25)$$

As a consequence, for every $k \geq 1$,

$$M(z_{k-1} - z_k) \in T(\tilde{z}_k),$$

where

$$z_k := (x_k, y_k, \gamma_k) \quad \forall k \geq 0, \quad \tilde{z}_k := (\tilde{x}_k, y_k, \tilde{\gamma}_k) \quad \forall k \geq 1, \quad (26)$$

and T and M are as in (18).

Proof. Inclusion in (23) follows trivially from (9) and the definition of x_k in (14). It follows from (12) and (14) that

$$\begin{aligned}\gamma_k - \gamma_{k-1} &= -\theta\beta(A\tilde{x}_k + By_k - b) - \tau\beta(A\tilde{x}_k + By_{k-1} - b) \\ &= -(\tau + \theta)\beta(A\tilde{x}_k + By_k - b) + \tau\beta B(y_k - y_{k-1}),\end{aligned}$$

which is equivalent to (25). Now, from the optimality condition for (13), we have

$$0 \in \partial g(y_k) - B^* \left[\gamma_{k-\frac{1}{2}} - \beta(A\tilde{x}_k + By_k - b) \right] + H(y_k - y_{k-1}). \quad (27)$$

On the other hand, using (11), we obtain

$$\begin{aligned}\gamma_{k-\frac{1}{2}} - \beta(A\tilde{x}_k + By_k - b) &= \gamma_{k-\frac{1}{2}} - \beta(A\tilde{x}_k + By_{k-1} - b) - \beta B(y_k - y_{k-1}) \\ &= \tilde{\gamma}_k + \gamma_{k-\frac{1}{2}} - \gamma_{k-1} - \beta B(y_k - y_{k-1}).\end{aligned}$$

From the definition of γ_k in (14), we find

$$\begin{aligned}\gamma_{k-\frac{1}{2}} - \gamma_{k-1} &= \gamma_{k-\frac{1}{2}} - \gamma_k + \gamma_k - \gamma_{k-1} = \theta\beta(A\tilde{x}_k + By_k - b) + \gamma_k - \gamma_{k-1} \\ &= \theta\beta \left[\frac{\tau}{\tau + \theta} B(y_k - y_{k-1}) - \frac{1}{(\tau + \theta)\beta} (\gamma_k - \gamma_{k-1}) \right] + \gamma_k - \gamma_{k-1} \\ &= \frac{\tau\theta\beta}{\tau + \theta} B(y_k - y_{k-1}) + \frac{\tau}{\tau + \theta} (\gamma_k - \gamma_{k-1}),\end{aligned}$$

where the last equality is due to (25). Combining the last two equalities, we have

$$\gamma_{k-\frac{1}{2}} - \beta(A\tilde{x}_k + By_k - b) = \tilde{\gamma}_k - \frac{(\tau - \tau\theta + \theta)\beta}{\tau + \theta} B(y_k - y_{k-1}) + \frac{\tau}{\tau + \theta} (\gamma_k - \gamma_{k-1}),$$

which, combined with (27), implies (24). \square

In the remaining part of this section, we will prove that the inequality in (17) holds. Toward this goal, we next establish three technical results.

Lemma 2.5. *Let $\{z_k\}$ and $\{\tilde{z}_k\}$ be as in (26). Then, for every $z^* \in T^{-1}(0)$, we have*

$$\|z^* - z_k\|_M^2 - \|z^* - z_{k-1}\|_M^2 \leq \|\tilde{z}_k - z_k\|_M^2 - \|\tilde{z}_k - z_{k-1}\|_M^2, \quad \forall k \geq 1.$$

Proof. As $M(z_{k-1} - z_k) \in T(\tilde{z}_k)$ (Theorem 2.4), T is monotone maximal (Proposition 2.2) and $0 \in T(z^*)$, we obtain $\langle M(z_{k-1} - z_k), \tilde{z}_k - z^* \rangle \geq 0$. Hence,

$$\begin{aligned}\|z^* - z_k\|_M^2 - \|z^* - z_{k-1}\|_M^2 &= \|z^* - \tilde{z}_k + \tilde{z}_k - z_k\|_M^2 - \|z^* - \tilde{z}_k + \tilde{z}_k - z_{k-1}\|_M^2 \\ &= \|\tilde{z}_k - z_k\|_M^2 + 2\langle M(z_{k-1} - z_k), z^* - \tilde{z}_k \rangle - \|\tilde{z}_k - z_{k-1}\|_M^2 \\ &\leq \|\tilde{z}_k - z_k\|_M^2 - \|\tilde{z}_k - z_{k-1}\|_M^2,\end{aligned}$$

concluding the proof. \square

Proposition 2.6. Define the matrix Q and the scalar ϑ as

$$Q = \begin{bmatrix} (3 - 3\tau - 2\tilde{\sigma})\beta I & 2(1 - \tau - \tilde{\sigma})I \\ 2(1 - \tau - \tilde{\sigma})I & \frac{4 - \tau - \theta - 2\tilde{\sigma}}{\beta}I \end{bmatrix}, \quad (28)$$

and

$$\vartheta = \sqrt{(3 - 3\tau - 2\tilde{\sigma})(4 - \tau - \theta - 2\tilde{\sigma})} - 2(1 - \tau - \tilde{\sigma}). \quad (29)$$

Then, Q is symmetric positive definite and $\vartheta > 0$. Moreover, for any $(y, \gamma) \in \mathbb{R}^p \times \mathbb{R}^m$

$$\|(y, \gamma)\|_Q^2 \geq -2\vartheta \langle y, \gamma \rangle.$$

Proof. Clearly Q is symmetric, and is positive definite iff

$$\widehat{Q} = \begin{bmatrix} (3 - 3\tau - 2\tilde{\sigma})\beta & 2(1 - \tau - \tilde{\sigma}) \\ 2(1 - \tau - \tilde{\sigma}) & \frac{4 - \tau - \theta - 2\tilde{\sigma}}{\beta} \end{bmatrix}$$

is positive definite. To show that $\widehat{Q} \in \mathbb{S}_{++}^2$ consider the scalars ϱ , $\tilde{\varrho}$, and $\hat{\varrho}$ defined by

$$\varrho = (3 - 3\tau - 2\tilde{\sigma})\beta, \quad \tilde{\varrho} = 2(1 - \tau - \tilde{\sigma}), \quad \text{and} \quad \hat{\varrho} = \frac{4 - \tau - \theta - 2\tilde{\sigma}}{\beta}.$$

Since $3 - 3\tau - 2\tilde{\sigma} = (1 - \tau) + 2(1 - \tau - \tilde{\sigma})$ and $4 - \tau - \theta - 2\tilde{\sigma} = (\tau + \theta) + 2(2 - \tau - \theta - \tilde{\sigma})$, we obtain, from (8), that $\varrho, \hat{\varrho} > 0$. Moreover,

$$\begin{aligned} \varrho\hat{\varrho} - \tilde{\varrho}^2 &= [(1 - \tau) + 2(1 - \tau - \tilde{\sigma})](4 - \tau - \theta - 2\tilde{\sigma}) - 4(1 - \tau - \tilde{\sigma})^2 \\ &= (1 - \tau - \tilde{\sigma})[(2 - \tau - \theta - \tilde{\sigma}) + 2(3 + \tau - \theta)] + \tilde{\sigma}(3 - \tilde{\sigma} - \theta). \end{aligned}$$

From (8), we have $(1 - \tau - \tilde{\sigma}) > 0$, $(2 - \tau - \theta - \tilde{\sigma}) > 0$ and $\theta < 2$. The latter inequality, together with the facts that $\tau > -1$ and $\tilde{\sigma} < 1$, yields $3 + \tau - \theta > 0$ and $3 - \tilde{\sigma} - \theta > 0$. Therefore, $\det(\widehat{Q}) > 0$ and $\text{Tr}(\widehat{Q}) > 0$, and we conclude that Q is positive definite. In addition, inequalities $\varrho\hat{\varrho} - \tilde{\varrho}^2 > 0$ and $(1 - \tau - \tilde{\sigma}) > 0$ clearly imply that $\vartheta > 0$.

Now, for a given $(y, \gamma) \in \mathbb{R}^p \times \mathbb{R}^m$, using (28), (29) and simple algebraic manipulations, we find

$$\|(y, \gamma)\|_Q^2 = \left\| \sqrt{(3 - 3\tau - 2\tilde{\sigma})\beta}y + \frac{\sqrt{4 - \tau - \theta - 2\tilde{\sigma}}}{\sqrt{\beta}}\gamma \right\|^2 - 2\vartheta \langle y, \gamma \rangle \geq -2\vartheta \langle y, \gamma \rangle,$$

which concluded the proof of the proposition. \square

Proposition 2.7. Consider the functions $\varphi, \widehat{\varphi}, \tilde{\varphi}, \overline{\varphi}: \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$\varphi(\sigma) = (1 - \tau)(\sigma - 1) + (1 - \tau - \tilde{\sigma})(\tau + \theta), \quad (30a)$$

$$\widehat{\varphi}(\sigma) = (1 - \tau)[(1 + \theta)\sigma - 1 + \tau] - \tilde{\sigma}(\tau + \theta), \quad (30b)$$

$$\tilde{\varphi}(\sigma) = \sigma - (1 - \tau - \theta)^2 - \tilde{\sigma}(\tau + \theta), \quad (30c)$$

$$\overline{\varphi}(\sigma) = [(1 + \tau)\widehat{\varphi}(\sigma) - 2\tau\varphi(\sigma)](1 + \tau)\tilde{\varphi}(\sigma) - (1 - \theta)^2(\varphi(\sigma))^2. \quad (30d)$$

Then, there exists a scalar $\sigma \in [\hat{\sigma}, 1)$ such that $\varphi(\sigma) \geq 0$, $\widehat{\varphi}(\sigma) \geq 0$, $\tilde{\varphi}(\sigma) > 0$ and $\overline{\varphi}(\sigma) \geq 0$.

Proof. Since

$$\varphi(1) = (1 - \tau - \tilde{\sigma})(\tau + \theta) = \widehat{\varphi}(1), \quad \widetilde{\varphi}(1) = (2 - \tau - \theta - \tilde{\sigma})(\tau + \theta),$$

and

$$\overline{\varphi}(1) = (1 - \tau - \tilde{\sigma})(\tau + \theta)^2 \left[(1 - \tau^2)(2 - \tau - \theta - \tilde{\sigma}) - (1 - \theta)^2(1 - \tau - \tilde{\sigma}) \right],$$

it follows from (8) that all functions defined in (30) are positive for $\sigma = 1$. Therefore, there exists $\sigma \in [\hat{\sigma}, 1)$ close to 1 such that the statements of the proposition hold. \square

The following lemma provides some estimates of the sequences $\{\|\tilde{z}_k - z_{k-1}\|_M^2\}$ and $\{\|\tilde{z}_k - z_k\|_M^2\}$, which appear in (17).

Lemma 2.8. *Let T , M , $\{p_k\}$, $\{q_k\}$, $\{z_k\}$ and $\{\tilde{z}_k\}$ be as in (18), (20) and (26). Then, for every $k \geq 1$,*

$$\|\tilde{z}_k - z_{k-1}\|_M^2 = \|\tilde{x}_k - x_{k-1}\|_G^2 + \|y_k - y_{k-1}\|_H^2 + a_k, \quad \|\tilde{z}_k - z_k\|_M^2 = \|\tilde{x}_k - x_k\|_G^2 + b_k, \quad (31)$$

where

$$a_k := \frac{(1 - \tau)(1 + \theta)\beta}{\tau + \theta} \|p_k\|^2 + \frac{2(1 - \tau)}{\tau + \theta} \langle p_k, q_k \rangle + \frac{1}{(\tau + \theta)\beta} \|q_k\|^2,$$

and

$$b_k := \frac{(1 - \tau)^2\beta}{\tau + \theta} \|p_k\|^2 + \frac{2(1 - \tau)(1 - \tau - \theta)}{\tau + \theta} \langle p_k, q_k \rangle + \frac{(1 - \tau - \theta)^2}{(\tau + \theta)\beta} \|q_k\|^2.$$

Proof. It follows from (26) and the first equality in (21) that

$$\|\tilde{z}_k - z_{k-1}\|_M^2 = \|(\tilde{x}_k - x_{k-1}, y_k - y_{k-1}, \beta p_k + q_k)\|_M^2.$$

Hence, using (18) and (20), we find

$$\|\tilde{z}_k - z_{k-1}\|_M^2 = \|\tilde{x}_k - x_{k-1}\|_G^2 + \|y_k - y_{k-1}\|_H^2 + \tilde{a}_k$$

where

$$\tilde{a}_k := \frac{(\tau - \tau\theta + \theta)\beta}{\tau + \theta} \|p_k\|^2 - \frac{2\tau}{\tau + \theta} \langle p_k, \beta p_k + q_k \rangle + \frac{1}{(\tau + \theta)\beta} \|\beta p_k + q_k\|^2.$$

By developing the right-hand side of the last expression, we have

$$\begin{aligned} \tilde{a}_k &= \frac{(\tau - \tau\theta + \theta - 2\tau + 1)\beta}{\tau + \theta} \|p_k\|^2 - \frac{2\tau - 2}{\tau + \theta} \langle p_k, q_k \rangle + \frac{1}{(\tau + \theta)\beta} \|q_k\|^2 \\ &= \frac{(1 - \tau)(1 + \theta)\beta}{\tau + \theta} \|p_k\|^2 + \frac{2(1 - \tau)}{\tau + \theta} \langle p_k, q_k \rangle + \frac{1}{(\tau + \theta)\beta} \|q_k\|^2 = a_k. \end{aligned}$$

Therefore, the first equation in (31) follows. Now, using (26), (20), the second equality in (21), and the definition of M in (18), we obtain

$$\begin{aligned} \|\tilde{z}_k - z_k\|_M^2 &= \|(\tilde{x}_k - x_k, 0, (1 - \tau)\beta p_k + (1 - \tau - \theta)q_k)\|_M^2 \\ &= \|\tilde{x}_k - x_{k-1}\|_G^2 + \frac{1}{(\tau + \theta)\beta} \|(1 - \tau)\beta p_k + (1 - \tau - \theta)q_k\|^2, \end{aligned}$$

which is equivalent to the second equation in (31). \square

Before proving the inequality in (17), we establish some other relations satisfied by the sequences generated by Algorithm 1. To do this, we consider the following constant

$$d_0 = \inf \left\{ \|z^* - z_0\|_M^2 : z^* \in T^{-1}(0) \right\}, \quad (32)$$

where M , T and z_0 are as in (18) and (26). Note that, if M is positive definite, then d_0 measures the squared distance in the norm $\|\cdot\|_M$ of the initial point $z_0 = (x_0, y_0, \gamma_0)$ to the solution set of (1).

Lemma 2.9. *Let $\{p_k\}$, $\{q_k\}$ and d_0 be as in (20) and (32). Then, the following hold:*

(a)
$$\min \left\{ 2\vartheta \langle p_1, q_1 \rangle, -\|y_1 - y_0\|_H^2 \right\} \geq -4d_0,$$

where ϑ is as in (29).

(b) for every $k \geq 2$, we have

$$2(1 + \tau) \langle p_k, q_k \rangle \geq 2(1 - \theta) \langle p_k, q_{k-1} \rangle - 2\tau\beta \|p_k\|^2 + \|y_k - y_{k-1}\|_H^2 - \|y_{k-1} - y_{k-2}\|_H^2.$$

Proof. (a) From (22) with $k = 1$, we have $\gamma_1 - \gamma_0 = \tau\beta p_1 + (\tau + \theta)q_1$. Then, using (26) (with $k = 1$) and the definition of M in (18), we find

$$\|z_1 - z_0\|_M^2 = \|(x_1 - x_0, y_1 - y_0, \gamma_1 - \gamma_0)\|_M^2 = \|x_1 - x_0\|_G^2 + \|y_1 - y_0\|_H^2 + c_1, \quad (33)$$

where

$$\begin{aligned} c_1 &:= \frac{(\tau - \tau\theta + \theta)\beta}{\tau + \theta} \|p_1\|^2 - \frac{2\tau}{\tau + \theta} \langle p_1, \gamma_1 - \gamma_0 \rangle + \frac{1}{(\tau + \theta)\beta} \|\gamma_1 - \gamma_0\|^2 \\ &= \frac{(\tau - \tau\theta + \theta - \tau^2)\beta}{\tau + \theta} \|p_1\|^2 + \frac{\tau + \theta}{\beta} \|q_1\|^2 = (1 - \tau)\beta \|p_1\|^2 + \frac{\tau + \theta}{\beta} \|q_1\|^2. \end{aligned} \quad (34)$$

Let $z^* = (x^*, y^*, \gamma^*)$ be an arbitrary solution of (16), i.e., $z^* \in T^{-1}(0)$ with T as in (18). Hence, it follows from (33) and the fact that $\|z - z'\|_M^2 \leq 2(\|z\|_M^2 + \|z'\|_M^2)$, for all z, z' , that

$$c_1 + \|y_1 - y_0\|_H^2 \leq \|z_1 - z_0\|_M^2 \leq 2 \left(\|z^* - z_1\|_M^2 + \|z^* - z_0\|_M^2 \right). \quad (35)$$

On the other hand, it follows from (10) with $k = 1$ and the definition of x_1 in (14) that

$$\|\tilde{x}_1 - x_1\|_G^2 \leq \hat{\sigma} \|\tilde{x}_1 - x_0\|_G^2 + \frac{\tilde{\sigma}}{\beta} \|\tilde{\gamma}_1 - \gamma_0\|^2 \leq \|\tilde{x}_1 - x_0\|_G^2 + \frac{\tilde{\sigma}}{\beta} \|\tilde{\gamma}_1 - \gamma_0\|^2,$$

where, in the second inequality, we use that $\hat{\sigma} < 1$. Thus, using the first identity in (21) with $k = 1$, we obtain

$$\|\tilde{x}_1 - x_0\|_G^2 - \|\tilde{x}_1 - x_1\|_G^2 \geq -\frac{\tilde{\sigma}}{\beta} \|\beta p_1 + q_1\|^2.$$

This inequality, together with Lemma 2.5 and Lemma 2.8 (with $k = 1$), implies that

$$\begin{aligned} \|z^* - z_0\|_M^2 - \|z^* - z_1\|_M^2 &\geq \|\tilde{z}_1 - z_0\|_M^2 - \|\tilde{z}_1 - z_1\|_M^2 \\ &\geq \|\tilde{x}_1 - x_0\|_G^2 - \|\tilde{x}_1 - x_1\|_G^2 + \frac{(1 - \tau)(\tau + \theta)\beta}{\tau + \theta} \|p_1\|^2 \\ &\quad + \frac{2(1 - \tau)(\tau + \theta)}{\tau + \theta} \langle p_1, q_1 \rangle + \frac{1 - [1 - (\tau + \theta)]^2}{(\tau + \theta)\beta} \|q_1\|^2 \\ &\geq (1 - \tau - \tilde{\sigma}) \left[\beta \|p_1\|^2 + 2 \langle p_1, q_1 \rangle \right] + \frac{2 - \tau - \theta - \tilde{\sigma}}{\beta} \|q_1\|^2. \end{aligned}$$

Combining this inequality with (35) and using the identity in (34), we find

$$\begin{aligned} 4 \|z^* - z_0\|_M^2 &\geq \|y_1 - y_0\|_H^2 + (3 - 3\tau - 2\tilde{\sigma}) \beta \|p_1\|^2 + 4(1 - \tau - \tilde{\sigma}) \langle p_1, q_1 \rangle + \frac{4 - \tau - \theta - 2\tilde{\sigma}}{\beta} \|q_1\|^2 \\ &= \|y_1 - y_0\|_H^2 + \|(p_1, q_1)\|_Q^2, \end{aligned}$$

where Q is as in (28). Hence, using Proposition 2.6 we conclude that

$$\max \left\{ -\vartheta \langle p_1, q_1 \rangle, \|y_1 - y_0\|_H^2 \right\} \leq 4 \|z^* - z_0\|_M^2.$$

Therefore, statement (a) follows from the definition of d_0 in (32).

(b) It follows from the definitions of γ_k and q_k in (14) and (20), respectively, that

$$\gamma_{k-\frac{1}{2}} - \beta (A\tilde{x}_k + By_k - b) = \gamma_k - (1 - \theta) \beta (A\tilde{x}_k + By_k - b) = \gamma_k + (1 - \theta) q_k.$$

Hence, since y_k is an optimal solution of (13), we obtain, for every $k \geq 1$,

$$\begin{aligned} 0 &\in \partial g(y_k) - B^* \left[\gamma_{k-\frac{1}{2}} - \beta (A\tilde{x}_k + By_k - b) \right] + H(y_k - y_{k-1}) \\ &= \partial g(y_k) - B^* [\gamma_k + (1 - \theta) q_k] + H(y_k - y_{k-1}). \end{aligned}$$

Thus, the monotonicity of ∂g and the definition of p_k in (20) imply that, for every $k \geq 2$,

$$\begin{aligned} 0 &\leq \langle \gamma_k - \gamma_{k-1} + (1 - \theta) (q_k - q_{k-1}), p_k \rangle - \|y_k - y_{k-1}\|_H^2 + \langle H(y_{k-1} - y_{k-2}), y_k - y_{k-1} \rangle \\ &= \langle \tau \beta p_k + (1 + \tau) q_k - (1 - \theta) q_{k-1}, p_k \rangle - \|y_k - y_{k-1}\|_H^2 + \langle H(y_{k-1} - y_{k-2}), y_k - y_{k-1} \rangle \\ &\leq \tau \beta \|p_k\|^2 + (1 + \tau) \langle p_k, q_k \rangle - (1 - \theta) \langle p_k, q_{k-1} \rangle - \frac{1}{2} \|y_k - y_{k-1}\|_H^2 + \frac{1}{2} \|y_{k-1} - y_{k-2}\|_H^2, \end{aligned}$$

where the second equality is due to (22) and the last inequality is due to the fact that $2 \langle Hy, y' \rangle \leq \|y\|_H^2 + \|y'\|_H^2$ for all $y, y' \in \mathbb{R}^p$. Therefore, the desired inequality follows immediately from the last one. \square

With the above propositions and lemmas, we now prove the inequality in (17).

Theorem 2.10. *Let $\{z_k\}$, $\{\tilde{z}_k\}$ and $\{q_k\}$ be as in (26) and (20) and assume that $\sigma \in [\hat{\sigma}, 1)$ is given by Proposition 2.7. Consider the sequence $\{\eta_k\}$ defined by*

$$\eta_0 = \frac{4(1 + \tau + \vartheta) \varphi(\sigma)}{(\tau + \theta)(1 + \tau) \vartheta} d_0, \quad \eta_k = \frac{\tilde{\varphi}(\sigma)}{(\tau + \theta) \beta} \|q_k\|^2 + \frac{\varphi(\sigma)}{(\tau + \theta)(1 + \tau)} \|y_k - y_{k-1}\|_H^2, \quad \forall k \geq 1, \quad (36)$$

where ϑ , d_0 , φ and $\tilde{\varphi}$ are as in (29), (32), (30a) and (30c), respectively. Then, for every $k \geq 1$,

$$\|\tilde{z}_k - z_k\|_M^2 + \eta_k \leq \sigma \|\tilde{z}_k - z_{k-1}\|_M^2 + \eta_{k-1}, \quad (37)$$

where M is as in (18).

Proof. It follows from Lemma 2.8 that

$$\begin{aligned} \sigma \|\tilde{z}_k - z_{k-1}\|_M^2 - \|\tilde{z}_k - z_k\|_M^2 &= \sigma \|\tilde{x}_k - x_{k-1}\|_G^2 - \|\tilde{x}_k - x_k\|_G^2 \\ &\quad + \sigma \|y_k - y_{k-1}\|_H^2 + \frac{(1-\tau)[(1+\theta)\sigma - (1-\tau)]\beta}{\tau+\theta} \|p_k\|^2 \\ &\quad + \frac{2(1-\tau)[\sigma - (1-\tau-\theta)]}{\tau+\theta} \langle p_k, q_k \rangle + \frac{\sigma - (1-\tau-\theta)^2}{(\tau+\theta)\beta} \|q_k\|^2. \end{aligned} \quad (38)$$

Using the inequality in (10), the definition of x_k in (14) and noting that $\sigma \geq \tilde{\sigma}$, we obtain

$$\sigma \|\tilde{x}_k - x_{k-1}\|_G^2 - \|\tilde{x}_k - x_k\|_G^2 \geq -\frac{\tilde{\sigma}}{\beta} \|\tilde{\gamma}_k - \gamma_{k-1}\|^2 = -\tilde{\sigma}\beta \|p_k\|^2 - 2\tilde{\sigma} \langle p_k, q_k \rangle - \frac{\tilde{\sigma}}{\beta} \|q_k\|^2$$

where the last equality is due to the first expression in (21). Combining the last inequality with (38) and definitions in (30), we find

$$\sigma \|\tilde{z}_k - z_{k-1}\|_M^2 - \|\tilde{z}_k - z_k\|_M^2 \geq \frac{\widehat{\varphi}(\sigma)\beta}{\tau+\theta} \|p_k\|^2 + \frac{2\varphi(\sigma)}{\tau+\theta} \langle p_k, q_k \rangle + \frac{\widetilde{\varphi}(\sigma)}{(\tau+\theta)\beta} \|q_k\|^2. \quad (39)$$

Let us now consider two cases: $k = 1$ and $k \geq 2$.

Case 1 ($k = 1$): From (39) with $k = 1$, Lemma 2.9(a) and the fact that $\varphi(\sigma) \geq 0$, we have

$$\sigma \|\tilde{z}_1 - z_0\|_M^2 - \|\tilde{z}_1 - z_1\|_M^2 \geq \frac{\widehat{\varphi}(\sigma)\beta}{\tau+\theta} \|p_1\|^2 - \frac{4\varphi(\sigma)}{(\tau+\theta)\vartheta} d_0 + \frac{\widetilde{\varphi}(\sigma)}{(\tau+\theta)\beta} \|q_1\|^2.$$

Hence, in view of the definitions of η_0 and η_1 in (36), we conclude that

$$\sigma \|\tilde{z}_1 - z_0\|_M^2 - \|\tilde{z}_1 - z_1\|_M^2 + \eta_0 - \eta_1 \geq \frac{\widehat{\varphi}(\sigma)\beta}{\tau+\theta} \|p_1\|^2 + \frac{\varphi(\sigma)}{(\tau+\theta)(1+\tau)} \left[4d_0 - \|y_1 - y_0\|_H^2 \right] \geq 0,$$

where the last inequality is due to Lemma 2.9(a) and Proposition 2.7. This implies that (37) holds for $k = 1$.

Case 2 ($k \geq 2$): It follows from Lemma 2.9(b) and (39) that

$$\begin{aligned} \sigma \|\tilde{z}_k - z_{k-1}\|_M^2 - \|\tilde{z}_k - z_k\|_M^2 &\geq \frac{\widehat{\varphi}(\sigma)\beta}{\tau+\theta} \|p_k\|^2 + \frac{\widetilde{\varphi}(\sigma)}{(\tau+\theta)\beta} \|q_k\|^2 \\ &\quad + \frac{\varphi(\sigma)}{(\tau+\theta)(1+\tau)} \left[2(1-\theta) \langle p_k, q_{k-1} \rangle - 2\tau\beta \|p_k\|^2 + \|y_k - y_{k-1}\|_H^2 - \|y_{k-1} - y_{k-2}\|_H^2 \right], \end{aligned}$$

which, combined with the definition of η_k in (36), yields

$$\begin{aligned} \sigma \|\tilde{z}_k - z_{k-1}\|_M^2 - \|\tilde{z}_k - z_k\|_M^2 + \eta_{k-1} - \eta_k &\geq \frac{[(1+\tau)\widehat{\varphi}(\sigma) - 2\tau\varphi(\sigma)]\beta}{(\tau+\theta)(1+\tau)} \|p_k\|^2 \\ &\quad + \frac{2(1-\theta)\varphi(\sigma)}{(\tau+\theta)(1+\tau)} \langle p_k, q_{k-1} \rangle + \frac{\widetilde{\varphi}(\sigma)}{(\tau+\theta)\beta} \|q_{k-1}\|^2. \end{aligned}$$

For simplicity, we define constants a , b , and c by

$$a = \frac{[(1+\tau)\widehat{\varphi}(\sigma) - 2\tau\varphi(\sigma)]\beta}{(\tau+\theta)(1+\tau)}, \quad b = \frac{(1-\theta)\varphi(\sigma)}{(\tau+\theta)(1+\tau)}, \quad \text{and} \quad c = \frac{\widetilde{\varphi}(\sigma)}{(\tau+\theta)\beta}.$$

Hence,

$$\sigma \|\tilde{z}_k - z_{k-1}\|_M^2 - \|\tilde{z}_k - z_k\|_M^2 + \eta_{k-1} - \eta_k \geq a \|p_k\|^2 - 2b \langle p_k, q_{k-1} \rangle + c \|q_{k-1}\|^2. \quad (40)$$

Now, note that

$$ac - b^2 = \frac{[(1 + \tau) \widehat{\varphi}(\sigma) - 2\tau\varphi(\sigma)](1 + \tau) \widetilde{\varphi}(\sigma) - (1 - \theta)^2 (\varphi(\sigma))^2}{(\tau + \theta)^2 (1 + \tau)^2} = \frac{\overline{\varphi}(\sigma)}{(\tau + \theta)^2 (1 + \tau)^2},$$

where $\overline{\varphi}$ is given in (30d). Therefore, it follows from Proposition 2.7 that $c > 0$ and $ac - b^2 \geq 0$, which, combined with (40), implies that (37) also holds for $k \geq 2$. \square

Remark 2.11. *If $\tau = 0$ (resp. $\theta = 1$), then Theorems 2.4 and 2.10 correspond to Lemma 3.1 and Theorem 3.3 in [2] (resp. [3, Proposition 1(a)]).*

2.2 Pointwise and ergodic convergence rates of Algorithm 1

In this section, we establish pointwise and ergodic convergence rates for Algorithm 1.

Theorem 2.12 (Pointwise convergence rate of Algorithm 1). *Consider the sequences $\{v_k\}$ and $\{w_k\}$ defined, for every $k \geq 1$, by*

$$v_k = \left(H + \frac{(\tau - \tau\theta + \theta)\beta}{\tau + \theta} B^* B \right) (y_{k-1} - y_k) - \frac{\tau}{\tau + \theta} B^* (\gamma_{k-1} - \gamma_k), \quad (41)$$

$$w_k = -\frac{\tau}{\tau + \theta} B (y_{k-1} - y_k) + \frac{1}{(\tau + \theta)\beta} (\gamma_{k-1} - \gamma_k). \quad (42)$$

Then, for every $k \geq 1$,

$$u_k \in \partial f(\tilde{x}_k) - A^* \tilde{\gamma}_k, \quad v_k \in \partial g(y_k) - B^* \tilde{\gamma}_k, \quad w_k = A\tilde{x}_k + By_k - b, \quad (43)$$

and there exists $i \leq k$ such that

$$\max \{\|u_i\|, \|v_i\|, \|w_i\|\} \leq \sqrt{\frac{2\lambda_M d_0 \mathcal{C}_1}{k}} \quad (44)$$

where $\mathcal{C}_1 := [1 + \sigma + 8(1 + \tau + \vartheta)\varphi(\sigma)/((\tau + \theta)(1 + \tau)\vartheta)]/[1 - \sigma]$, λ_M is the largest eigenvalue of the matrix M defined in (18), $\sigma \in [\hat{\sigma}, 1)$ is given by Proposition 2.7 and ϑ , φ and d_0 are as in (29), (30a) and (32), respectively.

Proof. By noting that $u_k = G(x_{k-1} - x_k)$ (see (14)), the expressions in (43) follow immediately from (41), (42) and Theorem 2.4. From Theorem 2.4, we have $(u_k, v_k, w_k) = M(z_{k-1} - z_k)$ and hence

$$\begin{aligned} \|(u_k, v_k, w_k)\|^2 &\leq \lambda_M \|z_{k-1} - z_k\|_M^2 \leq 2\lambda_M \left[\|z_{k-1} - \tilde{z}_k\|_M^2 + \|\tilde{z}_k - z_k\|_M^2 \right] \\ &\leq 2\lambda_M \left[(1 + \sigma) \|z_{k-1} - \tilde{z}_k\|_M^2 + \eta_{k-1} - \eta_k \right], \end{aligned}$$

where the last inequality is due to (37). On the other hand, from Lemma 2.5 and (37), we obtain

$$\|z^* - z_k\|_M^2 - \|z^* - z_{k-1}\|_M^2 \leq (\sigma - 1) \|\tilde{z}_k - z_{k-1}\|_M^2 + \eta_{k-1} - \eta_k,$$

where $z^* \in T^{-1}(0)$. The last two estimates and the fact that $\sigma < 1$ imply that, for every $k \geq 1$

$$\begin{aligned} \|(u_k, v_k, w_k)\|^2 &\leq 2\lambda_M \left[\frac{1+\sigma}{1-\sigma} \left(\|z^* - z_{k-1}\|_M^2 - \|z^* - z_k\|_M^2 + \eta_{k-1} - \eta_k \right) + \eta_{k-1} - \eta_k \right] \\ &= \frac{2\lambda_M}{1-\sigma} \left[(1+\sigma) \left(\|z^* - z_{k-1}\|_M^2 - \|z^* - z_k\|_M^2 \right) + 2(\eta_{k-1} - \eta_k) \right]. \end{aligned}$$

By summing the above inequality from $k = 1$ to k , we obtain

$$\sum_{l=1}^k \|(u_l, v_l, w_l)\|^2 \leq \frac{2\lambda_M}{1-\sigma} \left[(1+\sigma) \|z^* - z_0\|_M^2 + 2\eta_0 \right], \quad (45)$$

which, combined with the definitions of d_0 and η_0 (32) and (36), respectively, yields

$$k \left(\min_{l=1, \dots, k} \|(u_l, v_l, w_l)\|^2 \right) \leq \frac{2\lambda_M}{1-\sigma} \left[(1+\sigma) + \frac{8(1+\tau+\vartheta)\varphi(\sigma)}{(\tau+\theta)(1+\tau)\vartheta} \right] d_0.$$

Therefore, (44) follows now from the last inequality and the definition of \mathcal{C}_1 . \square

Remark 2.13. (a) It follows from Theorem 2.12 that, for a given tolerance $\rho > 0$, Algorithm 1 generates a ρ -approximate solution $(\tilde{x}_i, y_i, \tilde{\gamma}_i)$ of (16) with residual (u_i, v_i, w_i) , i.e.,

$$u_i \in \partial f(\tilde{x}_i) - A^* \tilde{\gamma}_i, \quad v_i \in \partial g(y_i) - B^* \tilde{\gamma}_i, \quad w_i = A\tilde{x}_i + By_i - b,$$

such that

$$\max\{\|u_i\|, \|v_i\|, \|w_i\|\} \leq \rho,$$

in at most

$$\bar{k} = \left\lceil \frac{2\lambda_M d_0 \mathcal{C}_1}{\rho^2} \right\rceil$$

iterations. (b) Theorem 2.12 encompasses many recently pointwise convergence rates of ADMM variants. Namely, (i) by taking $\tau = 0$ and $G = I/\beta$, we obtain the pointwise convergence rate of the partially inexact proximal ADMM established in [2, Theorem 3.1]. Additionally, if $\tilde{\sigma} = \hat{\sigma} = 0$, the pointwise rate of the FG-P-ADMM in [21, Theorem 2.1] is recovered. (ii) By choosing $\theta = 1$ and $G = I/\beta$, we have the pointwise rate of the inexact proximal generalized ADMM as in [3, Theorem 1]. Finally, if $\theta = 1$, $G = I/\beta$ and $\tilde{\sigma} = \hat{\sigma} = 0$, the pointwise convergence rate of the G-P-ADMM in [1, Theorem 3.4] is obtained.

Theorem 2.14 (Ergodic convergence rate of Algorithm 1). Consider the sequences $\{(x_k^a, y_k^a, \gamma_k^a, \tilde{x}_k^a, \tilde{\gamma}_k^a)\}$, $\{(u_k^a, v_k^a, w_k^a)\}$, and $\{(\varepsilon_k^a, \zeta_k^a)\}$ defined, for every $k \geq 1$, by

$$(x_k^a, y_k^a, \gamma_k^a, \tilde{x}_k^a, \tilde{\gamma}_k^a) = \frac{1}{k} \sum_{i=1}^k (x_i, y_i, \gamma_i, \tilde{x}_i, \tilde{\gamma}_i), \quad (u_k^a, v_k^a, w_k^a) = \frac{1}{k} \sum_{i=1}^k (u_i, v_i, w_i), \quad (46)$$

$$\varepsilon_k^a = \frac{1}{k} \sum_{i=1}^k \langle u_i + A^* \tilde{\gamma}_i, \tilde{x}_i - \tilde{x}_k^a \rangle, \quad \text{and} \quad \zeta_k^a = \frac{1}{k} \sum_{i=1}^k \langle v_i + B^* \tilde{\gamma}_i, y_i - y_k^a \rangle, \quad (47)$$

where v_i and w_i are as in (41) and (42), respectively. Then, for every $k \geq 1$, there hold $\varepsilon_k^a \geq 0$, $\zeta_k^a \geq 0$,

$$u_k^a \in \partial_{\varepsilon_k^a} f(\tilde{x}_k^a) - A^* \tilde{\gamma}_k^a, \quad v_k^a \in \partial_{\zeta_k^a} g(y_k^a) - B^* \tilde{\gamma}_k^a, \quad w_k^a = A\tilde{x}_k^a + By_k^a - b, \quad (48)$$

$$\max \{\|u_k^a\|, \|v_k^a\|, \|w_k^a\|\} \leq \frac{2\sqrt{\lambda_M d_0 \mathcal{C}_2}}{k}, \quad \max \{\varepsilon_k^a, \zeta_k^a\} \leq \frac{3d_0 \mathcal{C}_3}{2k}, \quad (49)$$

where $\mathcal{C}_2 := [1 + 4(1 + \tau + \vartheta) \varphi(\sigma) / ((\tau + \theta)(1 + \tau)\vartheta)]$ and $\mathcal{C}_3 := (3 - 2\sigma)\mathcal{C}_2 / (1 - \sigma)$, λ_M is the largest eigenvalue of the matrix M defined in (18), $\sigma \in [\hat{\sigma}, 1)$ is given by Proposition 2.7 and ϑ , φ and d_0 are as in (29), (30a) and (32), respectively.

Proof. For every $i \geq 1$, it follows from Theorem 2.12 that

$$u_i + A^* \tilde{\gamma}_i \in \partial f(\tilde{x}_i), \quad v_i + B^* \tilde{\gamma}_i \in \partial g(y_i), \quad w_i = A\tilde{x}_i + By_i - b.$$

Hence, using (46), we immediately obtain the last equality in (48). Furthermore, from the first two inclusions above, (46), (47) and [20, Theorem 2.1] we conclude that, for every $k \geq 1$, $\varepsilon_k^a \geq 0$, $\zeta_k^a \geq 0$, and the inclusions in (48) hold. To show (49), we recall again that $(u_i, v_i, w_i) = M(z_{i-1} - z_i)$ (see the proof of Theorem 2.12), which together with (46), yields $(u_k^a, v_k^a, w_k^a) = (1/k)M(z_0 - z_k)$. Then, for an arbitrary solution $z^* = (x^*, y^*, \gamma^*)$ of (16), we have

$$\|(u_k^a, v_k^a, w_k^a)\|^2 \leq \frac{\lambda_M}{k^2} \|z_0 - z_k\|_M^2 \leq \frac{2\lambda_M}{k^2} \left(\|z^* - z_0\|_M^2 + \|z^* - z_k\|_M^2 \right).$$

Combining Lemma 2.5 with (37), we obtain, for every $k \geq 1$, that

$$\|z^* - z_k\|_M^2 + \eta_k \leq \|z^* - z_{k-1}\|_M^2 + (\sigma - 1) \|\tilde{z}_k - z_{k-1}\|_M^2 + \eta_{k-1} \leq \|z^* - z_{k-1}\|_M^2 + \eta_{k-1}. \quad (50)$$

The last two expressions imply that

$$\|(u_k^a, v_k^a, w_k^a)\|^2 \leq \frac{4\lambda_M}{k^2} \left(\|z^* - z_0\|_M^2 + \eta_0 \right),$$

which, combined with the definitions of d_0 and η_0 in (32) and (36), respectively, implies the first inequality in (49). Let us now show the second inequality in (49). From definitions in (47), we have

$$\begin{aligned} \varepsilon_k^a + \zeta_k^a &= \frac{1}{k} \sum_{i=1}^k \left(\langle u_i, \tilde{x}_i - \tilde{x}_k^a \rangle + \langle v_i, y_i - y_k^a \rangle + \langle \tilde{\gamma}_i, A\tilde{x}_i + By_i - A\tilde{x}_k^a - By_k^a \rangle \right) \\ &= \frac{1}{k} \sum_{i=1}^k \left(\langle u_i, \tilde{x}_i - \tilde{x}_k^a \rangle + \langle v_i, y_i - y_k^a \rangle + \langle \tilde{\gamma}_i, w_i - w_k^a \rangle \right) \\ &= \frac{1}{k} \sum_{i=1}^k \left(\langle u_i, \tilde{x}_i - \tilde{x}_k^a \rangle + \langle v_i, y_i - y_k^a \rangle + \langle w_i, \tilde{\gamma}_i - \tilde{\gamma}_k^a \rangle \right), \end{aligned}$$

where the second equality is due to the expressions of w_i and w_k^a in (43) and (48), respectively, and the third follows from the fact that

$$\frac{1}{k} \sum_{i=1}^k \langle \tilde{\gamma}_i, w_i - w_k^a \rangle = \frac{1}{k} \sum_{i=1}^k \langle \tilde{\gamma}_i - \tilde{\gamma}_k^a, w_i - w_k^a \rangle = \frac{1}{k} \sum_{i=1}^k \langle w_i, \tilde{\gamma}_i - \tilde{\gamma}_k^a \rangle$$

(see the definitions of w_k^a and $\tilde{\gamma}_k^a$ in (46)). Hence, setting $\tilde{z}_k^a = (\tilde{x}_k^a, y_k^a, \tilde{\gamma}_k^a)$, and noting that $(u_i, v_i, w_i) = M(z_{i-1} - z_i)$ and $\tilde{z}_i = (\tilde{x}_i, y_i, \tilde{\gamma}_i)$, we obtain

$$\varepsilon_k^a + \zeta_k^a = \frac{1}{k} \sum_{i=1}^k \langle M(z_{i-1} - z_i), \tilde{z}_i - \tilde{z}_k^a \rangle. \quad (51)$$

On the other hand, using (37), we deduce that for all $z \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$

$$\begin{aligned} \|z - z_i\|_M^2 - \|z - z_{i-1}\|_M^2 &= \|\tilde{z}_i - z_i\|_M^2 - \|\tilde{z}_i - z_{i-1}\|_M^2 + 2 \langle M(z_{i-1} - z_i), z - \tilde{z}_i \rangle \\ &\leq (\sigma - 1) \|\tilde{z}_i - z_{i-1}\|_M^2 + \eta_{i-1} - \eta_i + 2 \langle M(z_{i-1} - z_i), z - \tilde{z}_i \rangle, \end{aligned}$$

and then, since $\sigma < 1$, we find

$$2 \sum_{i=1}^k \langle M(z_{i-1} - z_i), \tilde{z}_i - z \rangle \leq \|z - z_0\|_M^2 - \|z - z_k\|_M^2 + \eta_0 - \eta_k \leq \|z - z_0\|_M^2 + \eta_0.$$

Applying this result with $z := \tilde{z}_k^a$ and combining with (51), we find

$$2k(\varepsilon_k^a + \zeta_k^a) \leq \|\tilde{z}_k^a - z_0\|_M^2 + \eta_0 \leq \frac{1}{k} \sum_{i=1}^k \|\tilde{z}_i - z_0\|_M^2 + \eta_0 \leq \max_{i=1, \dots, k} \|\tilde{z}_i - z_0\|_M^2 + \eta_0, \quad (52)$$

where, in the second inequality, we used the convexity of $\|\cdot\|_M^2$ and the fact that $\tilde{z}_k^a = (1/k) \sum_{i=1}^k \tilde{z}_i$. Additionally, since $\|z + z' + z''\|_M^2 \leq 3(\|z\|_M^2 + \|z'\|_M^2 + \|z''\|_M^2)$, for all $z, z', z'' \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$, we also have

$$\|\tilde{z}_i - z_0\|_M^2 \leq 3 \left[\|\tilde{z}_i - z_i\|_M^2 + \|z^* - z_i\|_M^2 + \|z^* - z_0\|_M^2 \right], \quad \forall i \geq 1.$$

This, together with (37) and (50), implies that

$$\begin{aligned} \|\tilde{z}_i - z_0\|_M^2 &\leq 3 \left[\sigma \|\tilde{z}_i - z_{i-1}\|_M^2 + \eta_{i-1} + \|z^* - z_{i-1}\|_M^2 + \eta_{i-1} + \|z^* - z_0\|_M^2 \right] \\ &\leq 3 \left[\sigma \|\tilde{z}_i - z_{i-1}\|_M^2 + 2 \left(\|z^* - z_{i-1}\|_M^2 + \eta_{i-1} \right) + \|z^* - z_0\|_M^2 \right] \\ &\leq 3 \left[\sigma \|\tilde{z}_i - z_{i-1}\|_M^2 + 3 \|z^* - z_0\|_M^2 + 2\eta_0 \right], \end{aligned}$$

which, combined with (52), yields

$$2k(\varepsilon_k^a + \zeta_k^a) \leq 3 \left[3 \left(\|z^* - z_0\|_M^2 + \eta_0 \right) + \sigma \max_{i=1, \dots, k} \|\tilde{z}_i - z_{i-1}\|_M^2 \right].$$

Now, from (50), it is also possible to verify that

$$(1 - \sigma) \|\tilde{z}_i - z_{i-1}\|_M^2 \leq \|z^* - z_{i-1}\|_M^2 + \eta_{i-1} \leq \|z^* - z_0\|_M^2 + \eta_0,$$

and, therefore

$$\varepsilon_k^a + \zeta_k^a \leq \frac{3(3 - 2\sigma)}{2(1 - \sigma)k} \left(\|z^* - z_0\|_M^2 + \eta_0 \right). \quad (53)$$

Therefore, the second inequality in (49) now follows from the definitions of d_0 and η_0 in (32) and (36), respectively. \square

Remark 2.15. (a) It follows from Theorem 2.14 that, for a given tolerance $\rho > 0$, Algorithm 1 generates a ρ -approximate solution $(\tilde{x}_k^a, y_k^a, \tilde{\gamma}_k^a)$ of (16) with residuals (u_k^a, v_k^a, w_k^a) and $(\varepsilon_k^a, \zeta_k^a)$, i.e.,

$$u_k^a \in \partial_{\varepsilon_k^a} f(\tilde{x}_k^a) - A^* \tilde{\gamma}_k^a, \quad v_k^a \in \partial_{\zeta_k^a} g(y_k^a) - B^* \tilde{\gamma}_k^a, \quad w_k^a = A \tilde{x}_k^a + B y_k^a - b,$$

such that

$$\max\{\|u_k^a\|, \|v_k^a\|, \|w_k^a\|, \varepsilon_k^a, \zeta_k^a\} \leq \rho,$$

in at most $\bar{k} = \max\{k_1, k_2\}$ iterations, where

$$k_1 = \left\lceil \frac{2\sqrt{\lambda_M d_0 \mathcal{C}_2}}{\rho} \right\rceil, \quad \text{and} \quad k_2 = \left\lceil \frac{3d_0 \mathcal{C}_3}{2\rho} \right\rceil.$$

(b) Similarly to Theorem 2.12, Theorem 2.14 recovers, in particular, many recently ergodic convergence rates of ADMM variants. Namely, (i) by taking $\tau = 0$ and $G = I/\beta$, we obtain the ergodic convergence rate of the partially inexact proximal ADMM established in [2, Theorem 3.2]. Additionally, if $\tilde{\sigma} = \hat{\sigma} = 0$, the ergodic rate of the FG-P-ADMM with $\theta \in (0, (1 + \sqrt{5})/2)$ in [21, Theorem 2.2] is obtained. (ii) By choosing $\theta = 1$ and $G = I/\beta$, we have the ergodic rate of the inexact proximal generalized ADMM as in [3, Theorem 2]. Finally, if $\theta = 1$, $G = I/\beta$ and $\tilde{\sigma} = \hat{\sigma} = 0$, the ergodic convergence rate of the G-P-ADMM with $\tau \in (-1, 1)$ in [1, Theorem 3.6] is recuperated.

3 Numerical experiments

The purpose of this section is to assess the practical behavior of the proposed method. We first mention that the inexact FG-P-ADMM (Algorithm 1 with $\tau = 0$) and the inexact G-P-ADMM (Algorithm 1 with $\theta = 1$) have been shown very efficient in some applications. Indeed, as reported in [2], the inexact FG-P-ADMM with $\theta = 1.6$ outperformed other inexact ADMMs for two classes of problems, namely, LASSO and ℓ_1 -regularized logistic regression. On the other hand, the inexact G-P-ADMM, proposed later in [3], with $\tau = 0.9$ (or, $\alpha = 1.9$ in term of the relaxation factor α) showed to be even more efficient than the FG-P-ADMM with $\theta = 1.6$ for these same classes of problems. Therefore, our goal here is to investigate the efficiency of Algorithm 1, which combines both acceleration parameters τ and θ in a single method, for solving another real-life application. The computational results were obtained using MATLAB R2018a on a 2.4 GHz Intel(R) Core i7 computer with 8 GB of RAM.

We use as test problem the total variation (TV) regularization problem (a.k.a. TV/L2 minimization), first proposed by [28],

$$\min_{x \in \mathbb{R}^{m \times n}} \frac{\mu}{2} \|Kx - c\|^2 + \|x\|_{TV}, \quad (54)$$

where $x \in \mathbb{R}^{m \times n}$ is the original image to be restored, μ is a positive regularization parameter, $K : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$ is a linear operator representing some blurring operator, $c \in \mathbb{R}^{m \times n}$ is the degraded image and $\|\cdot\|_{TV}$ is the discrete TV-norm. Let us briefly recall the definition of TV-norm. Let $x \in \mathbb{R}^{m \times n}$ be given and consider D^1 and D^2 the first-order finite difference $m \times n$ matrices in the horizontal and vertical directions, respectively, which, under the periodic boundary condition, are defined by

$$(D^1 x)_{i,j} = \begin{cases} x_{i+1,j} - x_{i,j} & \text{if } i < m, \\ x_{1,j} - x_{m,j} & \text{if } i = m, \end{cases} \quad (D^2 x)_{i,j} = \begin{cases} x_{i,j+1} - x_{i,j} & \text{if } j < n, \\ x_{i,1} - x_{i,n} & \text{if } j = n, \end{cases}$$

for $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$. By defining $D = (D^1; D^2)$, we obtain

$$\|x\|_{TV} = \|x\|_{TV_s} := \|Dx\|_s := \sum_{i=1}^m \sum_{j=1}^n \left\| (Dx)_{i,j} \right\|_s, \quad (55)$$

where $(Dx)_{i,j} = \left((D^1x)_{i,j}, (D^2x)_{i,j} \right) \in \mathbb{R}^2$ and $s = 1$ or 2 . The TV norm is known as anisotropic and isotropic if $s = 1$ and $s = 2$, respectively. Here, we consider only the isotropic case.

By introducing an auxiliary variable $y = (y^1, y^2)$ where $y^1, y^2 \in \mathbb{R}^{m \times n}$ and, in view of the definition in (55), the problem in (54) can be written as

$$\min_{x,y} \frac{\mu}{2} \|Kx - c\|^2 + \|y\|_2 \quad s.t. \quad y = Dx, \quad (56)$$

which is obviously an instance of (1) with $f(x) = \frac{\mu}{2} \|Kx - c\|^2$, $g(y) = \|y\|_2$, $A = -D$, $B = I$, and $b = 0$. In this case, the pair (\tilde{x}_k, u_k) in (9) can be obtained by computing an approximate solution \tilde{x}_k with a residual u_k of the following linear system

$$\left(\mu K^\top K + \beta D^\top D \right) x = \mu K^\top c + D^\top (\beta y_{k-1} - \gamma_{k-1}).$$

In our implementation, the above linear system was reshaped as a linear system of size $mn \times 1$ and then solved by means of the conjugate gradient method [26] starting from the origin. Note that, by using the two-dimensional shrinkage operator [31, 36], the subproblem (13) has a closed-form solution $y_k = (y_k^1, y_k^2)$ given explicitly by

$$\left((y_k^1)_{i,j}, (y_k^2)_{i,j} \right) := \max \left\{ \left\| (w_{i,j}^1, w_{i,j}^2) \right\| - \frac{1}{\beta}, 0 \right\} \left(\frac{w_{i,j}^1}{\left\| (w_{i,j}^1, w_{i,j}^2) \right\|}, \frac{w_{i,j}^2}{\left\| (w_{i,j}^1, w_{i,j}^2) \right\|} \right),$$

for $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$, where

$$(w^1, w^2) := (D^1 \tilde{x}_k + (1/\beta) \gamma_{k-\frac{1}{2}}^1, D^2 \tilde{x}_k + (1/\beta) \gamma_{k-\frac{1}{2}}^2),$$

and the convention $0 \cdot (0/0) = 0$ is followed.

The initialization parameters in Algorithm 1 were set as follows: $(x_0, y_0, \gamma_0) = (0, 0, 0)$, $\beta = 1$, $G = I/\beta$, $H = \mathbf{0}$ and $\hat{\sigma} = 1 - 10^{-8}$. From (8) (see also Remark 2.1(a)), for given $\tau \in (-1, 1)$ and $\theta \in \left(-\tau, \left(1 - \tau + \sqrt{5 + 2\tau - 3\tau^2} \right) / 2 \right)$, the error tolerance parameter $\tilde{\sigma}$ was defined as

$$\tilde{\sigma} = 0.99 \times \begin{cases} \min \left\{ \frac{(1 + \tau + \theta - \tau\theta - \tau^2 - \theta^2)(\tau - 1)}{\tau^2 - 2\theta + \theta^2}, 1 - \tau, 1 \right\}, & \text{if } \tau^2 - 2\theta + \theta^2 < 0, \\ \min \{ 1 - \tau, 1 \}, & \text{if } \tau^2 - 2\theta + \theta^2 \geq 0. \end{cases}$$

Moreover, we used the following stopping criterion

$$\|M(z_{k-1} - z_k)\|_\infty < 10^{-2},$$

where $z_k = (x_k, y_k, \gamma_k)$ and M is as in (18).

We considered six test images, which were scaled in intensity to $[0, 1]$, namely, (a) Barbara (512×512), (b) baboon (512×512), (c) cameraman (256×256), (d) Einstein (225×225), (e) clock (256×256), and (f) moon (347×403). All images were blurred by a Gaussian blur of size 9×9 with standard deviation 5 and then corrupted by a mean-zero Gaussian noise with variance 10^{-4} . The regularization parameter μ was set equal to 10^3 . The quality of the images was measured by the peak signal-to-noise ratio (PSNR) in decibel (dB):

$$\text{PSNR} = 10 \log_{10} \left(\frac{\bar{x}_{\max}^2}{\text{MSE}} \right)$$

where $\text{MSE} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (\bar{x}_{i,j} - x_{i,j})^2$, \bar{x}_{\max} is the maximum possible pixel value of the original image and \bar{x} and x are the original image and the recovered image, respectively.

Tables 1–6 report the numerical results of Algorithm 1, with some choices of (τ, θ) satisfying (8), for solving the six TV regularization problem instances. In the tables, “Out” and “Inner” denote the number of iterations and the total of inner iterations of the method, respectively, whereas “Time” is the CPU time in seconds. We mention that, for each problem instance, the final PSNRs were the same for all (τ, θ) considered. We displayed these values in the tables as well as the PSNRs of the corrupted images.

Table 1: Baboon 512×512

PSNR: input 19.35dB, output 20.71dB					
τ	θ	$\tilde{\sigma}$	Out	Inner	Time
0.0	1.00	0.990	131	11723	503.12
0.0	1.60	0.062	100	11748	495.38
0.9	1.00	0.099	71	7408	314.51
0.7	1.12	0.175	73	7224	312.27
0.7	1.15	0.142	71	7120	303.42
0.7	1.18	0.107	70	7205	309.97
0.8	1.12	0.074	76	8322	387.54
0.8	1.15	0.040	75	8672	393.05

Table 2: Barbara 512×512

PSNR: input 22.59dB, output 23.81dB					
τ	θ	$\tilde{\sigma}$	Out	Inner	Time
0.0	1.00	0.990	142	12910	574.58
0.0	1.60	0.062	105	12403	538.17
0.9	1.00	0.099	80	8620	411.47
0.7	1.12	0.175	84	8643	394.74
0.7	1.15	0.142	82	8583	391.85
0.7	1.18	0.107	82	8835	392.69
0.8	1.12	0.074	79	8665	371.47
0.8	1.15	0.040	79	9110	400.56

From the tables, we can see clearly the numerical benefits of using acceleration parameters $\tau > 0$ and $\theta > 1$. Note that Algorithm 1 with the choice $(\tau, \theta) = (0, 1)$ had the worst performance, in terms of the three performance measurements, for all problem instances. Note also that Algorithm 1 with $(\tau, \theta) = (0.9, 1)$ (0.9 was the best value for τ in [3]) performed better than Algorithm 1 with $(\tau, \theta) = (0, 1.6)$ (1.6 was the best value for θ in [2]), such behavior was also observed in [3] for the LASSO and ℓ_1 -regularized logistic regression problems. We stress that Algorithm 1 with $(\tau, \theta) = (0.8, 1.12)$ was faster in four (Barbara, cameraman, clock and moon) of six instances. Fig. 2 plots the original and corrupted images as well as the restored image by Algorithm 1 with $(\tau, \theta) = (0.8, 1.12)$ for the six instances. As a summary, we can conclude that combinations of the acceleration parameters τ and θ can also be efficient strategies in the inexact ADMMs for solving real-life applications.

Table 3: Cameraman 256×256

PSNR: input 21.02dB, output 25.14dB					
τ	θ	$\tilde{\sigma}$	Out	Inner	Time
0.0	1.00	0.990	135	13684	87.92
0.0	1.60	0.062	85	10382	64.21
0.9	1.00	0.099	72	8472	54.83
0.7	1.12	0.175	75	8473	52.77
0.7	1.15	0.142	74	8429	52.02
0.7	1.18	0.107	74	8709	53.41
0.8	1.12	0.074	71	8460	51.83
0.8	1.15	0.040	71	8756	53.75

Table 4: Clock 256×256

PSNR: input 22.68dB, output 27.44dB					
τ	θ	$\tilde{\sigma}$	Out	Inner	Time
0.0	1.00	0.990	130	12666	82.03
0.0	1.60	0.062	84	10104	64.97
0.9	1.00	0.099	69	7746	53.39
0.7	1.12	0.175	72	7807	51.95
0.7	1.15	0.142	73	7985	52.01
0.7	1.18	0.107	70	7767	55.67
0.8	1.12	0.074	68	7740	50.03
0.8	1.15	0.040	67	7994	51.48

Table 5: Einstein 225×225

PSNR: input 23.70dB, output 28.24dB					
τ	θ	$\tilde{\sigma}$	Out	Inner	Time
0.0	1.00	0.990	120	10506	56.86
0.0	1.60	0.062	88	9968	52.84
0.9	1.00	0.099	72	7560	40.47
0.7	1.12	0.175	68	6573	34.54
0.7	1.15	0.142	74	7631	40.88
0.7	1.18	0.107	73	7566	40.69
0.8	1.12	0.074	71	7685	40.00
0.8	1.15	0.040	70	7830	40.27

Table 6: Moon 347×403

PSNR: input 25.57dB, output 28.28dB					
τ	θ	$\tilde{\sigma}$	Out	Inner	Time
0.0	1.00	0.990	128	11684	249.27
0.0	1.60	0.062	88	10239	215.91
0.9	1.00	0.099	72	7828	168.30
0.7	1.12	0.175	76	7921	170.00
0.7	1.15	0.142	74	7796	205.34
0.7	1.18	0.107	73	7909	194.64
0.8	1.12	0.074	68	7412	161.05
0.8	1.15	0.040	67	7711	181.89

4 Final remarks

We proposed an inexact symmetric proximal ADMM for solving linearly constrained optimization problems. Under appropriate hypotheses, the global $\mathcal{O}(1/\sqrt{k})$ pointwise and $\mathcal{O}(1/k)$ ergodic convergence rates of the proposed method were established for a domain of the acceleration parameters, which is consistent with the largest known one in the exact case. Numerical experiments were carried out in order to illustrate the numerical behavior of the new method. They indicate that the proposed scheme represents an useful tool for solving real-life applications. To the best of our knowledge, this was the first time that an inexact variant of the symmetric proximal ADMM was proposed and analyzed.

References

- [1] V. A. Adona, M. L. N. Gonçalves, and J. G. Melo. Iteration-complexity analysis of a generalized alternating direction method of multipliers. *J. Glob. Optim.*, 73(2):331–348, 2019.
- [2] 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.*,



Figure 2: Results on the images (top to bottom): “Baboon”, “Barbara”, “Cameraman”, “Clock”, “Einstein” and “Moon”. First column is the original images, the second is blurred and noisy images, and the third is the restored images by Algorithm 1 with $(\tau, \theta) = (0.8, 1.12)$.

- 182(2):640–666, 2019.
- [3] V. A. Adona, M. L. N. Gonçalves, and J. G. Melo. An inexact proximal generalized alternating direction method of multipliers. *Comput. Optim. Appl.*, 2020.
- [4] 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.
- [5] 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, 2008.
- [6] J. Bai, J. Li, F. Xu, and H. Zhang. Generalized symmetric ADMM for separable convex optimization. *Comput. Optim. Appl.*, 70(1):129–170, 2018.
- [7] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic Press, New York, 1982.
- [8] 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.
- [9] X. Chang, S. Liu, P. Zhao, and D. Song. A generalization of linearized alternating direction method of multipliers for solving two-block separable convex programming. *J. Comput. Appl. Math.*, 357:251 – 272, 2019.
- [10] J. Eckstein. Some saddle-function splitting methods for convex programming. *Optim. Method Softw.*, 4(1):75–83, 1994.
- [11] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Programming*, 55(3, Ser. A):293–318, 1992.
- [12] J. Eckstein and W. Yao. Approximate ADMM algorithms derived from Lagrangian splitting. *Comput. Optim. Appl.*, 68(2):363–405, 2017.
- [13] J. Eckstein and W. Yao. Relative-error approximate versions of Douglas–Rachford splitting and special cases of the ADMM. *Math. Programming*, 170(2):417–444, 2018.
- [14] 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. Elsevier, 1983.
- [15] 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. Elsevier, Amsterdam, 1983.

- [16] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.*, 2:17–40, 1976.
- [17] B. Gao and F. Ma. Symmetric alternating direction method with indefinite proximal regularization for linearly constrained convex optimization. *J. Optim. Theory Appl.*, 176(1):178–204, 2018.
- [18] R. Glowinski. *Numerical Methods for Nonlinear Variational Problems*. Springer Series in Computational Physics. Springer-Verlag, 1984.
- [19] R. Glowinski and A. Marroco. Sur l’approximation, par éléments finis d’ordre un, et la résolution, par pénalisation-dualité, d’une classe de problèmes de Dirichlet non linéaires. *R.A.I.R.O.*, 9(R2):41–76, 1975.
- [20] M. L. N. Gonçalves, M. M. Alves, and J. G. Melo. Pointwise and ergodic convergence rates of a variable metric proximal alternating direction method of multipliers. *J. Optim. Theory Appl.*, 177(2):448–478, 2018.
- [21] M. L. N. Gonçalves, J. G. Melo, and R. D. C. Monteiro. On the iteration-complexity of a non-Euclidean hybrid proximal extragradient framework and of a proximal ADMM. *Optimization*, 69(4):847–873, 2019.
- [22] B. He, H. Liu, Z. Wang, and X. Yuan. A strictly contractive Peaceman–Rachford splitting method for convex programming. *SIAM J. Optim.*, 24(3):1011–1040, 2014.
- [23] B. He, F. Ma, and X. Yuan. Convergence study on the symmetric version of ADMM with larger step sizes. *SIAM J. Imaging Sci.*, 9(3):1467–1501, 2016.
- [24] B. He and X. Yuan. On non-ergodic convergence rate of Douglas-Rachford alternating direction method of multipliers. *Numer. Math.*, 130(3):567–577, 2015.
- [25] M. Ng, F. Wang, and X. Yuan. Inexact alternating direction methods for image recovery. *SIAM J. Sci. Comput.*, 33(4):1643–1668, 2011.
- [26] J. Nocedal and S. J. Wright. *Numerical Optimization 2nd*. Springer, New York, 2006.
- [27] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.
- [28] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1):259 – 268, 1992.
- [29] Y. Shen, Y. Zuo, and A. Yu. A partial PPa S-ADMM for multi-block for separable convex optimization with linear constraints. *Optimization online*, 2020.
- [30] H. Sun, M. Tian, and M. Sun. The symmetric ADMM with indefinite proximal regularization and its application. *J. Inequal. Appl.*, 2017(1):1–22, 2017.
- [31] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM J. Imaging Sci.*, 1(3):248–272, 2008.
- [32] Z. Wu and M. Li. An LQP-based symmetric alternating direction method of multipliers with larger step sizes. *J. Oper. Res. Soc. China*, 7(2):365–383, 2019.

- [33] Z. Wu, M. Li, D. Z. W. Wang, and D. Han. A symmetric alternating direction method of multipliers for separable nonconvex minimization problems. *Asia-Pac. J. Oper. Res.*, 34(06):1750030, 2017.
- [34] Z. Wu, F. Liu, and M. Li. A proximal Peaceman–Rachford splitting method for solving the multi-block separable convex minimization problems. *Int. J. Comput. Math.*, 96(4):708–728, 2018.
- [35] 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.
- [36] J. Yang, W. Yin, Y. Zhang, and Y. Wang. A fast algorithm for edge-preserving variational multichannel image restoration. *SIAM J. Imaging Sci.*, 2(2):569–592, 2009.